



# Deep learning for semantic segmentation in multimodal medical images : application on brain tumor segmentation from multimodal magnetic resonance imaging

Tongxue Zhou

## ► To cite this version:

Tongxue Zhou. Deep learning for semantic segmentation in multimodal medical images : application on brain tumor segmentation from multimodal magnetic resonance imaging. Image Processing [eess.IV]. Normandie Université, 2022. English. NNT : 2022NORMIR01 . tel-03670606

**HAL Id: tel-03670606**

**<https://theses.hal.science/tel-03670606>**

Submitted on 17 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

# THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Informatique

Préparée au sein de INSA ROUEN NORMANDIE

## Deep Learning for Semantic Segmentation in Multimodal Medical Images: Application on Brain Tumor Segmentation from Multimodal Magnetic Resonance Imaging

Présentée et soutenue par

Tongxue ZHOU

Thèse soutenue publiquement le 31/01/2022  
devant le jury composé de

M. Pierrick COUPÉ,	Directeur de recherche CNRS à l'Université de Bordeaux	Rapporteur
M. Gaël VAROQUAUX,	Directeur de recherche INRIA à l'INRIA Saclay	Rapporteur
Mme Christine FERNANDEZ-MALOIGNE,	Professeure à l'Université de Poitiers	Présidente du jury
M. Nicolas DUCHATEAU,	Maitre de conférences à l'Université Lyon 1	Examineur
M. Stéphane CANU,	Professeur à l'INSA-Rouen Normandie	Directeur de thèse
Mme Su RUAN,	Professeure à l'Université de Rouen Normandie	Codirectrice de thèse

Thèse dirigée par STÉPHANE CANU et SU RUAN, laboratoire LITIS



# Abstract

In this thesis, we develop four deep learning based brain tumor segmentation methods with multimodal MRI images. The first two methods focus on the segmentation with complete modalities, and the last two methods aim to tackle the segmentation with missing modalities. We first present a multi-modal brain tumor segmentation network based on attention mechanism to fuse MRI images and context constraint to limit segmentation regions. Considering the correlation between different MR modalities, we propose a second method to model this correlation in the feature latent space with a non-linear model and KL divergence in order to fuse images in an efficient way. This method uses a novel tri-attention fusion block which consists of a modality attention module, a spatial attention module, and a correlation attention module. Since it's common to have missing imaging modalities in clinical practice, we propose two methods for brain tumor segmentation with missing MR modalities. The correlation between multimodal MRI images is exploited again to maintain the information of the missing images. In the first method, the network is trained with complete modalities via multi-source correlation, and tested with missing modalities in which the missing modalities are replaced by the most similar ones. To retrieve more precisely the lost information when modalities are missing, we proposed a second method to generate a feature-enhanced missing modality under the multi-source correlation condition. In addition, a KL divergence is applied to guarantee the similarity between the estimated correlated feature representation and the original feature representation. We evaluated these proposed methods on public multi-modal brain tumor segmentation datasets and demonstrated the effectiveness of these proposed methods.

**Keywords :** Deep learning, MRI, Brain tumor segmentation, Fusion, Multi-modality, Multi-source correlation, Missing data





# Résumé

Dans cette thèse, nous développons quatre méthodes de segmentation des tumeurs cérébrales à partir d'images IRM multimodales, basées sur l'apprentissage profond. Les deux premières méthodes se concentrent sur la segmentation tumorale avec des modalités complètes. Tandis que les deux dernières méthodes réalisent la segmentation lorsque certaines modalités sont manquantes. Nous présentons d'abord le premier réseau de neurones profond pour la segmentation d'images multimodales de tumeurs cérébrales. Celui-ci est basé à la fois sur un mécanisme d'attention afin de fusionner les images multimodales et sur une contrainte liée au contexte afin de limiter la région de la segmentation. Compte tenu qu'il existe une corrélation entre les modalités, nous proposons une seconde méthode exploitant cette corrélation multimodale dans l'espace des caractéristiques latentes, à l'aide d'un modèle non linéaire et de la divergence de KL, dans le but de fusionner les images de manière efficace. Cette méthode innovante utilise un bloc de fusion à « triple attention », qui comprend un module d'attention basé sur les modalités, un autre basé sur l'espace spatial, et un troisième basé sur la corrélation multimodale. Il est fréquent que certaines modalités soient manquantes en pratique clinique. Nous proposons donc deux méthodes permettant la segmentation de tumeurs cérébrales avec des modalités manquantes. Nous exploitons à nouveau la corrélation entre les modalités afin de conserver les informations des images manquantes. Dans notre première méthode, le réseau est entraîné avec la totalité des modalités à l'aide de la corrélation multimodale. Il est ensuite testé dans les cas où certaines modalités sont manquantes, et celles-ci sont remplacées par les modalités les plus proches. Afin de retrouver plus précisément les informations manquantes, nous proposons une seconde méthode permettant de générer les caractéristiques de la modalité manquante, grâce à une corrélation multimodale. De plus, nous appliquons une divergence KL pour garantir la similarité entre les caractéristiques générées et celles de la modalité originale. Nous avons démontré l'efficacité et la bonne performance des méthodes proposées en utilisant une base d'images multimodales publique.

**Mots clés :** Apprentissage profond, IRM, Segmentation de tumeur cérébrale, Fusion, Multi-modalités, Corrélation multi-source, Données manquantes



# Acknowledgment

First and foremost, I am extremely grateful to my supervisors, Prof. Stéphane Canu and Prof. Su Ruan for their invaluable advice and continuous support during my PhD study. Without their assistance and dedicated involvement, this thesis would have never been accomplished.

I would like to express my gratitude to Prof. Su Ruan for her constructive guidance during my PhD study. Her profound knowledge and instructive comments helped me in the research, including method development, experiment design and paper writing. Prof. Su Ruan is a hard-working researcher, which impresses me and motivates me to be a researcher like her. I am also very grateful for her help and support in my daily life, especially during the coronavirus pandemic period.

I would also thank my supervisor, Prof. Stéphane Canu, for his extensive knowledge, valuable advice and unwavering support. He provided me many opportunities to participate the research activities, such as supporting me to attend international conferences and summer schools.

I would like to express my sincere gratitude for all the members of my thesis committee. I would like to thank Dr. Pierrick Coupé, Director of Research at Université de Bordeaux, and Dr. Gaël Varoquaux, Director of Research at INRIA Saclay, for spending their valuable time to review my thesis. I would also like to thank Dr. Christine Fernandex-Maloigne, Prof. at Université de Poitiers, and Dr. Nicolas Duchateau, Associate Prof. at Université Lyon 1, for accepting to be examiners in my thesis defense.

I had great pleasure of working with my colleagues in LITIS QuantIF team. They gave me a great amount of assistance in research and daily life. I would like to extend my thanks to my friends in LITIS lab for their friendship and the warmth. I would also like to thank Madame Brigitte and Madame Sandra for their help during my PhD study.

Many thanks should also go to the China Scholarship Council for the financial support during my PhD study.

Last but not the least, I must express my very profound gratitude to my beloved parents with continuous support and encouragement throughout the years of my study. Meanwhile, special thanks go to my boyfriend Maël for his countless support and accompany. I would also like to thank his family and all the friends from Rennes for their kindness and warmheartedness.



# Contents

<b>Abstract</b>	<b>i</b>
<b>Résumé</b>	<b>iii</b>
<b>Acknowledgment</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xvii</b>
<b>Acronyms</b>	<b>xxi</b>
<b>1 Biomedical Background</b>	<b>1</b>
1.1 Introduction of Magnetic Resonance Imaging . . . . .	1
1.1.1 Principle of Magnetic Resonance Imaging . . . . .	1
1.1.2 MRI Sequences . . . . .	2
1.1.3 Variations of MRI . . . . .	3
1.1.4 Role of MRI in Medical Diagnosis . . . . .	4
1.2 Brain Tumor Segmentation . . . . .	5
1.2.1 Brain Tumor Overview . . . . .	5
1.2.2 Diagnosis of Brain Tumor . . . . .	7
1.2.3 Treatment Planning of Brain Tumor . . . . .	8
1.2.4 Challenges for Brain Tumor Segmentation . . . . .	9
1.3 Conclusion . . . . .	10
<b>2 A Review: Deep Learning for Medical Image Segmentation using Multi-modality Fusion</b>	<b>11</b>
2.1 Introduction . . . . .	12
2.2 Related Works . . . . .	15
2.2.1 Brief Introduction of Deep learning . . . . .	15
2.2.2 Multi-modal Medical Image Segmentation . . . . .	16
2.3 Data Preparation . . . . .	18

# CONTENTS

---

2.3.1	Data Dimension . . . . .	18
2.3.2	Pre-processing . . . . .	19
2.3.3	Data Augmentation . . . . .	19
2.3.4	Post-processing . . . . .	19
2.4	Multi-modal Medical Image Segmentation Networks . . . . .	19
2.4.1	Input-level Fusion Network . . . . .	20
2.4.2	Layer-level Fusion Network . . . . .	21
2.4.3	Decision-level Fusion Network . . . . .	23
2.5	Discussion and Methodologies to be Developed . . . . .	23
2.6	Conclusion . . . . .	26
<b>3</b>	<b>Fusion based on Attention Mechanism for Multi-modal MR Brain Tumor Segmentation</b>	<b>27</b>
3.1	Introduction . . . . .	28
3.2	Related Works . . . . .	29
3.3	Methodology . . . . .	31
3.3.1	The Three-stage Segmentation Network . . . . .	31
3.3.2	Initial Segmentation Network (Stage 1) . . . . .	31
3.3.3	Fusion Block based on Attention Mechanism (Stage 2) . . . . .	33
3.3.4	Final Segmentation Network (Stage 3) . . . . .	35
3.3.5	Loss Function . . . . .	35
3.4	Experimental Setup . . . . .	36
3.4.1	Data and Pre-processing . . . . .	36
3.4.2	Implementation Details . . . . .	36
3.4.3	Evaluation Metrics . . . . .	36
3.5	Experimental Results . . . . .	37
3.5.1	Quantitative Analysis . . . . .	37
3.5.1.1	Evaluation of Our Method . . . . .	37
3.5.1.2	Comparison with the State-of-the-art Methods . . . . .	39
3.5.2	Qualitative Analysis . . . . .	40
3.5.2.1	Evaluation of Our Method . . . . .	40
3.5.2.2	Comparison with the State-of-the-art Methods . . . . .	40
3.6	Discussion and Conclusion . . . . .	42
<b>4</b>	<b>Fusion based on Feature Correlation in Latent Space for Multi-modal Brain Tumor Segmentation</b>	<b>45</b>
4.1	Introduction . . . . .	46
4.2	Related work . . . . .	46
4.3	Methodology . . . . .	48
4.3.1	Network Architecture . . . . .	48
4.3.2	Feature Correlation and Tri-attention Fusion Strategy . . . . .	49
4.4	Experimental Setup . . . . .	52
4.4.1	Data and Pre-processing . . . . .	52
4.4.2	Implementation Details . . . . .	52

## CONTENTS

---

4.4.3	Evaluation Metrics . . . . .	54
4.5	Experimental Results . . . . .	54
4.5.1	Quantitative Analysis . . . . .	54
4.5.1.1	Ablation Study . . . . .	54
4.5.1.2	Comparison with the State-of-the-art Methods . . . . .	55
4.5.2	Qualitative Analysis . . . . .	59
4.6	Discussion . . . . .	59
4.6.1	Performance Analysis on Correlation Expression . . . . .	59
4.6.2	Performance Analysis on Correlation Attention Module . . . . .	60
4.6.3	Visualization of Feature Maps . . . . .	62
4.7	Conclusion . . . . .	63
<b>5</b>	<b>Multi-source Correlation Guided Brain Tumor Segmentation Network with Missing Modalities</b>	<b>67</b>
5.1	Introduction . . . . .	68
5.2	Related Works . . . . .	69
5.3	Methodology . . . . .	71
5.3.1	Motivation of the Proposed Method . . . . .	71
5.3.2	Overview of the Method 1 . . . . .	71
5.3.3	Modeling the Multi-source Correlation . . . . .	72
5.3.4	Overview of the Method 2 . . . . .	73
5.3.5	Synthesizing the Missing Modality . . . . .	74
5.3.6	Brain Tumor Segmentation . . . . .	77
5.3.7	Loss Functions . . . . .	77
5.4	Experimental Setup . . . . .	78
5.4.1	Dataset and Pre-processing . . . . .	78
5.4.2	Implementation Details . . . . .	78
5.4.3	Evaluation Metrics . . . . .	78
5.5	Experimental Results . . . . .	78
5.5.1	Quantitative Analysis . . . . .	79
5.5.1.1	Ablation Study . . . . .	79
5.5.1.2	Comparison with the State-of-the-art Methods . . . . .	80
5.5.2	Qualitative Analysis . . . . .	82
5.6	Discussion and Conclusion . . . . .	88
<b>6</b>	<b>General Conclusions and Perspectives</b>	<b>91</b>
6.1	Conclusions . . . . .	91
6.2	Perspectives . . . . .	92
	<b>List of publications</b>	<b>95</b>
	<b>Bibliography</b>	<b>97</b>





# List of Figures

1.1	The schematic diagram of MRI scanner [3]. . . . .	2
1.2	Conceptual diagram of obtaining signals from protons in the body. . . . .	3
1.3	Examples of FLAIR, T1, T1c and T2 images from the same patient [7]. . . . .	3
1.4	Schematic diagram of primary brain tumors [18]. . . . .	6
1.5	Schematic diagram of secondary brain tumors [18]. . . . .	6
2.1	The tendency of multi-modal medical image segmentation in deep learning from 2014 to 2020. . . . .	14
2.2	The tendency of relative research field with/without deep learning. . . . .	14
2.3	The multi-modal medical images, (a)-(c) are the commonly used multi-modal medical images [76] and (d)-(g) are the different sequences of brain MRI [7]. . . . .	15
2.4	The pipeline of multi-modal medical image segmentation based on deep learning. . . . .	17
2.5	The generic categorization of the fusion strategy. . . . .	20
2.6	The generic network architecture of the input-level fusion. . . . .	20
2.7	The generic network architecture of the layer-level fusion. . . . .	22
2.8	The generic network architecture of the decision-level fusion. . . . .	23
3.1	An example from BraTS 2017 dataset [137]. The first four images from left to right show the MRI modalities: Fluid Attenuated Inversion Recover (FLAIR), contrast enhanced T1-weighted (T1c), T1-weighted (T1), T2-weighted (T2) images, and the fifth image is the ground-truth labels, Net&Ncr is shown in red, edema in orange and enhancing tumor in white, Net refers non-enhancing tumor and Ncr refers necrotic tumor. . . . .	30
3.2	The graphical concept of our proposed method, to simplify the presentation, we ignore the deep supervision part in these three segmentation networks, the details are shown in [120]. . . . .	31

---

LIST OF FIGURES

---

3.3	Top: The architecture scheme of initial segmentation network. The four modalities are concatenated channel by channel in input space and the output is the segmentation predictions of the three tumor sub-regions. Bottom: The architecture scheme of our proposed res_dil block (left) and deep supervision (right). IN refers instance normalization, Dil_conv refers the dilated convolution (rate = 2, 4, respectively), we refer to the vertical depth as level, with higher levels being higher spatial resolution. In the deep supervision part, $Input_n$ refers the output of res_dil block of the $n_{th}$ level in the decoder, $Output_n$ refers the segmentation result of the $n_{th}$ level in the decoder. . . . .	32
3.4	The architecture scheme of fusion network. Each imaging modality (FLAIR, T1, T1c, T2) is encoded by a single encoder to obtain the individual latent representations $(z_1, z_2, z_3, z_4)$ , and the context constraint can provide boundary information to refine the segmentation result. Then, the five encoders are fused into the shared representation space with the fusion block. Finally the fused latent representation $Z_f$ is decoded by the decoder to obtain the segmentation result. Here we present the segmentation network architecture of enhancing tumor, it is same for other tumor regions.	33
3.5	The architecture scheme of fusion block. The individual latent representations $(z_1, z_2, z_3, z_4)$ are first concatenated as the input of the attention mechanism $Z$ . Then, they are recalibrated along channel attention module and spatial attention module to achieve the $Z_s$ and $Z_c$ . Finally, they are added to obtain the fused latent representation $Z_f$ . . . . .	34
3.6	Qualitative comparison among different strategies of our method in Stage 2 on several examples. We denote the DSC on each result. Net&Ncr is shown in red, edema in orange and enhancing tumor in white. . . . .	41
3.7	Qualitative results in the three stages of our method on several examples. We denote the DSC on each result in Stage 1 and Stage 3. Label 1 (red): net&ncr, Label 2 (orange): edema, Label 3 (white): enhancing tumor. The green bounding box emphasizes the differences of segmentation results among different methods. . . . .	41
3.8	Qualitative results between our method (8 initial filters) and method from [86] (16 initial filters) on several examples. We denote the DSC on each result. Net&Ncr is shown in red, edema in orange and enhancing tumor in white. . . . .	42
4.1	An example from BraTS 2018 dataset[7]. The first four images from left to right show the MRI modalities: T1-weighted (T1), FLAIR, contrast enhanced T1-weighted (T1c), T2-weighted (T2) images, and the fifth image is the ground-truth labels created by experts. The color is used to distinguish the different tumor regions: red: necrotic and non-enhancing tumor, yellow: edema, green: enhancing tumor. . . . .	47

## LIST OF FIGURES

---

4.2	Overview of our proposed segmentation network. The backbone is a multi-encoder based 3D U-Net, the separate encoders enable the network to extract the independent feature representations. The proposed dual-attention fusion block is to re-weight the feature representations along modality and space paths. The tri-attention fusion block consists of the dual-attention fusion and a correlation attention module. . . . .	49
4.3	Joint intensity distributions of MR images: (a) T1-FLAIR, (b) T1-T1c, (c) T1-T2, (d) FLAIR-T1c, (e) FLAIR-T2, (f) T1c-T2. The intensity of the first modality is read on abscissa axis and that of the second modality on the ordinate axis. . . . .	50
4.4	Architecture of the tri-attention fusion strategy. The individual feature representations ( $Z_1, Z_2, Z_3, Z_4$ ) are first concatenated, then they are re-weighted by dual-attention fusion block along modality attention module and spatial attention module to achieve the modality attention representation $Z_{im}$ and spatial attention representation $Z_{is}$ . In addition, the correlation attention module is used to constrain the spatial-attention representations to learn segmentation-related representation. Finally, the $Z_{im}$ and $Z_{is}$ are added to obtain the fused feature representation $Z_{if}$ . . . .	51
4.5	Comparison of different weight coefficients in the loss function. Average DSC Score vs $\lambda$ and Average HD vs $\lambda$ . . . . .	53
4.6	Box plots of DSC for the three compared methods in Table 4.1 with regard to the three tumor regions: Enhancing Tumor (ET), Whole Tumor (WT) and Tumor Core (TC). Method (1) is shown in pink, method (2) in blue and method (3) in green. . . . .	56
4.7	Box plots of Hausdorff Distance for the three compared methods in Table 4.1 with regard to the three tumor regions: ET, WT and TC. Method (1) is shown in pink, method (2) in blue and method (3) in green. . . . .	56
4.8	Visualization of several segmentation results, and the related DSC of whole tumor is presented. (a) Baseline (b) Baseline with dual attention fusion (c) Baseline with tri-attention fusion (d) Ground-truth. Red: necrotic and non-enhancing tumor core; Yellow: edema; Green: enhancing tumor. Blue arrow emphasizes the mis-segmentation of the methods. . . . .	60
4.9	Box plots of DSC for the two compared correlation expressions in Table 4.5 with regard to the three tumor regions: ET, WT and TC. Linear expression is shown in pink, Non-linear expression in blue. . . . .	61
4.10	Box plots of Hausdorff Distance for the two compared correlation expressions in Table 4.5 with regard to the three tumor regions: ET, WT and TC. Linear expression is shown in pink, Non-linear expression in blue. . .	62
4.11	Visualization of effectiveness of proposed correlation attention module. First column: input images, second column: baseline, third column: baseline + dual attention module, fourth column: baseline + tri-attention module (added on fused feature representation), fifth column: (Ours, added on spatial-attention feature representation), sixth column: ground-truth. .	64

## LIST OF FIGURES

---

- 5.1 A training sample from BraTS 2018 dataset. From left to right: FLAIR, contrast enhanced T1-weighted (T1c), T1-weighted (T1), T2-weighted (T2) images, and the ground-truth (GT). Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor. . . . . 69
- 5.2 A schematic overview of the proposed network. Each input modality is encoded by individual encoder to obtain the individual representation. The proposed Correlation Model (CM) block and Fusion block (Fusion) project the individual representations into a fused representation, which is finally decoded to form the reconstruction modalities and the segmentation result. 72
- 5.3 Architecture of correlation model. MPE Module first maps the individual representation  $f_i(X_i|\theta_i)$  to a set of correlation parameters  $\Gamma_i$ , under these parameters, LCE Module transforms all the individual representations to form a latent multi-source correlation representation  $F_i(X_i|\theta_i)$ . . . . . 73
- 5.4 An overview of the proposed network, consisting of a Feature-enhanced Generator (FeG), a Correlation Constraint (CC) block and a segmentation network. The feature-enhanced generator utilizes the available modalities to generate a feature-enhanced missing modality  $X_5$ . Then the complete modalities are input to the segmentation network for the final prediction. In addition, the correlation constraint block are used to guide both the feature-enhanced generator and segmentation network via exploiting the multi-source correlation. . . . . 74
- 5.5 Architecture of the Feature-enhanced Generator (FeG). Left: multi-encoder feature-enhanced generator shown in Figure 5.4; Right: we take one encoder and the decoder as an example. The left series of blocks connected by the blue arrows represent the encoder, and right series of blocks connected by the yellow arrows represent the decoder. In order to maintain the spatial information, we replace the pooling operation with the convolution block with stride=2. The res\_dil block used in both encoder and decoder is to increase the receptive field to capture more features, and the rate denotes the dilated rate. . . . . 75
- 5.6 Architecture of the Correlation Constraint (CC) block.  $X_1, X_2, X_3, X_4$  are the input modalities, which are possible to be missing.  $X_5$  is the generated average modality. CPEM first maps the individual representation  $f_i$  to a set of independent parameters  $\Gamma_i$ , under these parameters, LCEM transforms all the individual representations to form correlated representation  $F_i$ . In addition, the KL based CCL is employed to guide the whole training process. 76
- 5.7 Examples of the segmentation results on full modalities of the first proposed method. (1) denotes the baseline, (2) denotes baseline with reconstruction. Red: necrotic and non-enhancing tumor core; Orange: edema; White: enhancing tumor. . . . . 85

## LIST OF FIGURES

---

5.8	Examples of the segmentation results on missing modalities of the first proposed method. Red: necrotic and non-enhancing tumor core; Orange: edema; White: enhancing tumor. . . . .	85
5.9	Segmentation results of the second proposed method, the DSC of whole tumor, tumor core and enhancing tumor is denoted under each image. On the top, the first row shows the four MR modalities, FLAIR, T1c, T1 and T2. On the bottom, the three rows show the segmentation results of different methods. The first four columns show the different missing modalities situations. The last column shows the ground-truth segmentation. Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor. . . . .	86
5.10	Generation and segmentation results of the second proposed method. On the top, the first row shows the four MR modalities (FLAIR, T1c, T1 and T2) and the segmentation ground-truth. On the bottom, the four columns show the different missing modalities situations. The first row shows the ground-truth average modality. The second row shows the corresponding generated average modality. The last row shows the segmentation results of our method. Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor. . .	87



# List of Tables

1.1	The gray distribution of different tissues in multiple MRI sequences [8] . . .	4
2.1	Summary of deep learning network architectures, ILSVRC: ImageNet Large Scale Visual Recognition Challenge. . . . .	16
2.2	Summary of the multi-modal medical segmentation datasets. . . . .	17
2.3	Summary of the commonly used evaluation metrics. False Positive (FP), True Positive (TP), False Negative (FN) and True Negative (TN) are the number of false positive voxels, true positive voxels, false negative voxels and true negative voxels, respectively. $d$ is the Euclidean distance, $S$ and $R$ are the two sets of surface points of the prediction and the real annotation. $ X $ and $ Y $ are the number of voxels of the prediction and real annotation, respectively. . . . .	18
2.4	Summary of the deep learning approaches for multi-modal medical image segmentation, the bold presents the best performance in the challenge. The acronyms in results are: cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), the symbol * indicates the method has available code.	24
3.1	Segmentation results of fusion network, bold results show the best scores for each tumor region, backbone refers to the four-encoder based network without new loss and context constraint, Net refers non-enhancing tumor, Necr refers necrotic tumor, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	38
3.2	The choice of the coefficient $\alpha$ in the new loss function. . . . .	38
3.3	Segmentation results of our method on BraTS 2017 dataset, bold results show the best scores for each tumor region, Stage 1 refers to initial segmentation network, Stage 2 refers to fusion network and Stage 3 refers to final segmentation network, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	38
3.4	Comparison of our proposed method and other related methods on BraTS 2017, bold results show the best scores for each tumor region, and underline results refer the second best results, AVG denotes the average results on the three tumor regions. . . . .	39



## LIST OF TABLES

---

3.5	Computational complexity comparison of different methods including data dimension, input size, number of layer, number of initial filter, data augmentation, post-processing, GPU and training time. "-" indicates that the information is not provided in the published paper. . . . .	40
4.1	Evaluation of our proposed method on BraTS 2018 training dataset, (1) Baseline (2) Baseline + Dual attention fusion (3) Baseline + Tri-attention fusion, ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	55
4.2	Evaluation of our proposed method on BraTS 2017 training dataset, (1) Proposed method in Chapter 3, (2) Proposed method in Chapter 4, ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	55
4.3	Comparison of different methods on BraTS 2018 validation dataset, ET, WT, TC denote enhancing tumor, whole tumor, tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, underline results denote the second best results. . . . .	58
4.4	Computational complexity comparison of different methods including data dimension, input size, number of layer, number of initial filter, data augmentation, post-processing, GPU and training time. "-" indicates that the information is not provided in the published paper. . . . .	58
4.5	Comparison of segmentation accuracy of different correlation expressions. Avg denotes the average results on the three tumor regions, bold results denote the best results, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	61
4.6	Comparison of segmentation accuracy of correlation attention module in different layer of the network. ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results. . . . .	63
5.1	Comparison of different strategies in our proposed method 1 on full modalities on BraTS 2018 training set, $\uparrow$ denotes the improvement compared to the previous method, bold results show the best scores for each tumor region, AVG denotes the average results on the three target regions, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	79

## LIST OF TABLES

---

5.2	Comparison of different strategies in our proposed method 2 in terms of DSC (%) on BraTS 2018 dataset, ● denotes the present modality and ○ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).	80
5.3	Comparison of different strategies in our proposed method 2 in terms of HD on BraTS 2018 dataset, ● denotes the present modality and ○ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).	81
5.4	Comparison of different methods on full modalities on BraTS 2018 validation set, bold results show the best scores, and underline results refer the second best results. . . . .	82
5.5	Comparison of different strategies in our proposed method 1 in the terms of DSC (%) on BraTS 2018 dataset, ○ denotes the missing modality and ● denotes the present modality, WoCM denotes our method without CM, bold results denote the best score, Average denotes the average results on one target region across all the situations, * denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ). . . . .	83
5.6	Comparison of different methods in terms of DSC (%) on BraTS 2018 dataset, ● denotes the present modality and ○ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations. . . . .	84



# Acronyms

<b>ARVD</b>	Absolute Relative Volume Difference.	17
<b>BF</b>	Blood Flow.	4
<b>BV</b>	Blood Volume.	4
<b>CD</b>	Correlation Description.	50
<b>CNN</b>	Convolutional Neural Network.	13, 16
<b>CNS</b>	Central Nervous System.	5
<b>CPU</b>	Central Processing Unit.	11
<b>CRF</b>	Conditional Random Field.	18, 57
<b>CSF</b>	Cerebro Spinal Fluid.	2
<b>CT</b>	Computed Tomography.	5, 13, 20, 21, 92
<b>CT-MRI</b>	Computed Tomography - Magnetic Resonance Imaging.	16
<b>DSC</b>	Dice Similarity Coefficient.	17, 36, 53, 78
<b>ED</b>	Edema.	22
<b>ET</b>	Enhancing Tumor.	xiii, 22, 36, 52, 55, 59, 78
<b>FCN</b>	Fully Convolutional Network.	18, 29, 30, 57
<b>FLAIR</b>	Fluid Attenuated Inversion Recover.	xi, xii, xiii, 2, 13, 28, 30, 36, 46, 69
<b>fMRI</b>	Functional MRI.	3
<b>FN</b>	False Negative.	xvii, 17
<b>FP</b>	False Positive.	xvii, 17
<b>GAN</b>	Generative Adversarial Network.	20, 21, 25, 70, 73, 88, 93

- GPU** Graphics Processing Unit. 11
- HD** Hausdorff Distance. 17, 36, 37, 39, 53, 78
- MRA** Magnetic Resonance Angiography. 4
- MRI** Magnetic Resonance Imaging. 1, 2, 3, 4, 5, 7, 8, 9, 13, 16, 18, 20, 21, 22, 28, 29, 31, 36, 46, 52, 68, 69, 70, 78, 89, 91
- MRV** Magnetic Resonance Venography. 4
- MTT** Mean Transit Time. 4
- NCR** Necrotic. 22
- NET** Non-Enhancing Tumor. 22
- PET** Positron Emission Tomography. 7, 8, 13, 92
- ReLU** Rectified Linear Unit. 11, 16
- RF** Radio Frequency. 2, 5
- RNN** Recurrent Neural Network. 18
- scSE** spatial and channel SE. 30
- SE** Squeeze and Excitation. 30
- SGD** Stochastic Gradient Descent. 16
- SSIM** Structural Similarity Index Metric. 77
- T** Tesla. 1
- TC** Tumor Core. xiii, 36, 52, 55, 59, 78
- TE** Time to Echo. 2
- TN** True Negative. xvii, 17
- TP** True Positive. xvii, 17
- TR** Repetition Time. 2
- TTP** Time To Peak. 4
- WHO** World Health Organization. 5
- WT** Whole Tumor. xiii, 36, 52, 55, 59, 78

# Chapter 1

## Biomedical Background

### Contents

<b>1.1</b>	<b>Introduction of Magnetic Resonance Imaging</b>	<b>1</b>
1.1.1	Principle of Magnetic Resonance Imaging	1
1.1.2	MRI Sequences	2
1.1.3	Variations of MRI	3
1.1.4	Role of MRI in Medical Diagnosis	4
<b>1.2</b>	<b>Brain Tumor Segmentation</b>	<b>5</b>
1.2.1	Brain Tumor Overview	5
1.2.2	Diagnosis of Brain Tumor	7
1.2.3	Treatment Planning of Brain Tumor	8
1.2.4	Challenges for Brain Tumor Segmentation	9
<b>1.3</b>	<b>Conclusion</b>	<b>10</b>

## 1.1 Introduction of Magnetic Resonance Imaging

### 1.1.1 Principle of Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a medical imaging technique used in radiology to produce three dimensional detailed anatomical images of the body. The first MRI scanner used to image the human body was built in New York in 1977. Since then, the technology has come a long way and now MRI is widely used in hospitals and clinics for medical diagnosis and treatment monitoring. The MRI scanner is essentially a giant magnet. The strength of the magnet is measured in a unit called Tesla (T). Most MRI scanners used in hospitals and medical research clinics are 1.5 or 3T. A 3T MRI scanner is around 50,000 times stronger than the earth's magnetic field which is around 0.00006T [1]. Figure 1.1 presents the schematic diagram of MRI scanner. To obtain an MRI image, a patient is placed inside a large magnet. Contrast agents (often containing the element Gadolinium) may be given to a patient intravenously before or during the MRI to increase the speed at

which protons realign with the magnetic field. The faster the protons realign, the brighter the image. During scanning, the patient must keep still to avoid blurring the image [2].

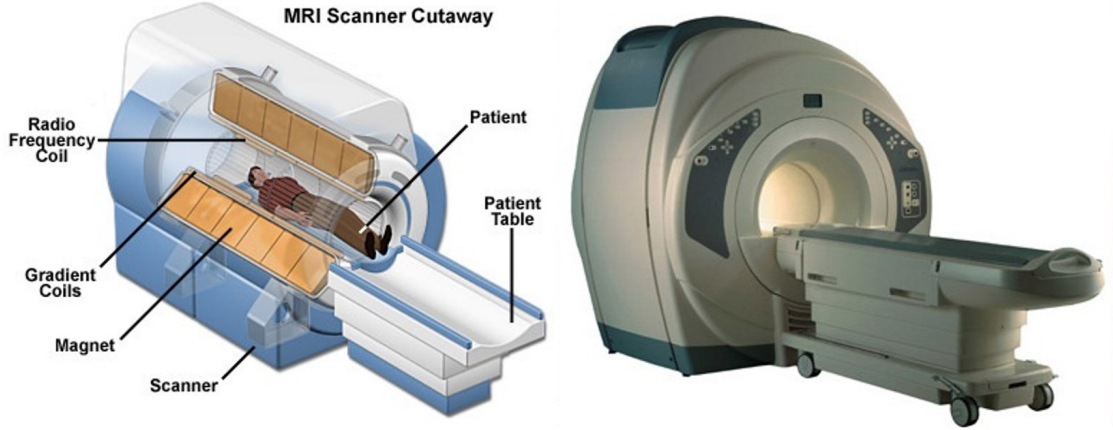


Figure 1.1 – The schematic diagram of MRI scanner [3].

The principle of MRI is described as follows. The human body contains a lot of water and hydrocarbons, which contain many hydrogen atoms. The hydrogen nuclei (protons), called nuclear spin. They act like a tiny magnet. Figure 1.2 details the process of obtaining signals from protons in the body. When outside of an external magnetic field, protons will point randomly in any direction, thus yielding a null magnetization moment. When it is placed in a strong magnetic field,  $B_0$ , protons exhibit a weak tendency to align with the direction of  $B_0$  (Figure 1.2 (a)). When the Radio Frequency (RF) pulse is applied, (Figure 1.2 (b)),  $B_1$ , perpendicular to  $B_0$ , the protons become deflected by  $90^\circ$ . When the RF signal stops, the deflected protons return to equilibrium such that it is parallel again to  $B_0$  (Figure 1.2 (c)). The recovery process is called the relaxation process. It emits their stored energy which can be received by a coil as a signal, and an image is formed by analyzing this signal and the time duration of the return [5].

### 1.1.2 MRI Sequences

MRI is a multimodal imaging approach. It can express various contrasts depending on the parameters of the excitation used. This contrast is due to the various relaxation properties of the protons in the tissues. The major image contrasts are T1- and T2-weighted images, which reflect two different relaxation times (T1 longitudinal relaxation time and T2 transverse relaxation time). Another commonly used sequence is the FLAIR. The FLAIR sequence is similar to a T2-weighted image except that the Time to Echo (TE) and Repetition Time (TR) are very long. TR is the amount of time between successive pulse sequences applied to the same slice. TE is the time between the delivery of the RF pulse and the receipt of the echo signal. Figure 1.3 presents the examples of FLAIR, T1-weighted (T1), contrast-enhanced T1-weighted (T1c) and T2 weighted (T2) images from the same patient. Table 1.1 presents the gray distribution of different tissues in

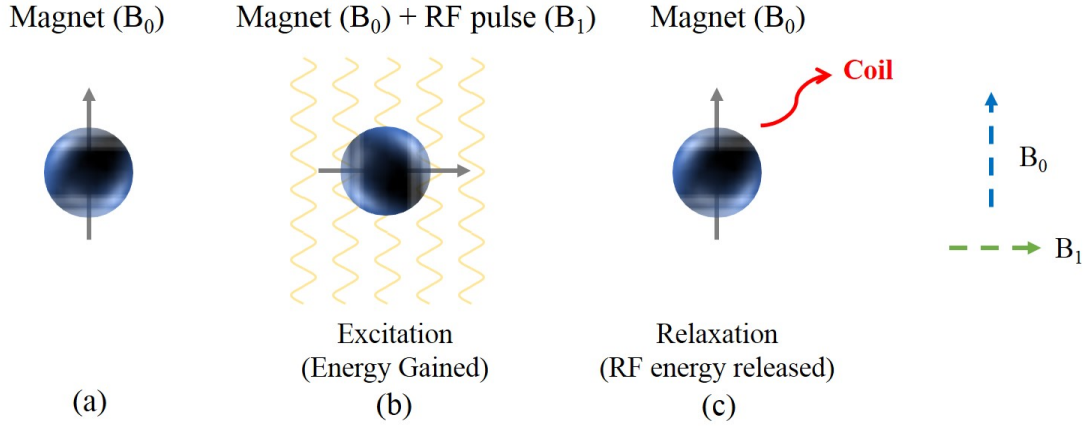


Figure 1.2 – Conceptual diagram of obtaining signals from protons in the body. Protons are generally oriented in random directions within the body. (a) When entering a strong magnetic field  $B_0$ , they change to being aligned within the magnetic field  $B_0$ . (b) If an RF pulse ( $B_1$ ) is applied, the protons become deflected by  $90^\circ$ . The deflected protons then continue to store the energy of the RF pulse. (c) If the RF pulse stops, the deflected protons return to their previous orientation while emitting their stored energy which can be received by a coil placed around the object. The figure is inspired by [4].

these three MRI sequences. In general, T1 and T2 images can be easily distinguished by comparing the Cerebro Spinal Fluid (CSF). CSF is dark on T1 image while bright on T2-weighted image. For FLAIR images, the abnormalities remain bright but normal CSF fluid is attenuated and made dark. FLAIR is very sensitive to pathology and makes the differentiation between CSF and an abnormality much easier [6].

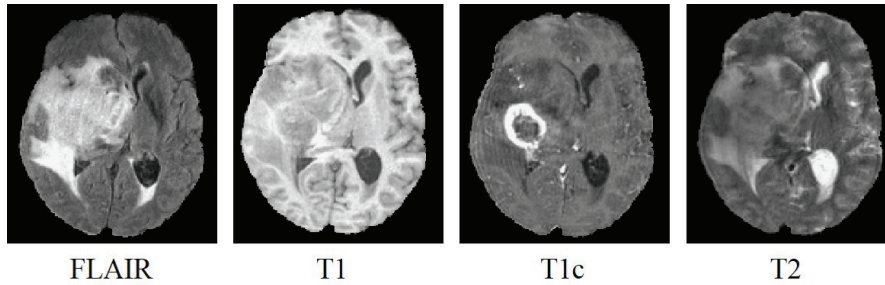



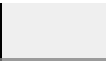










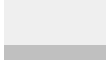





Figure 1.3 – Examples of FLAIR, T1, T1c and T2 images from the same patient [7].

### 1.1.3 Variations of MRI

There are a variety of MRIs available, the most common MRIs include [9]:



Table 1.1 – The gray distribution of different tissues in multiple MRI sequences [8]

Tissue	T1	T2	Flair
CSF			
Gray Matter			
White Matter			
Edema			
Fat			
Cartilage			

- Functional MRI (fMRI): fMRI can measure the changes in cerebral blood flow and the oxygenation of brain cells that is correlated to brain activity. It can produce an activation map showing which parts of the brain are involved in a particular mental process. It can be used to examine the functional anatomy of brain, evaluate the effects of diseases, and also guide the brain treatment planning.
- Perfusion MRI: It is perfusion scanning by using a particular MRI sequence. The acquired data are then post-processed to obtain perfusion maps with different parameters, such as Blood Volume (BV), Blood Flow (BF), Mean Transit Time (MTT) and Time To Peak (TTP) [10].
- Magnetic Resonance Venography (MRV): It is combined with contrast dye to produce clear images of internal organs and other structures inside the body. The dye highlights the veins, so that they appear translucent and show up well in images.
- Magnetic Resonance Angiography (MRA): It is similar to an MRV. It combines images with an intravenous contrast dye, but focuses on blood vessels instead of veins. The physician will be able to evaluate the blood vessels that run through the heart and soft tissues in the body.

#### 1.1.4 Role of MRI in Medical Diagnosis

MRI has a wide range of applications in medical diagnosis. In neuroimaging, MRI is the effective imaging tool, since it provides better visualization of the posterior cranial fossa, containing the brainstem and the cerebellum. In cardiovascular, cardiac MRI can be used to assess the structure and the function of the heart [11]. Also, MRI can be used to diagnose imaging of systemic muscle diseases [12]. In generally, MRI is a safe imaging technique, although injuries may occur as a result of failed safety procedures or human error [13].

The benefits of MRI can be summarized as follows [14]:

- Non-invasive: MRI does not use radiation and it is considered as a non-invasive procedure.
- No Ionizing Radiation: RF pulses used in MRI do not cause ionization and have no harmful effects of ionizing radiation.
- Contrast resolution: It can manipulate the contrast between different tissues by altering the pattern of RF pulses.
- Multiplanar image: We can obtain axial, sagittal and coronal images directly with MRI, which is impossible with radiography and Computed Tomography (CT).

There are some limitations of MRI, which are summarized as follows:

- Motion artifact: It is one of the most common artifacts in MR imaging. Motion can cause either ghost images or diffuse image noise in the phase-encoding direction.
- Low contrast: The low contrast of MRI results in the fuzzy tumor boundary, making the following segmentation challenging.

## 1.2 Brain Tumor Segmentation

### 1.2.1 Brain Tumor Overview

Tumor is an uncontrolled growth of cancer cells in any part of the body. Brain tumor is one of the most aggressive cancers in the world [15, 16]. World Health Organization (WHO) assigned brain tumors grades, ranging from Grade I (least malignant) to Grade IV (most malignant), which signifies the rate of growth. In general, grade I and grade II are benign brain tumor (low-grade); grade III and grade IV are malignant brain tumor (high-grade). Brain tumors can be classified as primary tumors (see Figure 1.4) and secondary tumors (brain metastasis tumors) (see Figure 1.5). The former one starts within the brain, such as meninges, brain cells, nerve cells, glands. The latter one begins as a cancer elsewhere and spreads to brain. For example, lung cancer, skin cancer, breast cancer, and kidney cancer can metastasize to the brain. Most of the brain tumors are secondary forms of tumors. Survival rate is one indicator commonly used to reflect of the severity of the disease and the effectiveness of the treatment process. The 5-year survival rate tells what percent of people live at least 5 years after the tumor is found. The 5-year survival rate for people with a cancerous brain or Central Nervous System (CNS) tumor is 36%. The 10-year survival rate is about 31%. Survival rates decrease with age. The 5-year survival rate for people younger than age 15 is more than 75%. For people age 15 to 39, the 5-year survival rate is more than 72%. The 5-year survival rate for people age 40 and over is more than 21%. However, survival rates vary widely and depend on several factors, including the type of brain or spinal cord tumor [17].

Meningiomas and gliomas are the two most common types of primary tumors that occur in the brain or spinal cord. Other common primary brain tumors are pituitary adenomas, ependymomas, schwannomas and craniopharyngiomas. Meningiomas originate

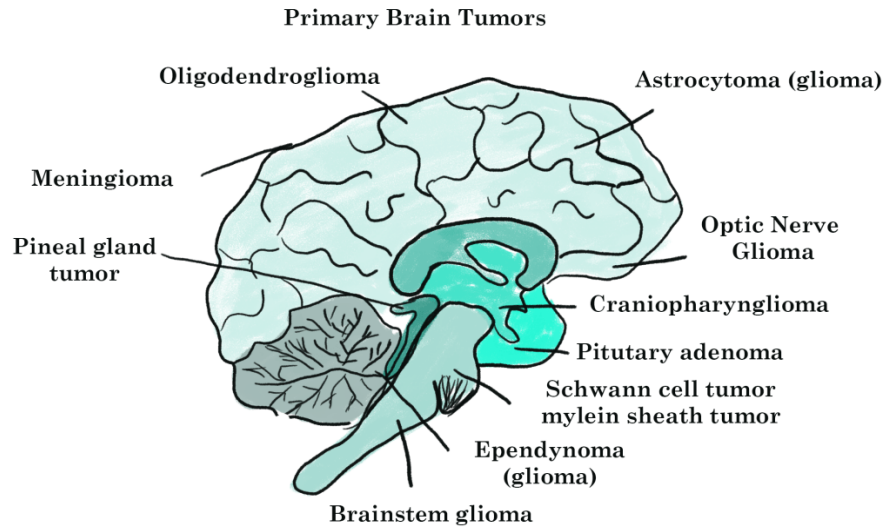


Figure 1.4 – Schematic diagram of primary brain tumors [18].

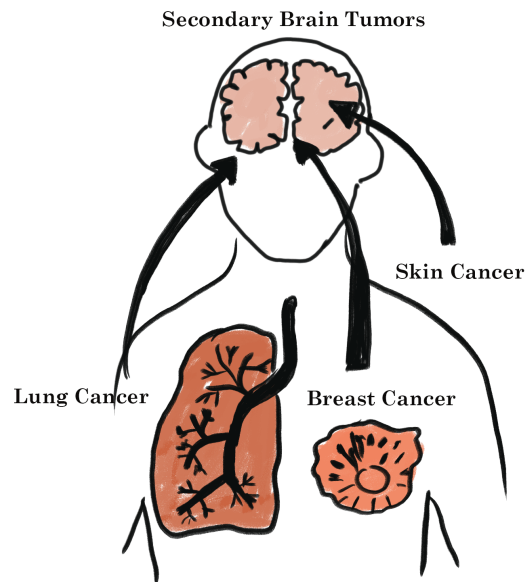


Figure 1.5 – Schematic diagram of secondary brain tumors [18].

from the meninges, membranes that cover the brain and spinal cord. Most meningiomas are benign, while their growth may affect the brain by causing various disabilities such as vision and hearing impairment, memory loss, or even seizures. Gliomas begin in glial cells which make up the supportive tissue of the brain. These include glioblastomas, astrocytomas, oligodendrogliomas, and ependymomas. Astrocytoma is the most common type of glioma tumor. It originates in star-shaped cells called astrocytes, which are located in the brain's cerebrum. Sometimes they grow slowly (Grade II) and sometimes they come on more aggressively (Grade III) [19].

Factors that can cause brain tumor can be summarized as follows [18, 20]:

- Age: Brain tumors are more common in children and older adults, the risk of getting brain cancer increases with age.
- Family history: Brain cancer is usually not genetically inherited. Only 5-10% of cancers are due to genetics.
- Head injury and seizures: Serious head trauma has long been studied for its relationship to brain tumors.
- Exposure to chemicals: Exposure to these cancer-causing chemicals increases the risk of brain cancer.
- Ionizing radiation: Previous treatment to the brain or head with ionizing radiation, including x-rays, can be a risk factor for a brain tumor.

The appearance of symptoms depends on the size and the location of the tumor. Some tumors directly destroy the brain tissue, and some indirectly put pressure on the surrounding area of the brain. Symptoms can include headache, memory loss, difficulty in writing and reading, changes in perception of senses, uncontrolled movement, dizziness or vertigo, difficulty in walking, weakness in arms, legs, and face, tremors and so on.

### 1.2.2 Diagnosis of Brain Tumor

In general, diagnosing a brain tumor usually begins with MRI. Once MRI shows that there is a tumor in the brain, the most common way to determine the type of brain tumor is to look at the results from a sample of tissue after a biopsy or surgery. The commonly used diagnosis procedures are described below [18, 21]:

- MRI: MRI uses magnetic fields to produce detailed images of the body. The MRI may be of the brain, spinal cord, or both, depending on the type of tumor suspected and the likelihood that it will spread in the CNS. The types of MRI to use will depend on the results of a neuro-examination.
- Positron Emission Tomography (PET). A PET scan is a way to create pictures of organs and tissues inside the body using various substances, such as sugars or proteins. A PET scan is usually combined with a CT scan, called a PET-CT scan. A small amount of a radioactive substance is injected into the patient's body. This

substance is taken up by cells that are actively dividing. Because tumor cells are more likely to be actively dividing, they absorb more of the radioactive substance. A scanner then detects this substance to produce images of the inside of the body.

- **Angiography:** A dye is injected into the artery, generally in the groin region, which reaches the arteries of the brain. It enables the doctors to see the blood supply in the brain. This information is then used at the time of surgery.
- **Biopsy:** A biopsy is the removal of a small amount of tissue for examination under a microscope. A neuropathologist then analyzes the sample. The biopsy helps to identify whether the brain tumors are malignant or benign, it also helps in determining the origin of cancer.
- **X-ray:** The pressure put by tumors on the skull can cause breaks or fractures in the bones, which can be identified with the help of a specific type of X-rays. X-rays can also show if some calcium deposits are present in a tumor. These deposits can be present in your blood vessels if the tumor has spread to the bones.

### 1.2.3 Treatment Planning of Brain Tumor

In brain tumor treatment, a multidisciplinary doctor team often works together to create a patient's overall treatment plan that combines different types of treatment. The team may include a variety of other health care professionals, such as physician assistants, nurse practitioners, oncology nurses, social workers, pharmacists, counselors, dietitians, rehabilitation specialists, and others. Treatment options include surgery, radiation therapy, chemotherapy, and targeted therapy [22].

Surgery is to remove the tumor and some surrounding healthy tissue during an operation. It is usually the first treatment used for brain tumor. Radiation therapy is to use high-energy x-rays or other particles to destroy tumor cells. Doctors may use radiation therapy to slow or stop the growth of a brain tumor. It is typically given after surgery and possibly along with chemotherapy. Accurate tumor delineation is particularly important in radiation therapy to avoid under-treatment of tumor or overtreatment of surrounding normal tissues. The imaging modalities, such as MRI and PET can provide additional information to more precisely delineate tumors during radiation therapy [23, 24]. Chemotherapy is using drugs to destroy tumor cells, usually keeping the tumor cells from growing, dividing, and making more cells. In addition to standard chemotherapy, targeted therapy is a treatment that targets the tumor's specific genes, proteins, or the tissue environment that contributes to a tumor's growth and survival. This type of treatment blocks the growth and spread of tumor cells and limits the damage to healthy cells. For a low-grade brain tumor, surgery may be the only treatment needed especially if all of the tumor can be removed. If there is visible tumor remaining after surgery, radiation therapy and chemotherapy may be used. For higher-grade tumors, treatment usually begins with surgery, followed by radiation therapy and chemotherapy.

#### 1.2.4 Challenges for Brain Tumor Segmentation

Brain tumor segmentation takes an important role in tumor diagnosis and treatment planning in clinical routine [25]. Manually delineating tumors is the most intuitive and common way in clinical practice. Since manual segmentation of brain tumors is a highly time-consuming, expensive and subjective task. Developing automatic segmentation methods is necessary for tumor segmentation. However, there are some challenges for brain tumor segmentation.

First, brain tumors have properties making the accurate segmentation challenging [26]:

- The brain anatomy structure varies from patients to patients.
- The brain tumor can appear at variable locations in almost any size and shape.
- The brain tumor are very heterogeneous, the intensity value of brain tumor may overlap with the intensity value of the healthy brain tissue.

Second, the segmentation results strongly rely on the imaging technology.

- The brain tumor boundary is often unclear due to the low contrast of MRI.
- It's challenging to fuse different MRI sequences to utilize the complementary information to improve the segmentation performance.
- It's common to have some missing MRI sequences in clinical practice, which can heavily decrease the segmentation precision compared with the complete sequences.

Third, a well-processed brain tumor image dataset plays an import role in brain tumor segmentation [27].

- Image quality has a critical impact on segmentation performance, while hospitals often have different scanners and image protocols, which complicates the standardization and data quality.
- Image pre-processing steps have an import impact on the segmentation performance. For example, image registration and intensity normalization across cases are critical for brain tumor segmentation.
- Data imbalance is common and poses another intricate challenge for brain tumor segmentation in deep learning. For example, the normal brain region is larger than the abnormal regions, and the edema region is generally larger than other regions. Training with the imbalanced data can cause an unstable segmentation network.

### 1.3 Conclusion

In this chapter, we presented the medical background about brain tumor segmentation in MRI, consisting of principle of MRI, the application of MRI in medical diagnosis. We also introduce the knowledge about brain tumor, the corresponding diagnosis and treatment process and the challenges of brain tumor segmentation. In the next chapter, we will present the state-of-the-art methods.

## Chapter 2

# A Review: Deep Learning for Medical Image Segmentation using Multi-modality Fusion

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>12</b>
<b>2.2</b>	<b>Related Works</b>	<b>15</b>
2.2.1	Brief Introduction of Deep learning	15
2.2.2	Multi-modal Medical Image Segmentation	16
<b>2.3</b>	<b>Data Preparation</b>	<b>18</b>
2.3.1	Data Dimension	18
2.3.2	Pre-processing	19
2.3.3	Data Augmentation	19
2.3.4	Post-processing	19
<b>2.4</b>	<b>Multi-modal Medical Image Segmentation Networks</b>	<b>19</b>
2.4.1	Input-level Fusion Network	20
2.4.2	Layer-level Fusion Network	21
2.4.3	Decision-level Fusion Network	23
<b>2.5</b>	<b>Discussion and Methodologies to be Developed</b>	<b>23</b>
<b>2.6</b>	<b>Conclusion</b>	<b>26</b>

---

---

---

Tongxue Zhou, Su Ruan, and Stéphane Canu. “A review: Deep learning for medical image segmentation using multi-modality fusion”. In: *Array* (2019), p. 100004

---

---



## 2.1 Introduction

Segmentation using multi-modality has been widely studied with the development of medical image acquisition systems. Different strategies for image fusion, such as probability theory [28, 29, 30], fuzzy concept [31, 32, 33, 34, 35], belief functions [36, 37, 38, 39], and machine learning [40, 41, 42, 43, 44] have been developed with success. For the methods based on the probability theory and machine learning, different data modalities have different statistical properties which makes it difficult to model them using shallow models. For the methods based on the fuzzy concept, the fuzzy measure quantifies the degree of membership relative to a decision for each source. The fusion of several sources is achieved by applying the fuzzy operators to the fuzzy sets. For the methods based on the belief function theory, each source is first modeled by an evidential mass, the DempsterShafer rule is then applied to fuse all sources. The main difficulty to use the belief function theory and the fuzzy set theory relates to the choice of the evidential mass, the fuzzy measure and the fuzzy conjunction function. However, a deep learning based network can directly encode the mapping. Therefore, the deep learning based method has a great potential to produce better fusion results than conventional methods. Some new works on deep learning combined with belief function have been developed [45]. Since 2012, several deep convolutional neural network models have been proposed such as AlexNet [46], ZFNet [47], VGG [48], GoogleNet [49], Residual Net [50], DenseNet [51], FCN [52] and U-Net [53]. These models have not only provided state-of-the-art performance for image classification, segmentation, object detection and tracking tasks, but also provide a new point of view for image fusion. There are mainly four reasons contributing to their success: Firstly, the main reason behind the amazing success of deep learning over traditional machine learning models is the advancements in neural networks. It learns high-level features from data in an incremental manner, which eliminates the need of domain expertise and hard feature extraction. And it solves the problem in an end to end manner. Secondly, the appearance of Graphics Processing Unit (GPU) and GPU-computing libraries make the model can be trained 10 to 30 times faster than on Central Processing Unit (CPU). And the open source software packages provide efficient GPU implementations. Thirdly, publicly available datasets such as ImageNet [54], can be used for training, which allow researchers to train and test new variants of deep learning models. Finally, several available efficient optimization techniques also contribute to the final success of deep learning, such as dropout, batch normalization, Adam optimizer and others, Rectified Linear Unit (ReLU) activation function and its variants, with that, we can update the weights and obtain the optimal performance.

Medical image segmentation is an important field in medical image analysis and is a necessary step for diagnosis, monitoring and treatment. The goal of segmentation is to assign the label to each pixel in an image. It generally includes two phases, firstly, detect the unhealthy tissue or areas of interest; secondly, delineate the different anatomical structures or areas of interest. Motivated by the success of deep learning, researches in medical image field have also attempted to apply deep learning based approaches to medical image segmentation in the brain [55, 56, 57, 58, 59], lung [60, 61, 62], pancreas [63, 64, 65, 66], prostate [67, 68, 69] and multi-organ [70, 71, 72, 73]. In order to obtain

more accurate segmentation for better diagnosis, using multi-modal medical images has been a growing trend strategy. A thorough analysis of the literature with the keywords ‘deep learning’, ‘medical image segmentation’ and ‘multi-modality’ on Google Scholar is performed in Figure 2.1, which was queried on July 01, 2021. We can observe that the number of papers increases every year from 2014 to 2020, which means multi-modal medical image segmentation in deep learning are obtaining more and more attention in recent years. To have a better understanding of the dimension of this research field, we compare the scientific production of the image segmentation community, the medical image segmentation community, and the medical image segmentation using multi-modality fusion with and without deep learning in Figure 2.2. We can observe that the amount of papers has a descent or even tendency in the methods without deep learning, while there is an increase number of papers using deep learning method in each research field.

The principal modalities in medical images analysis are CT, PET and MRI, which are presented in Figure 2.3. As pointed out in [74], the CT image can diagnose muscle and bone disorders, such as bone tumors and fractures, while the MR image can offer a good soft tissue contrast without radiation. Functional images, such as PET, lack anatomical characterization, while can provide quantitative metabolic and functional information about diseases. MRI modality can provide complementary information due to its dependence on variable acquisition parameters, such as T1-weighted (T1), contrast-enhanced T1-weighted (T1c), T2-weighted (T2) and FLAIR images. Compared to single images, multi-modal images help to extract features from different views and bring complementary information, contributing to better data representation and discriminative power of the network. For example, T2 and FLAIR are suitable to detect the tumor with peritumoral edema, while T1 and T1c to detect the tumor core without peritumoral edema. Therefore, applying multi-modal images can reduce the information uncertainty and improve clinical diagnosis and segmentation accuracy [75].

The multi-modal fusion methods can be categorized into earlier fusion and later fusion. The earlier fusion is simple and most works use the fusion strategy to do the segmentation. It focuses on the subsequent complex segmentation network architecture designs without considering the relationship between different modalities. However, the later fusion pays more attention on the fusion problem, since each modality is employed as an input of one network which can learn complex and complementary feature information of each modality. In general, compared to the earlier fusion, the later fusion can achieve better segmentation performance if the fusion method is effective enough. And the selection of fusion method depends on the specific problem.

The rest of the chapter is structured as follows. In Section 2.2, we introduce the general principle of deep learning and multi-modal medical image segmentation. In Section 2.3, we present how to prepare the data before feeding to the network. In Section 2.4, we describe the detailed multi-modal medical image segmentation network based on different fusion strategies. In Section 2.5, we give a discussion about the chapter and the methodologies to be developed. In Section 2.6, we summarize the chapter.

## CHAPTER 2. A REVIEW: DEEP LEARNING FOR MEDICAL IMAGE SEGMENTATION USING MULTI-MODALITY FUSION

---

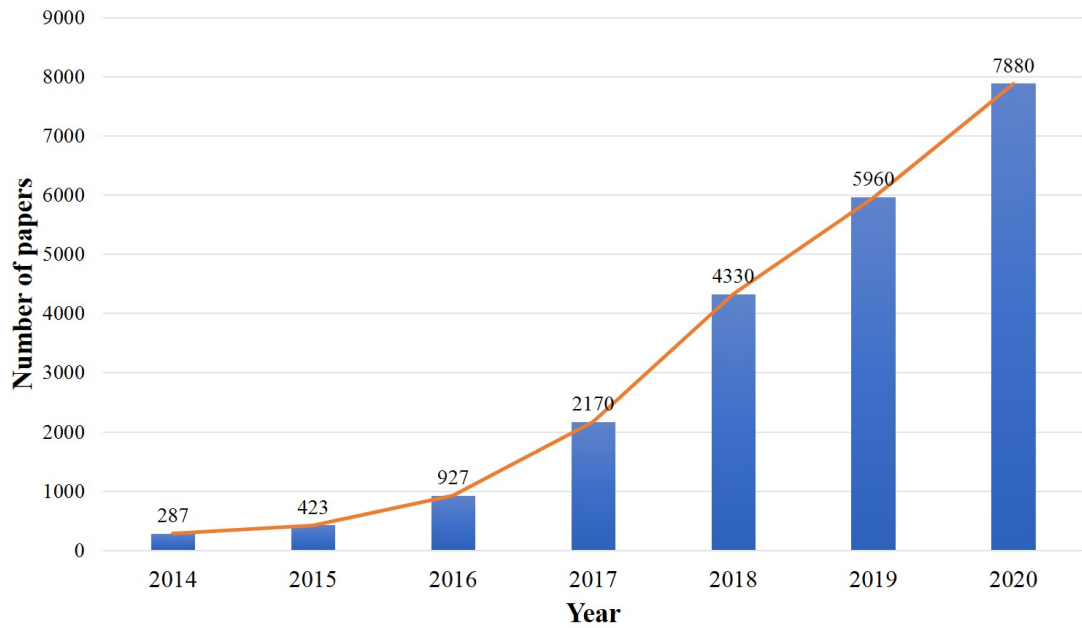


Figure 2.1 – The tendency of multi-modal medical image segmentation in deep learning from 2014 to 2020.

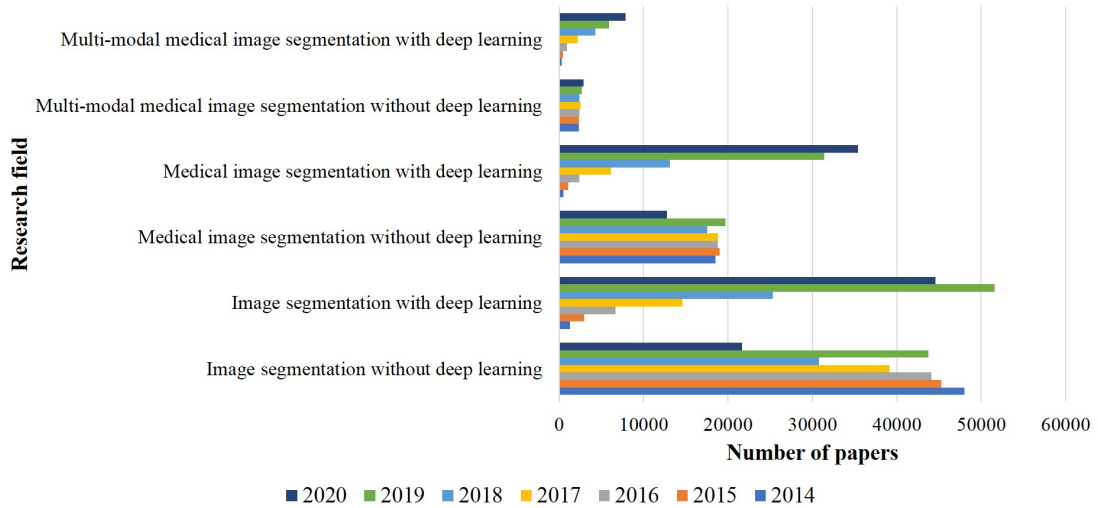


Figure 2.2 – The tendency of relative research field with/without deep learning.

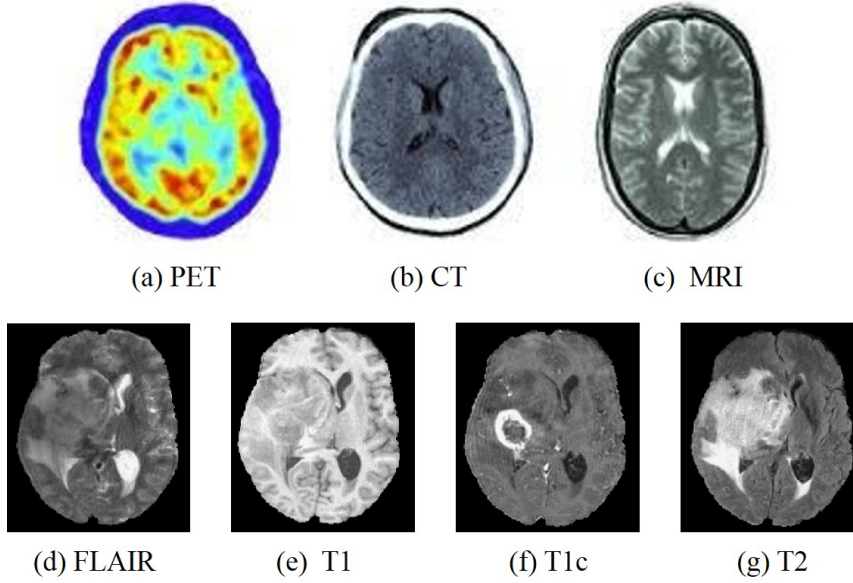


Figure 2.3 – The multi-modal medical images, (a)-(c) are the commonly used multi-modal medical images [76] and (d)-(g) are the different sequences of brain MRI [7].

## 2.2 Related Works

### 2.2.1 Brief Introduction of Deep learning

Deep learning refers to a neural network with multiple layers of nonlinear processing units [77]. Each successive layer uses the output of the previous layer as input. The network can extract the complex hierarchy features from a large amount of data using these layers. In recent years, deep learning has made significant improvements in image classification, recognition, object detection and medical image analysis, where they have produced excellent results comparable to human experts. Among the known deep learning algorithms, such as stacked auto-encoders [78], deep Boltzmann machines [79], and convolutional neural networks [80], the most successful one for image segmentation is Convolutional Neural Network (CNN). It was first proposed in 1989 by LeCun and the first successful real-world application [81] is the hand-written digit recognition in 1998 by LeCun, where he presented a five-layer fully-adaptive architecture. Due to its accuracy results (1% error rate and 9% reject rate from a dataset of 2007 handwritten characters), the neural networks can be applied into a real-world problem. However, it did not gather much attention until the contribution of Krizhevsky et al. to the ImageNet challenge in 2012. The proposed AlexNet [46], similar to LeNet but deeper, outperformed all the competitors and won the challenge by reducing the top-5 error (The percentage of times that the classifier did not involve the correct class among the top 5 predicted classes.) from 26% to 15.3%. In the subsequent years, other based on CNN architectures are

## CHAPTER 2. A REVIEW: DEEP LEARNING FOR MEDICAL IMAGE SEGMENTATION USING MULTI-MODALITY FUSION

Architecture	Author	Rank on ILSVRC	Top-5 error rate	Number of parameters
LeNet[81]	LeCun et al. 1998	N/A	N/A	60 thousand
AlexNet[46]	Krizhevsky et al. 2012	1st	16.4%	60 million
ZFNet[47]	Zeiler et al. 2013	1st	11.7%	N/A
VGG Net[48]	Simonyan et al. 2014	2nd	7.3%	138 million
GoogleNet[49]	Szegedy et al. 2015	1st	6.7%	5 million (V1) & 23 million (V2)
ResNet[50]	He. Kaiming et al. 2016	1st	3.57%	25.6 million (ResNet-50)
DenseNet[51]	Huang et al. 2017	N/A	N/A	6.98 million (DenseNet-100, k=12)

Table 2.1 – Summary of deep learning network architectures, ILSVRC: ImageNet Large Scale Visual Recognition Challenge.

proposed, including VGGNet [48], GoogleNet [49], Residual Net [50] and DenseNet [51]. Table 2.1 describes the details of these network architectures.

CNN is a multi-layer neural network containing convolution, pooling, activation and fully connected layers. Convolution layers are the core of CNN and are used for feature extraction. The convolution operation can produce different feature maps depending on the filters used. Pooling layer performs a downsampling operation by using maximum or average of the defined neighbourhood as the value to reduce the spatial size of each feature map. Non-linear rectified layer (ReLU) and its modifications such as Leaky ReLU are among the most commonly used activation functions [82], which transforms data by clipping any negative input values to zero or having a small slope for negative values, while positive input values are passed as output. Neurons in a fully connected layer are fully connected to all activations in the previous layer. They are placed before the classification output of a CNN and are used to flatten the results before a prediction is made using linear classifiers. While training the CNN architecture, the model predicts the class scores for training images, computes the loss using the selected loss function and finally updates the weights using the gradient descent method by back-propagation. The cross-entropy loss is one of the most widely used loss functions and Stochastic Gradient Descent (SGD) is the most popular method to operate gradient descent.

### 2.2.2 Multi-modal Medical Image Segmentation

Due to the variable size, shape and location of target tissue, medical image segmentation is one of the most challenging tasks in the field of medical image analysis. Despite the variety of proposed segmentation network architectures, it is still hard to compare the performance of different algorithms, because most of the algorithms are evaluated on different sets of data and reported in different metrics. In order to obtain accurate segmentation and compare different state-of-the-art methods, some well-known publicly challenges for segmentation are created, such as Brain Tumour Segmentation (BraTS) [55], Ischemic Stroke Lesion Segmentation(ISLES)<sup>1</sup>, MR Brain Image Segmentation (MRBrainS) [83], Neonatal Brain Segmentation (NeoBrainS) [84], Combined Computed Tomography - Magnetic Resonance Imaging (CT-MRI) Healthy Abdominal Organ Seg-

<sup>1</sup><http://www.isles-challenge.org>

## CHAPTER 2. A REVIEW: DEEP LEARNING FOR MEDICAL IMAGE SEGMENTATION USING MULTI-MODALITY FUSION

Table 2.2 – Summary of the multi-modal medical segmentation datasets.

Dataset	Train	Validation	Test	Segmentation Task	Modality	Image Size
BraTS 2012	35	N/A	15	Brain tumor	T1, T1c, T2, FLAIR	$160 \times 216 \times 176$ $176 \times 176 \times 216$
BraTS 2013	35	N/A	25	Brain tumor	T1, T1c, T2, FLAIR	$160 \times 216 \times 176$ $176 \times 176 \times 216$
BraTS 2014	200	N/A	38	Brain tumor	T1, T1c, T2, FLAIR	$160 \times 216 \times 176$ $176 \times 176 \times 216$
BraTS 2015	200	N/A	53	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
BraTS 2016	200	N/A	191	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
BraTS 2017	285	46	146	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
BraTS 2018	285	66	191	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
BraTS 2019	335	125	166	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
BraTS 2020	369	125	166	Brain tumor	T1, T1c, T2, FLAIR	$240 \times 240 \times 155$
ISLES2015	28	N/A	36	Ischemic stroke lesion	T1, T2, TSE, FLAIR, DWI, TFE/TSE	$230 \times 230 \times 154$
	30	N/A	20		T1c, T2, DWI, CBF, CBV, TTP, Tmax	N/A
MRBrainS13	5	N/A	15	Brain Tissue	T1, T1_1mm, T1_IR, FLAIR	$256 \times 256 \times 192$ $240 \times 240 \times 48$
NeoBrainS12	20	N/A	5	Brain Tissue	T1, T2	$384 \times 384 \times 50$ $512 \times 512 \times 110$ $512 \times 512 \times 50$
iSeg-2017	10	N/A	13	Brain Tissue	T1, T2	N/A
CHAOS	20	N/A	20	Abdominal Organs	CT, T1-DUAL, T2-SPIR	N/A
IVD	16	N/A	8	Intervertebral Disc	In-phase, Opposed-phase, Fat, Water	N/A

mentation (CHAOS)<sup>2</sup>, 6-month Infant Brain MRI Segmentation (Iseg-2017) [85] and Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality MR (M3) Images (IVDM3Seg)<sup>3</sup>. Table 2.2 describes the detailed datasets information mentioned above. Table 2.3 shows the main evaluation metrics in these datasets.

We describe a pipeline of multi-modal medical image segmentation based on deep learning, shown in Figure 2.4. The pipeline consists of four parts: data preparation, network architecture, fusion strategy and data post-processing. In the data preparation stage, the data dimension is firstly chosen, and the pre-processing is used to reduce the variation between images, and data augmentation strategy can also be used to increase the training data to avoid the over-fitting problem. In the network architecture and fusion strategy stages, the basic network and detailed multi-modal images fusion strategies are presented to train the segmentation network. In the data post-processing stage, some post-processing techniques such as morphological techniques and conditional random field are implanted to refine the final segmentation result. In the task of multi-modal medical image segmentation, fusing multiple modalities is the key problem of the task. According to the level in the network architecture where the fusion is performed, the fusion strategies can be categorized into three groups: input-level fusion, layer-level fusion, and decision-level fusion, the details refers to Section 2.4.

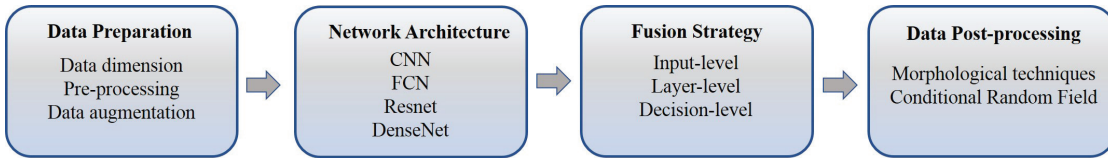


Figure 2.4 – The pipeline of multi-modal medical image segmentation based on deep learning.

<sup>2</sup><https://chaos.grand-challenge.org>

<sup>3</sup><https://ivdm3seg.weebly.com>

CHAPTER 2. A REVIEW: DEEP LEARNING FOR MEDICAL IMAGE  
SEGMENTATION USING MULTI-MODALITY FUSION

Evaluation metric	Mathematical description
Dice Similarity Coefficient (DSC)	$DSC = \frac{2TP}{2TP+FP+FN}$
Sensitivity	$Sensitivity = \frac{TP}{TP+FN}$
Specificity	$Specificity = \frac{TN}{TN+FP}$
Precision	$Precision = \frac{TP}{TP+FP}$
Hausdorff Distance (HD)	$HD = \max\{\max_{s \in S} \min_{r \in R} d(s, r), \max_{r \in R} \min_{s \in S} d(r, s)\}$
Absolute Relative Volume Difference (ARVD)	$ARVD(X, Y) = \left  100 \times \left( \frac{ X }{ Y } - 1 \right) \right $

Table 2.3 – Summary of the commonly used evaluation metrics. FP, TP, FN and TN are the number of false positive voxels, true positive voxels, false negative voxels and true negative voxels, respectively.  $d$  is the Euclidean distance,  $S$  and  $R$  are the two sets of surface points of the prediction and the real annotation.  $|X|$  and  $|Y|$  are the number of voxels of the prediction and real annotation, respectively.

## 2.3 Data Preparation

This section will describe the data processing including data dimension selection, image pre-processing, data augmentation and post-processing techniques. This step is important in deep learning based segmentation network.

### 2.3.1 Data Dimension

Medical image segmentation usually deals with 3D images. Some models directly use the 3D images to train models [86, 87, 88, 89], while some models process the 3D image slice by slice [56, 90, 91, 92, 93]. The 3D approach takes the 3D image as input and applies the 3D convolution kernel to exploit the spatial contextual information of the image. The main drawback is its expensive computational cost. Compared to utilizing the whole volume image to train the model, some 3D small patches can be used to reduce the computational cost. For instance, Kamnitsas et al. [94] extracts 10k random 3D patches at regular intervals to train a brain tumor segmentation network. The 2D approach takes the image slice or patch extracted from the 3D image as input and applies the 2D convolutional kernel. The 2D approach can efficiently reduce the computational cost, while it ignores the spatial information of the image in z direction. For example, Zhao et al. [95] first trained a Fully Convolutional Network (FCN) using image patches and then a Conditional Random Field (CRF) as Recurrent Neural Network (RNN) using image slices. Finally, they fine-tuned the two networks using image slices. To exploit the feature information of the 2D image and 3D image, Mlynarski et al. [96] described a CNN-based model for brain tumor segmentation. It first extracts the 2D features of the image from axial, coronal and sagittal views and then takes them as the additional input of the 3D CNN-based model. The method can learn rich feature information in three dimensions, which achieve good performance with median Dice scores of 0.918 (whole tumor), 0.883 (tumor core) and 0.854 (enhancing core).

### 2.3.2 Pre-processing

Pre-processing plays an important role in subsequent segmentation task, especially for the multi-modal medical image segmentation because there are variant intensity, contrast and noise in the images. Therefore, to make the images appear more similar and make the network training smooth and quantifiable, some pre-processing techniques are applied before feeding to the segmentation network. The typical pre-processing techniques consist of image registration, bias field correction and intensity normalization. For BraTS dataset, the image registration has already done before provided to the public. [56, 90, 94, 95, 97] used the N4ITK method to correct the distortion of MRI data. [57, 56, 86, 90, 93, 94] proposed to normalize each modality independently by subtracting the mean and dividing by the standard deviation of the brain region.

### 2.3.3 Data Augmentation

Most of the time, a large number of labels for training is not available for several reasons. Labelling the dataset requires an expert in this field which is expensive and time-consuming. Training a neural network from limited training data, the over-fitting problem needs to be considered. [98] Data augmentation is a way to reduce over-fitting and increase the amount of training data. It creates new images by transforming (rotated, translated, scaled, flipped, distorted and adding some noise such as Gaussian noise) the ones in training dataset. Both the original image and created images are fed into the neural network. For example, Isensee et al. [86] proposed to address over-fitting by utilizing a large variety of data augmentation techniques like random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring on the fly during training.

### 2.3.4 Post-processing

Post-processing is applied to refine the final result in segmentation network. The small isolated predicted regions are prone to artefacts and the largest volume are usually kept in the final segmentation. In this case, morphological techniques are preferred to remove incorrect small fragments and keep the largest volume. And some post-processing techniques can be designed according to the structure of detected region. For example, considering LGG patients may don't have enhancing tumor, Isensee et al. [87] proposed to replace all enhancing tumor voxels with necrosis if the number of predicted enhancing tumor is less than a threshold. In addition, a 3D fully connected Condition Random Field (CRF) [94] can be applied for post-processing to effectively remove false positives to refine the segmentation result.

## 2.4 Multi-modal Medical Image Segmentation Networks

Over the years, various semi-automated and automated techniques have been proposed for multi-modal medical image segmentation using deep learning based methods, such as CNN



[81] and FCN [52], especially U-Net [53]. According to the multi-modal fusion strategies, we category the network architectures into input-level fusion network, layer-level fusion network and decision-level fusion network. For each fusion strategy, we conclude some common used menthods shown in Figure 2.5 and the details are described in the following sections.

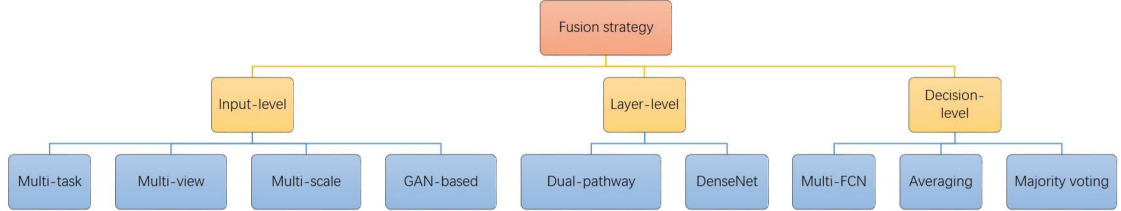


Figure 2.5 – The generic categorization of the fusion strategy.

### 2.4.1 Input-level Fusion Network

In the input-level fusion network, multi-modality images are fused channel-wise, and then fed to the segmentation network. Most of the existing multi-modal medical image segmentation networks adopt the input-level fusion strategy [56, 86, 87, 90, 93, 94, 95, 99, 100]. Figure 2.6 describes the generic network architecture of the input-level fusion segmentation network. We take CT and MRI as two input modalities, convolutional neural network as the segmentation network and the brain tumor segmentation as the segmentation task. This kind of fusion network usually adopts four architectures, multi-task segmentation, multi-view segmentation, multi-scale segmentation and Generative Adversarial Network (GAN)-based segmentation.

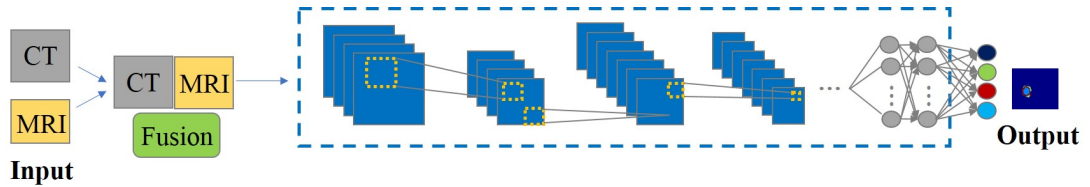


Figure 2.6 – The generic network architecture of the input-level fusion.

To name a few, Wang et al. [93] proposed a multi-modal brain tumor segmentation network. It uses multi-task and multi-view architectures. In order to obtain a united feature set, it directly integrates the four modalities (T1, T1c, T2 and FLAIR of MRI) as the multi-channel inputs in the input space. Then it separates the complex multi-class segmentation task into several simpler segmentation tasks according to the hierarchical structure of the brain tumor. The whole tumor is firstly segmented and then the bounding box including the whole tumor is used for the tumor core segmentation. Based on the obtained bounding box of the tumor core, the enhancing tumor core is finally segmented.

Furthermore, to take advantage of 3D contextual information, for each individual task, they fused the segmentation results from three different orthogonal views (axial, coronal and sagittal) by averaging the softmax outputs of the individual task. Experiments with the testing set of BraTS 2017 data show that the proposed method achieves an average Dice scores of 0.7831, 0.8739, and 0.7748 for enhancing tumor core, whole tumor and tumor core, respectively, which won the second place on BraTS 2017 challenge. The multi-task segmentation separates the complex task of multiple class segmentation into several simpler segmentation tasks and takes advantage of the hierarchical structure of tumour sub-regions to improve segmentation accuracy.

It's likely to require different receptive field when segmenting different regions in an image. Qin et al. [88] proposed the autofocus convolutional layer to enhance the abilities of neural networks by using multi-scale processing. After integrating the multi-modal images in the input space, they applied an autofocus convolutional layer by using multiple convolutional layers with different dilation rates to change the size of the receptive field. Autofocus convolutional layer can indicate the importance of each scale when processing different locations of an image. Also, they used an attention mechanism to choose the optimal scale. The proposed autofocus layer can be easily integrated into existing networks to improve a model's performance. The proposed method gained promising performance on the challenging tasks of multi-organ segmentation in pelvic CT and brain tumor segmentation in MRI.

Motivated by the success of GAN [101], which models a mini-max game between the generator and the discriminator, some methods propose to apply the discriminator as the extra constraint to improve the segmentation performance [102, 103]. In [102], by fusing the multi-modal images as multi-channel inputs, they trained two separate networks: a residual U-net as the generative network and a discriminator network. The segmentation network will generate a segmentation, while the discriminator network will distinguish between the generated segmentations and ground truth. The proposed method was evaluated on the BraTS 2018 dataset and achieved competitive results. Huo et al. [103] employed the PatchGAN [104] as an additional discriminator to supervise the training procedure of the network. The method based on GAN can obtain a robust segmentation due to the extra constraint of discriminator, while it costs more memory to train the extra discriminator.

The input-level fusion strategy can maximumly keep the original image information and learn the intrinsic image feature. Sequential segmentation networks allow to take different strategies, such as multi-task, multi-view, multi-scale and GAN-based segmentation network, to fully exploit the feature representation from multi-modal images.

#### 2.4.2 Layer-level Fusion Network

In the layer-level fusion network, each input image is used to train an individual network. Then these learned individual feature representations will be fused in the layers of the network. The fused result will be fed to the decision layer to obtain the final segmentation result. The layer-level fusion network can effectively integrate and leverage multi-modal

images [89, 91, 105, 106]. Figure 2.7 describes the generic network architecture of layer-level fusion work.

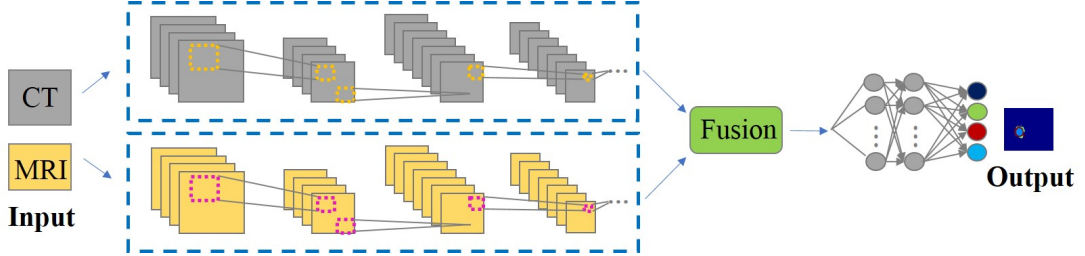


Figure 2.7 – The generic network architecture of the layer-level fusion.

We still take the brain tumor segmentation in multi-sequence of MRI to illustrate this kind of fusion. It is well known that T1 weighed MRI and T1c are suitable to segment the tumor core without the peritumoral edema, while T2 and FLAIR are suitable to segment the peritumoral edema. Chen et al. [105] proposed a dual-pathway multi-modal brain tumor segmentation network. The first pathway uses T2 and FLAIR images to extract the relative feature to segment the whole tumor from the background. The second pathway uses the T1 and T1c images to train another segmentation network to learn other features. Then, the features from both pathways are fused and finally fed into a four-class classifier to segment the background, Edema (ED), ET and Necrotic (NCR)/Non-Enhancing Tumor (NET). The dual-pathway segmentation network can exploit the effective feature information of different modalities and achieve the accurate segmentation results.

Dolz et al. [89] proposed a 3D fully convolutional neural network based on DenseNets. Each imaging modality is fed into an individual path, and the dense connections occur not only between the pairs of layers within the same path, but also between those across different paths. In this way, the proposed network can learn more complex feature representations between modalities. The extensive experiment results on two different segmentation challenges: iSEG 2017 [85] and MRBrainS 2013 [83], demonstrate that the proposed method can achieve significant performance.

To summarize, in the layer-level fusion network, DenseNets are the commonly used networks, which can bring the three following benefits. First, direct connections between all layers help to improve the flow of information and gradients through the entire network, alleviating the problem of vanishing gradient. Second, short paths to all the feature maps in the architecture introduce implicit deep supervision. Third, dense connections have a regularizing effect, which reduces the risk of over-fitting on tasks with smaller training sets. In the layer-level fusion network, the connection among different layers can capture complex relationships between modalities, which can help to learn the effective feature representations for segmentation.

### 2.4.3 Decision-level Fusion Network

In decision-level fusion network, like the layer-level fusion, each modality image is used as a single input in an individual network. The individual network can exploit the unique information of the corresponding modality. The outputs of these individual networks will then be integrated to achieve the final segmentation result. Figure 2.8 describes the generic network architecture of layer-level fusion segmentation work.

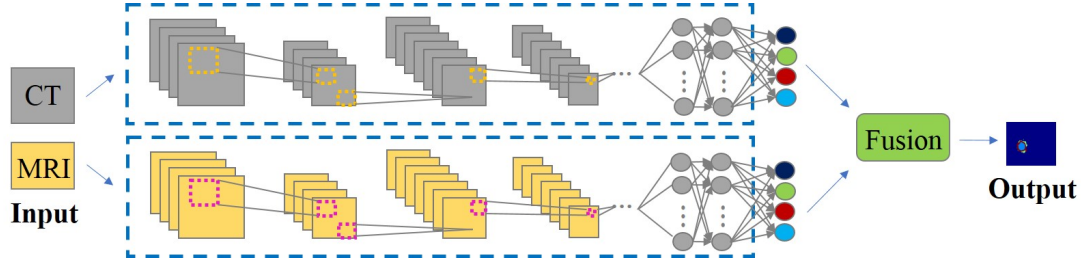


Figure 2.8 – The generic network architecture of the decision-level fusion.

For the decision-level fusion, many fusion strategies have been proposed [106]. Most of them are based on averaging and majority voting. For averaging strategy, Kamnitsas et al. [97] trained three networks separately and then averaged the confidence of the individual networks. The final segmentation is obtained by assigning each voxel with the highest confidence. For majority voting strategy, the final label of a voxel depends on the majority of the labels of the individual networks.

The statistical properties of different modalities are different, which make it difficult for a single model to directly find correlations across modalities. Therefore, in decision-level fusion segmentation network, the multiple segmentation networks can be trained to fully exploit the multi-modal features. Aygün et al. [107] investigates different fusion methods on the brain tumor segmentation problem in terms of memory and performance. In terms of memory usage, the decision-level fusion strategies require more memory since the model fuses the features later and more parameters are needed for layers to perform convolution and other operations. However, the later fusion can achieve better performance, because each modality is employed as an input in one network, which can learn complementary feature information compared to input-level fusion network.

## 2.5 Discussion and Methodologies to be Developed

In this chapter, we presented a large set of state-of-the-art multi-modal medical image segmentation networks based on deep learning. They are summarized in Table 2.4. Publicly available multi-modal medical image datasets for segmentation task are rare, the most commonly used dataset is BraTS dataset. For their segmentation, the current best method is from [111], they use the input-level fusion strategy to directly integrate the different modalities in the input space. The network architecture is based on U-Net. Also,

CHAPTER 2. A REVIEW: DEEP LEARNING FOR MEDICAL IMAGE  
SEGMENTATION USING MULTI-MODALITY FUSION

---

Table 2.4 – Summary of the deep learning approaches for multi-modal medical image segmentation, the bold presents the best performance in the challenge. The acronyms in results are: cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), the symbol \* indicates the method has available code.

Article	Pre-processing	Data	Network	Fusion level	Results (DSC)	Dataset
[94]*	Normalization Bias Field Correction	3D Patch	CNN CRF	Input	whole/core/enhanced 0.84/0.66/0.63	<b>BraTS15</b>
[56]	Normalization Bias Field Correction	2D Patch	CNN	Input	whole/core/enhanced 0.84/0.71/0.57	BraTS13
[86]*	Normalization Data Augmentation	3D Patch	U-Net ResNet	Input	whole/core/enhanced 0.85/0.74/0.64 0.85/0.77/0.64	BraTS15 BraTS17
[95]	Normalization Bias Field Correction	3D Patch	FCN CRF RNN	Input	whole/core/enhanced 0.86/0.73/0.62 0.84/0.73/0.62 4/3/2(rank)	BraTS13 BraTS15 BraTS16
[108]	Normalization	3D	U-Net ResNet	Input	whole/core/enhanced 0.87/0.75/0.64	BraTS15
[93]	Normalization Bias Field Correct	2D Slice	U-Net ResNet	Input	whole/core/enhanced 0.87/0.77/0.78	BraTS17
[87]	Normalization Data Augmentation	3D Patch	U-Net ResNet	Input	whole/core/enhanced 0.87/0.80/0.77	BraTS18
[90]	Normalization Data Augmentation Bias Field Correction	2D Patch	CNN FCN	Input	whole/core/enhanced 0.89/0.77/0.80	BraTS15
[56]	Normalization Bias Field Correction	3D	CNN	Input	whole/core/enhanced 0.88/0.81/0.76	<b>BraTS13</b>
[109]	Normalization Data Augmentation	2D	FCN	Input	whole/core/enhanced 0.87/0.81/0.72	<b>BraTS16</b>
[97]*	Normalization Bias Field Correction	3D	U-Net FCN DeepMedic	Input	whole/core/enhanced 0.88/0.78/0.72	<b>BraTS17</b>
[99]*	Normalization Data Augmentation	3D	U-Net VAE	Input	whole/core/enhanced 0.88/0.81/0.76	<b>BraTS18</b>
[110]	Normalization Data Augmentation	3D	U-Net	Input	whole/core/enhanced 0.88/0.83/0.83	<b>BraTS19</b>
[111]*	Normalization Data Augmentation	3D	U-Net	Input	whole/core/enhanced 0.88/0.85/0.82	<b>BraTS20</b>
[112]	N/A	2D Slice	CNN ResNet	Input	0.9112	<b>IVD</b>
[100]*	Normalization	3D Patch	U-Net ResNet	Input	0.59 ± 0.31 0.84 ± 0.10	<b>ISLES15</b> (SISS/SPES)
[113]*	Normalization	3D Patch	SVM	Input	CSF/WM/GM 0.78 0.88 0.84	<b>MRBrainS13</b>
[89]*	N/A	3D Patch	CNN DenseNet	Layer	CSF/WM/GM 0.95/0.91/0.90 0.84/0.90/0.86	iSEG-2017 MRBrainS13
[114]*	Normalization	3D Patch	3D DenseNet	Layer	CSF/WM/GM 0.96/0.91/0.91	<b>iSEG-2017</b>
[91]	N/A	2D Slice	U-Net DenseNet	Layer	0.9191 ± 0.0179	<b>IVD</b>
[92]	N/A	2D Patch	FCN	Decision	CSF/WM/GM 0.85/0.88/0.87	Private data

to adapt to the BraTS dataset, some modifications regarding post-processing, region-based training, data augmentation as well as several minor modifications are used to improve the segmentation accuracy.

For multi-modal medical image segmentation, the fusion strategy takes an important role in order to achieve an accurate segmentation result. Deep learning based methods outperform the conventional methods in three aspects. First, deep learning based networks learn a complex and abstract hierarchical feature representation for image data to overcome the difficulty of manual feature design. Second, deep learning based networks can present the complex relationships between different modalities by using the hierarchical network layer, such as the layer-level fusion strategy. Third, the image transform and fusion strategy in the conventional fusion strategy can be jointly generated by training a deep learning model, in this way some potential deep learning network architectures can be investigated for designing an effective image fusion strategy. Therefore, the deep learning based method has a great potential to produce better fusion results than conventional methods.

Choosing an effective deep learning fusion strategy is still an important issue. In 2013-2020 BraTS Challenge, all the methods applied the input-level fusion to directly integrate the different MR images in the input space, which is simple and can remain the intrinsic image feature and allow the method to focus on the subsequent segmentation network architecture designs, such as multi-task, multi-view, multi-scale and GAN-based strategies. While the strategy just concatenates the modalities in the input space, but it does not exploit the relationships among the different modalities. For layer-level fusion, the fusion strategy often takes the DenseNet as the basic network. The dense connections among different layers can capture complex relationships between modalities, which can help the segmentation network learn more valuable information and achieve better performance than input-level fusion. For decision-level fusion, it can achieve better performance compared to the input-level fusion, since each modality is employed to train a single network to learn independent feature representation, while this requires more memory and computational time. In summary, the layer-fusion strategy seems better. However, the results of these three fusion strategies are not obtained from the same dataset, it's difficult to compare their performance. Methodologically, each strategy has its advantages and disadvantages.

Although we observed the advantages of these fusion strategies based on deep learning. Regarding to the previous works, we can still observe that there are some locks to lift in multi-modal medical image segmentation based on deep learning. It is known that multi-modal fusion networks generally perform better than single-modal network for segmentation task. The problem is how to fuse different modalities to get the best compromise for a precise segmentation. Hence, how to design multi-modal networks to efficiently combine different modalities, how to exploit the latent relationship between different modalities, and how to integrate the multi-information into the segmentation network to improve the segmentation performance can be the topics of future works.

In this thesis, we will focus on deep learning based methods with the layer-level fusion. The objective is to exploit features in the latent space of different modalities and to fuse

the complementary information to improve the segmentation performance. The main applications are brain tumor segmentation from multimodal MRIs in the case of either the complete disposition of multimodal MRIs or their partial disposition.

## 2.6 Conclusion

In this chapter, we briefly describe the principle of deep learning, elaborated the multi-modal medical image segmentation methods and multi-modal fusion methods.

Compared with conventional methods, deep learning based methods can learn effective features and achieve superior performance in many fields. Therefore, in this work, deep learning approaches are exploited in multi-modal fusion and brain tumor segmentation. It is known that using multi-modalities can achieve better segmentation results than using single modality. However, multi-modal medical image segmentation remains challenging due to the different image characteristics of different modalities. A key challenge is to exploit the latent correlation between modalities and to fuse the complementary information to improve the segmentation performance. As we introduced in Section 2.4, there are three types of fusion methods, input-level fusion, layer-level fusion and decision-level fusion. Since the input-level fusion is to simply combine the different inputs. And decision-level fusion aims at the fusion of the outputs obtained from each source, which usually requires more memory. In this work, we choose to develop the layer-level fusion methods. All the proposed methods will be introduced in the following chapters.

## Chapter 3

# Fusion based on Attention Mechanism for Multi-modal MR Brain Tumor Segmentation

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>28</b>
<b>3.2</b>	<b>Related Works</b>	<b>29</b>
<b>3.3</b>	<b>Methodology</b>	<b>31</b>
3.3.1	The Three-stage Segmentation Network	31
3.3.2	Initial Segmentation Network (Stage 1)	31
3.3.3	Fusion Block based on Attention Mechanism (Stage 2)	33
3.3.4	Final Segmentation Network (Stage 3)	35
3.3.5	Loss Function	35
<b>3.4</b>	<b>Experimental Setup</b>	<b>36</b>
3.4.1	Data and Pre-processing	36
3.4.2	Implementation Details	36
3.4.3	Evaluation Metrics	36
<b>3.5</b>	<b>Experimental Results</b>	<b>37</b>
3.5.1	Quantitative Analysis	37
3.5.1.1	Evaluation of Our Method	37
3.5.1.2	Comparison with the State-of-the-art Methods	39
3.5.2	Qualitative Analysis	40
3.5.2.1	Evaluation of Our Method	40
3.5.2.2	Comparison with the State-of-the-art Methods	40
<b>3.6</b>	<b>Discussion and Conclusion</b>	<b>42</b>

---



1. **Tongxue Zhou**, Su Ruan, Yu Guo, and Stéphane Canu. “A Multi-Modality Fusion Network Based on Attention Mechanism for Brain Tumor Segmentation”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 377–380
  2. **Tongxue Zhou**, Su Ruan, Haigen Hu, and Stéphane Canu. “Deep Learning Model Integrating Dilated Convolution and Deep Supervision for Brain Tumor Segmentation in Multi-parametric MRI”. in: *International Workshop on Machine Learning in Medical Imaging (MLMI)*. Springer. 2019, pp. 574–582
  3. **Tongxue Zhou**, Stéphane Canu, and Su Ruan. “Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation”. In: *Computerized Medical Imaging and Graphics* 86 (2020), p. 101811
- 

### 3.1 Introduction

Brain tumor is one of the most aggressive cancers in the world [15, 16]. MRI is a widely used imaging technique to assess brain tumor, it is non-invasive and has good soft tissue contrast. It can provide invaluable information about shape, size, and localization of brain tumors without exposing the patient to high ionization radiation [115, 116, 117]. The commonly used MRI sequences are T1-weighted (T1), contrast enhanced T1-weighted (T1c), T2-weighted (T2) and FLAIR images. In this thesis, we refer to these images of different sequences as modalities. Different modalities can provide complementary information to analyze different sub-regions of gliomas. For example, T2 and FLAIR can highlight the tumor with peritumoral edema, designated whole tumor. T1 and T1c can highlight the tumor without peritumoral edema, designated tumor core. An enhancing region of the tumor core with hyper-intensity can also be observed in T1c, designated enhancing tumor core. Therefore applying multi-modal images can reduce the information uncertainty and improve clinical diagnosis and segmentation accuracy.

Inspired by the attention mechanism [118], in this chapter, we propose a three-stage multi-modality fusion network based on attention mechanism and additional constraint information for brain tumor segmentation. The conference versions of this work have been presented in [119] and [120]. The journal version has been presented in [121]. In addition, this work has also been extended to segment Covid-19 infection regions in CT imaging [122]. The main contributions of our method are four folds:

- A novel three-stage network is presented for multi-modal brain tumor segmentation.
- An attention mechanism is used to fuse different modalities to achieve the most important feature representations.
- An additional constraint information is introduced and integrated to the multi-encoder based network architecture to enhance the segmentation accuracy.
- A new loss function is proposed to solve the multi-class segmentation problem.

The rest of the chapter is organized as follows. Section 3.2 introduces the related works on brain tumor segmentation. Section 3.3 elaborates the proposed method. Section 3.4 presents the experimental setup. Section 3.5 describes and analyzes the experimental results. Section 3.6 discusses and concludes the proposed method.

## 3.2 Related Works

A wide range of approaches for brain tumor segmentation, such as probability theory [29], kernel feature selection [43], belief function [123], random forest [124], conditional random field [125], support vector machine [126] and random walk [127] have been developed with success. However, brain tumor segmentation is still a challenging task due to three reasons: (1) The brain anatomy structure varies from patients to patients. (2) The variability across size, shape, and texture of tumors. (3) The variability in intensity range and low contrast in qualitative MR imaging modalities. This is particularly true for brain tumor segmentation, where the tumor contour is fuzzy due to low contrast (see Figure 3.1).

Recently, with a strong feature learning ability, deep learning based approaches have become more prominent for brain tumor segmentation. Cui et al. [90] proposed a cascaded deep learning convolutional neural network consisting of two sub-networks. The first network is to define the tumor region from the MRI slice and the second network is used to label the defined tumor region into multiple sub-regions. Mlynarski et al. [96] introduced a CNN-based model to efficiently combine the advantages of the short-range 3D context and the long-range 2D context for brain tumor segmentation. Wang et al. [128] proposed a novel 2D fully convolution segmentation network WRN-PPNet based on the pyramid pooling module. Zhao et al. [95] integrated FCN [52] and conditional random fields to segment brain tumor. Havaei et al. [57] implemented a two-pathway architecture that learns about the local details of the brain tumor as well as the larger context features. Kamnitsas et al. [94] proposed an efficient fully connected multi-scale CNN architecture named DeepMedic, which reassembles a high resolution and a low resolution pathway to obtain the segmentation results. Furthermore, they used a 3D fully connected conditional random field to effectively remove false positives. Isensee et al. [86] modified the U-Net to brain tumor segmentation and use data augmentation to prevent the over-fitting. Kamnitsas et al. [97] introduced EMMA, an ensemble of multiple models and architectures including DeepMedic, FCNs and U-Net, and won the first position in BraTS 2017 competition.

However, these mentioned methods above only directly integrate the four MRI modalities in the input space to achieve segmentation. For multi-modal medical image segmentation task, the fusion strategy takes an important role in achieving the accurate segmentation results. In general, we can category the network architectures into single-encoder-based method and multi-encoder-based method, as presented in [129]. The single-encoder-based method [86, 97] directly integrates the different multi-modality images by channel in the input space, classified as input-level fusion, as we introduced in Section 2.4.1 in Chapter 2, while the correlation among different modalities are not well exploited. However, the multi-encoder-based method [130] allows to separately

extract individual feature information by applying multiple modality-specific encoders, and to fuse them with specific fusion strategy to emphasize the useful information for segmentation. According to [131], multi-encoder-based method has better performance than single-encoder-based method, which can learn more complementary and cross-modal interdependent features. However, not all features extracted from the encoder are useful for segmentation. Therefore, it is necessary to find an effective way to fuse features, we focus on the extraction of the most informative features for segmentation.

To this end, we propose to use the attention mechanism, which can be viewed as a tool being capable to take into account the most informative feature representation. Channel attention module and spatial attention module are the commonly used attention mechanisms. The former one uses attention mechanism to select meaningful features along channel path. For example, Hu et al. [132] introduced the Squeeze and Excitation (SE) block to perform dynamic channel-wise feature recalibration to improve the representational power of a network. Li et al. [133] proposed to combine attention mechanism and spatial pyramid to extract precise dense features for pixel labeling in semantic segmentation. Oktay et al. [134] proposed an attention U-net, which uses the channel attention mechanism to fuse the high-level and low-level features for CT abdominal segmentation. The latter one, spatial attention modules, calculate the feature representation in each position by weighted summation of features at all positions. For example, Roy et al. [118] proposed to use both spatial and channel SE (scSE) and demonstrated that scSE blocks can yield an improvement on three different FCN architectures. Recently, Roy et al. [135] applied scSE blocks to the few-shot segmentation. Fu et al. [136] presented a dual attention network using the channel and spatial attention mechanisms to adaptively integrate local semantic features with global dependencies for scene segmentation. However, the methods mentioned above evaluated the attention mechanism only on the single-modal image datasets and didn't consider the fusion issue on the multi-modal medical images. Therefore, we propose to integrate the attention mechanism to the multi-modal brain tumor segmentation.

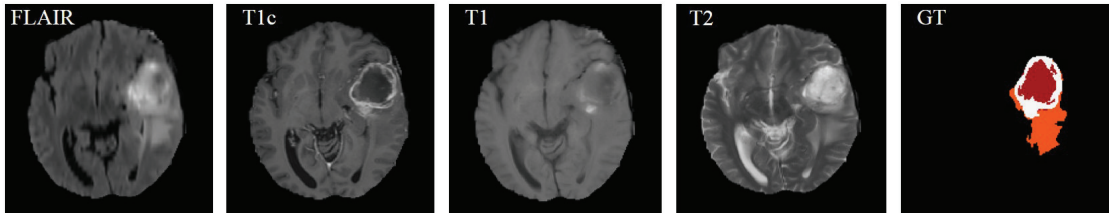


Figure 3.1 – An example from BraTS 2017 dataset [137]. The first four images from left to right show the MRI modalities: FLAIR, contrast enhanced T1-weighted (T1c), T1-weighted (T1), T2-weighted (T2) images, and the fifth image is the ground-truth labels, Net&Ncr is shown in red, edema in orange and enhancing tumor in white, Net refers non-enhancing tumor and Ncr refers necrotic tumor.

### 3.3 Methodology

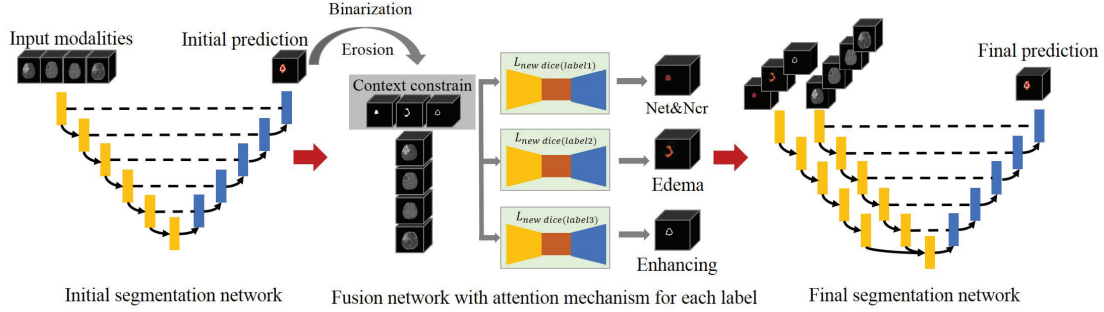


Figure 3.2 – The graphical concept of our proposed method, to simplify the presentation, we ignore the deep supervision part in these three segmentation networks, the details are shown in [120].

#### 3.3.1 The Three-stage Segmentation Network

The proposed network is a three-stage segmentation network, the graphical concept of the proposed network is illustrated in Figure 3.2. In the first stage, our previous work, the 3D U-Net architecture [120], is used as the initial segmentation network to get the rough prediction results. Then, the binarization and erosion operations are applied to each initial prediction result to get the context constraint for the following multi-encoder based fusion network. Here, the context constraint is defines as the contour of all other tumor regions except the target tumor region, which can provide some boundary information to guide the target tumor segmentation. In the second stage, to learn complementary features and cross-modal inter-dependencies, we applied the multi-encoder based framework for each label. It takes four MRI modalities and the context constraint as input in each encoder, respectively. Each encoder can produce a latent representations for the input data, and then these modality-specific features and constraint-specific feature are concatenated to the fusion block at each level. With the assistance of the attention mechanism, the feature representations will be separated along channel-wise and space-wise, and the most informative feature is obtained as the shared latent representation. Finally, the shared latent representation is projected by decoder to the label space to obtain the segmentation result for each label. In the third stage, a two-encoder based 3D U-Net segmentation network is applied to combine and refine the three single prediction results.

#### 3.3.2 Initial Segmentation Network (Stage 1)

The initial segmentation network is a 3D U-Net architecture, which has the same architecture but half initial convolutional filters than our previous work [120], which reduces the burden on graphic memory and accelerate the training process. The network architecture is described in Figure 3.3. Since standard U-Net can't get enough semantic features due

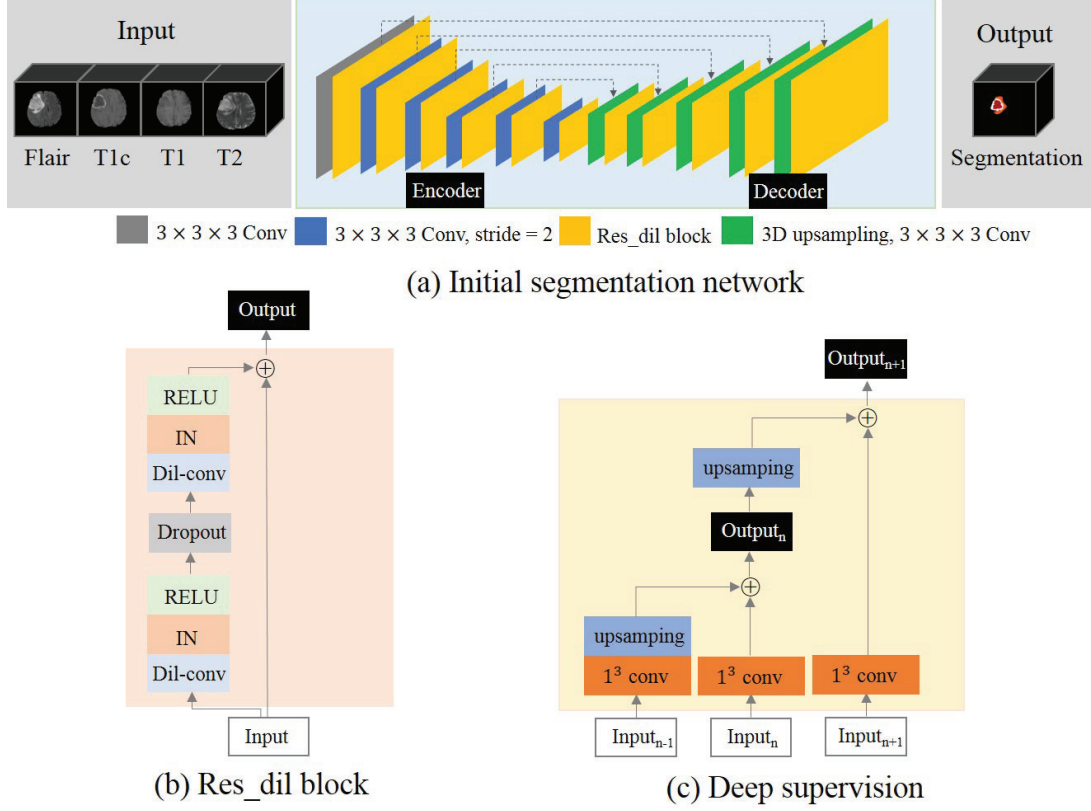


Figure 3.3 – Top: The architecture scheme of initial segmentation network. The four modalities are concatenated channel by channel in input space and the output is the segmentation predictions of the three tumor sub-regions. Bottom: The architecture scheme of our proposed res\_dil block (left) and deep supervision (right). IN refers instance normalization, Dil\_conv refers the dilated convolution (rate = 2, 4, respectively), we refer to the vertical depth as level, with higher levels being higher spatial resolution. In the deep supervision part,  $Input_n$  refers the output of res\_dil block of the  $n_{th}$  level in the decoder,  $Output_n$  refers the segmentation result of the  $n_{th}$  level in the decoder.

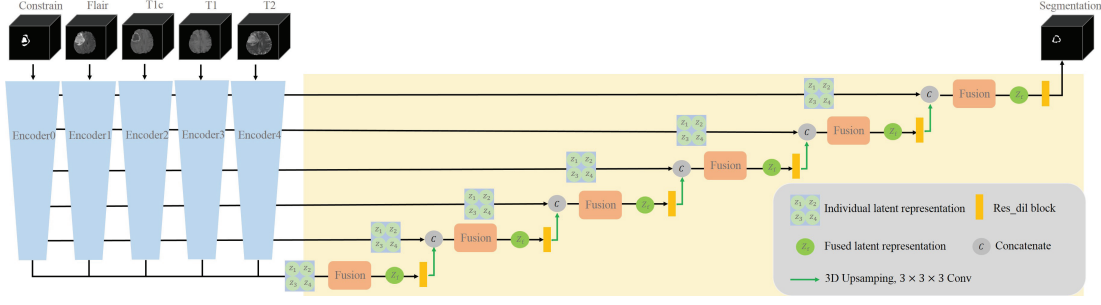


Figure 3.4 – The architecture scheme of fusion network. Each imaging modality (FLAIR, T1, T1c, T2) is encoded by a single encoder to obtain the individual latent representations ( $z_1, z_2, z_3, z_4$ ), and the context constraint can provide boundary information to refine the segmentation result. Then, the five encoders are fused into the shared representation space with the fusion block. Finally the fused latent representation  $Z_f$  is decoded by the decoder to obtain the segmentation result. Here we present the segmentation network architecture of enhancing tumor, it is same for other tumor regions.

to the limited receptive field. Inspired by dilated convolution [138], we use residual block with dilated convolutions, denoted as `res_dil`, on both encoder part and decoder part to obtain features at multiple scales. It can obtain more extensive local information to help retain information and fill details during training process. The encoder is used to get the latent feature representation from the four modalities. It includes a convolutional block and a `res_dil` block with skip connections. The decoder is used to recover the image details. It begins with a up-sampling layer followed by a  $3 \times 3 \times 3$  convolution to adjust the number of features, then the upsampled features are combined with the features from the corresponding level of the encoder part using concatenation. After the concatenation, a  $3 \times 3 \times 3$  convolution and a `res_dil` block is used to increase the receptive field. In addition, deep supervision [86] is used to combine multi-scale segmentation results at different layers to improve the final segmentation.

### 3.3.3 Fusion Block based on Attention Mechanism (Stage 2)

Multi-modality fusion network can capture more specific and effective information for different modalities than the single-encoder based network [119]. Therefore, three five-encoder based networks are used to compose the fusion network, where each network is used to segment a single tumor region. The architecture of the fusion network is presented in Figure 3.4. To learn complementary features and cross-modal inter-dependencies from multi-modalities, we applied the attention mechanism to the fusion block. Since the three tumor regions are close to each other, which can lead to produce more falsely predicted pixels in the neighbor regions. Therefore, the context constraint is proposed to provide more boundary information to benefit the segmentation. We set  $L$  the label set,  $L = \{l_1, l_2, \dots, l_M\}$ , where  $M$  is the number of labels. For example, when segmenting label  $l_i$ ,  $i \in M$ , all other initial prediction labels  $l_j$ ,  $j \in M$ ,  $j \neq i$ , are processed as the

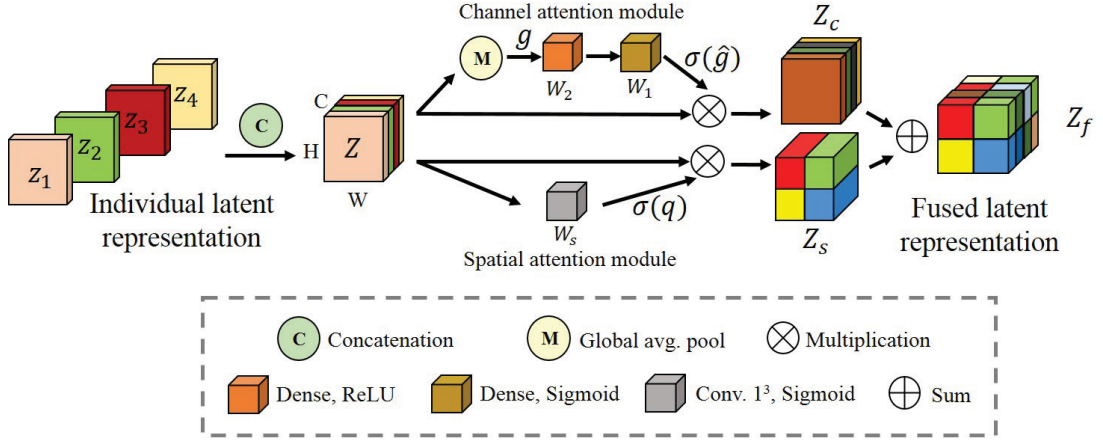


Figure 3.5 – The architecture scheme of fusion block. The individual latent representations ( $z_1, z_2, z_3, z_4$ ) are first concatenated as the input of the attention mechanism  $Z$ . Then, they are recalibrated along channel attention module and spatial attention module to achieve the  $Z_s$  and  $Z_c$ . Finally, they are added to obtain the fused latent representation  $Z_f$ .

context constraint to guide the segmentation of target tumor label (label  $l_i$ ) and refine the segmentation result. In addition, considering the location relationship among the three tumor regions, a new multi-class loss function (Equation 3.6) is proposed to take into account the more constraint information to boost the segmentation results. In addition, to make the network more efficient, during training, the three single label segmentation networks are running in parallel.

The purpose of fusion block is to stand out the most important features from different modalities to highlight regions that are greatly relevant to brain tumor segmentation. One simple way to fuse the independent latent representations is to average over them, while it could lose some valuable information in the latent representation. To this end, we propose a fusion block, described in Figure 3.5. The individual latent representations ( $z_1, z_2, z_3, z_4, \dots, z_n$ ) with  $n$  modalities are first concatenated to obtain the input feature map  $Z = [z_1, z_2, z_3, z_4, \dots, z_n]$ ,  $z_k \in R^{H \times W}$ . Note that, in the lowest level of the network, there are 4 modality-specific features ( $z_1, z_2, z_3, z_4$ ) for the fusion block in our case. In the other levels, the result of the previous level in the decoder is also concatenated with the modality-specific features to obtain the input feature map  $Z = [z_1, z_2, z_3, z_4, z_5]$ ,  $z_k \in R^{H \times W}$  of the fusion block. In the following, we describe the fusion block with the 4 modality-specific features. In the channel attention module, a global average pooling is first performed to produce a tensor  $g \in R^{1 \times 1 \times 4}$ , which represents the global spatial information of the feature map, with its  $k^{th}$  element

$$g_k = \frac{1}{H \times W} \sum_i^H \sum_j^W z_k(i, j) \quad (3.1)$$

Then two fully-connected layers are applied to encode the channel-wise dependencies,  $\hat{g} = W_1(\delta(W_2g))$ , with  $W_1 \in R^{4 \times 2}$ ,  $W_2 \in R^{2 \times 4}$ , being weights of two fully-connected layers and the ReLU operator  $\delta(\cdot)$ .  $\hat{g}$  is then passed through the sigmoid layer ( $\sigma$ ) to obtain the channel-wise weights, which will be applied to the input map  $Z$  through multiplication to achieve the channel-wise features  $Z_c$ , the  $\sigma(\hat{g}_k)$  indicates the importance of the  $i$  channel of the feature map.

$$Z_c = [\sigma(\hat{g}_1)z_1, \sigma(\hat{g}_2)z_2, \sigma(\hat{g}_3)z_3, \sigma(\hat{g}_4)z_4,] \quad (3.2)$$

In the spatial attention module, the feature map can be considered as  $Z = [z^{1,1}, z^{1,2}, \dots, z^{i,j}, \dots, z^{H,W}]$ ,  $z^{i,j} \in R^{1 \times 1 \times 4}$ ,  $i \in 1, 2, \dots, H$ ,  $j \in 1, 2, \dots, W$ , and then a convolution operation  $q = W_s \star Z$ ,  $q \in R^{H \times W}$  with weight  $W_s \in R^{1 \times 1 \times 4 \times 1}$ , is used to squeeze the spatial domain, and to produce a projection tensor, which represents the linearly combined representation for all channels for a spatial location. The tensor is finally passed through a sigmoid layer to obtain the space-wise weights  $\sigma(q_{i,j})$ , which indicates the importance of the spatial information  $(i, j)$  of the feature map. The space-wise weights  $\sigma(q_{i,j})$  is multiplied with the input map  $Z$  to obtain the space-wise features  $Z_s$ .

$$Z_s = [\sigma(q_{1,1})z^{1,1}, \dots, \sigma(q_{i,j})z^{i,j}, \dots, \sigma(q_{H,W})z^{H,W}] \quad (3.3)$$

The fused feature representation is finally obtained by adding the channel-wise feature and space-wise feature.

$$Z_f = Z_c + Z_s \quad (3.4)$$

### 3.3.4 Final Segmentation Network (Stage 3)

To combine the three single segmentation results and achieve the final segmentation result, a two-encoder based 3D U-Net segmentation network is applied. The architectures of the encoder and decoder are the same as the fusion network. One encoder takes the concatenation of the three predicted probability maps in the second stage as input, and another encoder takes the concatenation of the four original modalities as input. In this way, the final segmentation network can not only combine the three single segmentation results but also take advantage of them to refine the final segmentation performance.

### 3.3.5 Loss Function

Dice loss function (Equation 3.5) is commonly used for medical image segmentation problem. However, for multi-class segmentation problem, the location relationship of multiple classes should be considered. We introduce a new loss function (Equation 3.6) to the multi-encoder based segmentation network (Stage 2). To avoid other labels to be falsely predicted into the target label region, we take the Dice score of all other labels into the loss function to constrain the target label, and enhance the segmentation accuracy.

$$L_{dice} = 1 - 2 \frac{\sum_{l \in L} \sum_{i \in N} y_i^{(l)} \hat{y}_i^{(l)} + \epsilon}{\sum_{l \in L} \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \epsilon} \quad (3.5)$$



$$L_{new\ dice}(l_j) = 1 - 2 \left[ \frac{\sum_{i \in N} y_i^{(l_j)} \hat{y}_i^{(l_j)} + \epsilon}{\sum_{i \in N} (y_i^{(l_j)} + \hat{y}_i^{(l_j)}) + \epsilon} - \alpha \sum_{l_k \in L, j \neq k} \frac{\sum_{i \in N} y_i^{(l_k)} \hat{y}_i^{(l_j)} + \epsilon}{\sum_{i \in N} (y_i^{(l_k)} + \hat{y}_i^{(l_j)}) + \epsilon} \right] \quad (3.6)$$

where  $N$  is the set of samples,  $L$  is the set of all labels, in our task, there are three labels: net&ncr (label 1), edema (label 2) and enhancing tumor (label 3).  $y_i^{(l)}$  is the ground-truth for the sample  $i$  and label  $l$ ,  $\hat{y}_i^{(l)}$  is the predicted probability for the same sample and label pair,  $\epsilon$  is a small constant to avoid dividing by 0. We use  $L_{dice}$  as the loss function for the initial segmentation network and final segmentation network, and  $L_{new\ dice}$  for the multi-encoder based fusion network.

## 3.4 Experimental Setup

### 3.4.1 Data and Pre-processing

The datasets used in the experiments come from BraTS 2017 dataset [137]. The training set includes 210 HGG patients and 75 LGG patients. Each patient has four image modalities including T1-weighted (T1), contrast enhanced T1-weighted (T1c), T2-weighted (T2) and FLAIR images. All data used in the experiments have been pre-processed with a standard procedure. The N4ITK [139] method is first used to correct the distortion of MRI data, and intensity normalization is applied to normalize each modality to a zero-mean, unit-variance space. To exploit the spatial contextual information of the image, we use the 3D image, clip and resize the images from  $155 \times 240 \times 240$  to  $128 \times 128 \times 128$ . Following the challenge, four intra-tumor structures have been grouped into three mutually inclusive tumor regions: (a) The WT, consisting of all tumor tissues. (b) The TC, consisting of the enhancing tumor, necrotic and non-enhancing tumor core. (c) The ET, consisting of the enhancing tumor.

### 3.4.2 Implementation Details

The experiment is implemented in Keras with a single Nvidia GPU Quadro P5000 (16G). The network is optimized using the Adam optimizer (initial learning rate =  $5e-4$ ) with a decreasing learning rate factor 0.5 with patience of 10 epochs. To avoid over-fitting, early stopping is used when the validation loss isn't improved for 50 epoch. We randomly split the dataset into 80% training and 20% testing.

### 3.4.3 Evaluation Metrics

To evaluate the proposed methods, two evaluation metrics: DSC and HD are used to obtain quantitative measurements of the segmentation accuracy.

1) DSC : It is designed to evaluate the overlap rate of prediction results and ground truth. Dice ranges from 0 to 1, and the better predict results will have a larger value.

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (3.7)$$

where  $TP$  represents the number of true positive voxels,  $FP$  represents the number of false positive voxels, and  $FN$  represents the number of false negative voxels.

2) HD: It is computed between the boundaries of prediction results and ground-truth, which is an indicator of the largest segmentation error. The better prediction result will have a smaller value.

$$HD = \max\{\max_{s \in S} \min_{r \in R} d(s, r), \max_{r \in R} \min_{s \in S} d(r, s)\} \quad (3.8)$$

where  $S$  and  $R$  are the two sets of the surface points of the prediction and the ground-truth, and  $d$  is the Euclidean distance.

## 3.5 Experimental Results

We conduct the comparative experiments to demonstrate the effectiveness of our proposed method. In Section 3.5.1.1, we evaluate the effectiveness of the proposed components. In Section 3.5.1.2, we compare our method with the state-of-the-art methods. In Section 3.5.2.1, we carry out the qualitative experiment to further demonstrate the advantages of our proposed method. In Section 3.5.2.2, we visualize the segmentation results compared with the state-of-the-art methods.

### 3.5.1 Quantitative Analysis

#### 3.5.1.1 Evaluation of Our Method

We first evaluate the proposed fusion network. The network applied context constraint and a new loss function to guide the network to obtain the three tumor sub-regions: non-enhancing and necrotic, edema and enhancing tumor, respectively. To see the impact of the components of the fusion network (Stage 2), including context constraint and the new loss function. We refer to the network without context constraint and the new loss function as backbone. From Table 3.1, we can observe the backbone method achieves DSC of 64.58%, 68.58%, 58.31% for enhancing, edema and net&ncr tumor, respectively. When the Dice loss function is replaced by our proposed new loss function, we can see an increase of DSC across all tumor regions, the new loss function can help to constrain the target label not to be falsely predicted to the neighbor tumor regions. To choose the optimal coefficient in the new loss function, we did a grid search between  $[0, 2]$  and found the best coefficient is 0.1. Table 3.2 reports the performance of the coefficient  $\alpha$  in the new loss function. In addition, when the context constraint is integrated to one of the encoder of the fusion network, the segmentation results of all the tumor regions are

CHAPTER 3. FUSION BASED ON ATTENTION MECHANISM FOR  
MULTI-MODAL MR BRAIN TUMOR SEGMENTATION

Table 3.1 – Segmentation results of fusion network, bold results show the best scores for each tumor region, backbone refers to the four-encoder based network without new loss and context constraint, Net refers non-enhancing tumor, Ncr refers necrotic tumor, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC (%)		
	Enhancing	Edema	Net&Ncr
Backbone	64.58	68.58	58.31
+ New loss	66.18*	72.00*	60.09*
+ New loss + Context constraint	<b>68.70*</b>	<b>75.28*</b>	<b>61.42*</b>

Table 3.2 – The choice of the coefficient  $\alpha$  in the new loss function.

$\alpha$	0	0.05	<b>0.1</b>	0.5	1	2
DSC (%)	64.58	64.41	<b>66.18</b>	63.27	61.25	44.49

improved. We conclude that the context constraint can definitely boost the segmentation results.

To evaluate the effectiveness of each stages in our method, we compare their results in Table 3.3. Compared to the results of Stage 1, except the enhancing tumor, which decreases 1.0% of the DSC, all the tumor regions have a large improvement in DSC in Stage 2. The main reason for the decrease of the DSC on enhanced tumor is that, the enhancing tumor usually locates between the edema and net&ncr regions, the contours are usually diffused and there are no clear cut with the other two regions. However, with assistance of the fusion network (Stage 3), which can help to refine the three single label segmentations, the enhancing tumor region increases 6.26% of the DSC compared to Stage 2. And all the DSC of the other tumor regions are also further improved.

Table 3.3 – Segmentation results of our method on BraTS 2017 dataset, bold results show the best scores for each tumor region, Stage 1 refers to initial segmentation network, Stage 2 refers to fusion network and Stage 3 refers to final segmentation network, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC (%)				
	WT	TC	ET	ED	Net&Ncr
Stage 1	86.84	75.88	69.40	72.02	58.28
Stage 2	N/A	N/A	68.70	75.28*	61.42*
Stage 3	<b>89.40</b>	<b>81.64</b>	<b>73.00*</b>	<b>75.59</b>	<b>64.63*</b>

### CHAPTER 3. FUSION BASED ON ATTENTION MECHANISM FOR MULTI-MODAL MR BRAIN TUMOR SEGMENTATION

Table 3.4 – Comparison of our proposed method and other related methods on BraTS 2017, bold results show the best scores for each tumor region, and underline results refer the second best results, AVG denotes the average results on the three tumor regions.

Methods	DSC (%)				HD			
	WT	TC	ET	AVG	WT	TC	ET	AVG
Ronneberger et al. [53]	79.1	49.9	8.0	45.7	18.80	21.10	38.00	30.0
Pereira et al. [140]	86.6	76.6	69.8	77.7	8.48	10.51	<b>6.13</b>	8.37
Isensee et al. [86]	<b>89.5</b>	<b>82.8</b>	<u>70.7</u>	<u>81.0</u>	<u>6.04</u>	<u>6.95</u>	<u>6.24</u>	<b>6.41</b>
Jesson et al. [141]	88.6	78.9	68.2	78.6	6.58	7.11	8.11	7.27
Ours	<u>89.4</u>	<u>81.6</u>	<b>73.0</b>	<b>81.3</b>	<b>5.73</b>	<b>6.79</b>	7.68	<u>6.73</u>

#### 3.5.1.2 Comparison with the State-of-the-art Methods

To demonstrate the performance of our method, we compare our segmentation results with the state-of-the-art methods in the MICCAI 2017 challenge. In BraTS 2017, the top performing method used an ensemble of FCN. In principle, building an ensemble network will certainly lead to better results. Since we evaluate the effect of our proposed network, so we compare our method with other approaches in single model. The compared algorithms are given as follows:

- (1) Ronneberger et al. [53] proposed U-Net, the widely used and effective approach to segment the medical images.
- (2) Pereira et al. [140] proposed segmentation SE (SegSE) block to create more complex features for feature recalibration in brain tumor segmentation.
- (3) Jesson et al. [141] employed a multi-scale loss function to combine higher resolution features with the lower level segmentation results for brain tumor segmentation.
- (4) Isensee et al. [86] modified the U-Net for 3D brain tumor segmentation, which used both context and location pathways to learn the complex feature representation.

The compared results are summarized in Table 3.4, we also compare the computational complexity among these methods shown in Table 3.5. The best result in single model is from [86], which achieves 89.5%, 82.8% and 70.7% on the DSC on whole tumor, tumor core and enhancing tumor regions, respectively. However, it uses 16 initial convolution filters and data augmentation, which is time-consuming (training time is about five days). We can observe that our proposed method has a superior result in the terms of average DSC. Specially, compared to the best approach [86], we can achieve almost the same DSC on whole tumor (decreased 0.1%) and a slightly decrease can be seen on tumor core (decreased 1.2%). However, we can achieve the best DSC on enhancing tumor, the most challenging tumor region, and also the best average DSC. In addition, our method yields the best HD on both whole tumor and tumor core regions, which indicates that our method has a minimum segmentation error on the two regions. Therefore, the proposed method can achieve the competitive results.

## CHAPTER 3. FUSION BASED ON ATTENTION MECHANISM FOR MULTI-MODAL MR BRAIN TUMOR SEGMENTATION

Table 3.5 – Computational complexity comparison of different methods including data dimension, input size, number of layer, number of initial filter, data augmentation, post-processing, GPU and training time. ”-” indicates that the information is not provided in the published paper.

Methods	Dimension	Size	Layer	Filter	Aug	Post	GPU	Time
Ronneberger et al. [53]	3D	$128 \times 128 \times 128$	4	32	No	No	16G	10h
Pereira et al. [140]	3D	$128 \times 128 \times 128$	4	12	Yes	Yes	-	-
Isensee et al. [86]	3D	$128 \times 128 \times 128$	5	16	Yes	No	-	120h
Jesson et al. [141]	3D	$184 \times 200 \times 152$	4	32	No	No	-	-
Ours	3D	$128 \times 128 \times 128$	6	8	No	No	16G	40h

### 3.5.2 Qualitative Analysis

In order to evaluate the effectiveness of our model, we randomly select several examples on BraTS 2017 dataset and visualize the results in Figure 3.6 - Figure 3.8. Since the quantitative results of U-Net is not good, so we’ll not compare our results with it. Except the best method in single model [86], the other related works didn’t provide the available code, so we can’t do the related comparison.

#### 3.5.2.1 Evaluation of Our Method

Figure 3.6 shows the comparison results between different methods in Stage 2: backbone, backbone with new loss, backbone with new loss and context constraint. We can observe the backbone network generates many false predictions, for example, it detects some false pixels in the three tumor sub-regions. However, the false predictions are corrected when the new loss is applied, but there are still some false pixels, like in the enhancing tumor and edema regions, some isolated pixels are failed to be detected. However, when the context constraint is applied to the network, it provides some boundary information for target tumor segmentation, the results are further refined which achieves the best results.

Figure 3.7 shows the comparison results in the three stages of our method on several examples. Firstly, the initial segmentation network can obtain a rough segmentation in Stage 1, but with some falsely predicted pixels. While in Stage 2, the context constraint can provide more boundary information and help to refine each tumor region. Finally, in stage 3, the fusion network combines the three single regions to the form the final segmentation result. We can observe that by using the three-stage segmentation network, we can achieve the best segmentation results in multi-modal brain tumor segmentation task.

#### 3.5.2.2 Comparison with the State-of-the-art Methods

We compare our results with the best method in single model in Brats 2017, which uses twice more convolution filters than us, the comparison results are shown in Figure 3.8. For the first example, the method of [86] predicted many false isolated edema regions and failed to detect some the net&ncr regions. In the second example, the method of [86]

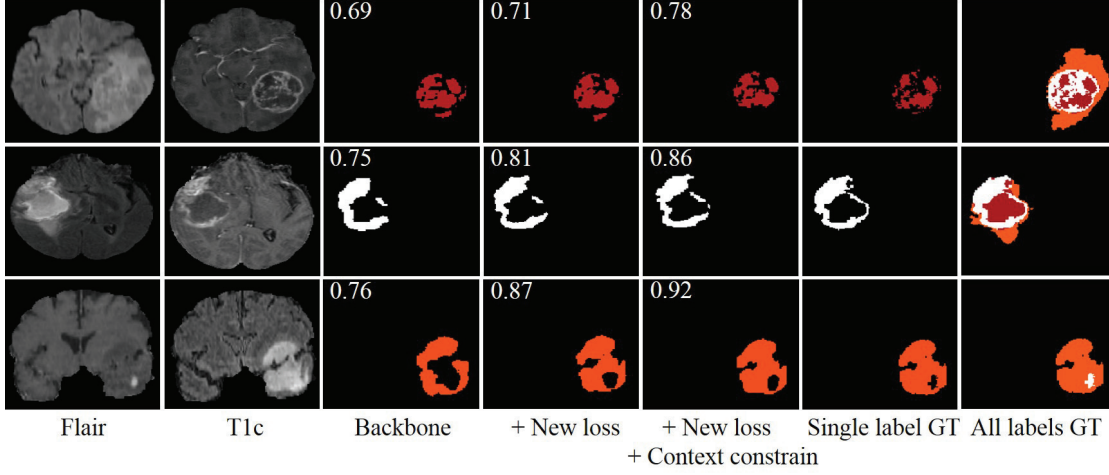


Figure 3.6 – Qualitative comparison among different strategies of our method in Stage 2 on several examples. We denote the DSC on each result. Net&Ncr is shown in red, edema in orange and enhancing tumor in white.

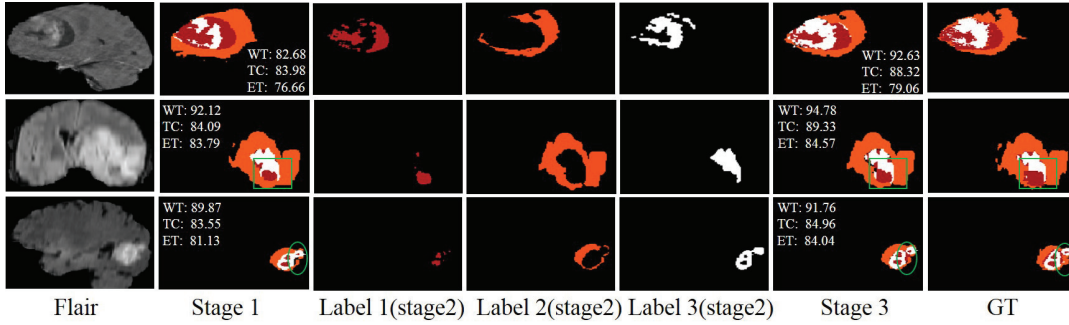


Figure 3.7 – Qualitative results in the three stages of our method on several examples. We denote the DSC on each result in Stage 1 and Stage 3. Label 1 (red): net&ncr, Label 2 (orange): edema, Label 3 (white): enhancing tumor. The green bounding box emphasizes the differences of segmentation results among different methods.

over-segments an edema region on the left top. And in the third example, it not only fails to detect the small enhancing tumor on the boundary but also produces many false isolated predictions on edema region. However, our three-stage segmentation network can gradually refine the results from previous stage, and finally achieve the superior results on all examples.

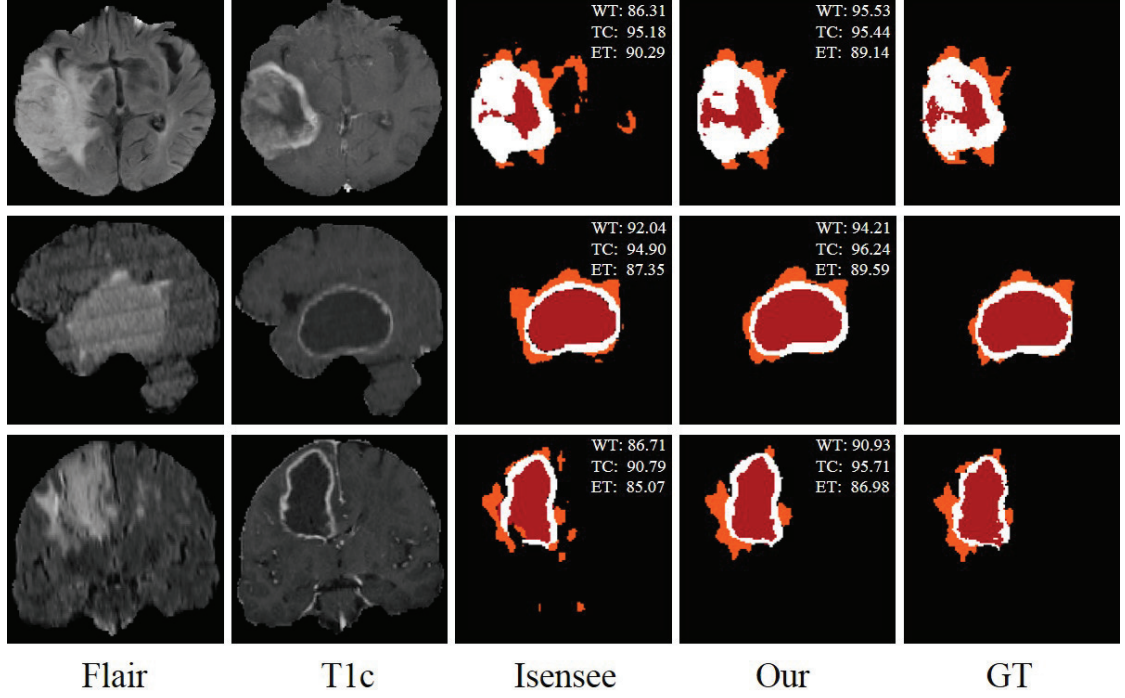


Figure 3.8 – Qualitative results between our method (8 initial filters) and method from [86] (16 initial filters) on several examples. We denote the DSC on each result. Net&Ncr is shown in red, edema in orange and enhancing tumor in white.

### 3.6 Discussion and Conclusion

In this chapter, we propose a novel three-stage network based on context constraint and attention mechanism for multi-modal brain tumor segmentation. To decrease the influence of the fuzzy contour in the brain tumors, we first used a initial segmentation network to produce a context constraint for each tumor region. Then, under the constraint information, we applied a multi-encoder based network to achieve three single tumor region segmentations. Specifically, the attention mechanism is introduced to achieve the fusion of different modalities. In addition, considering the location relationship of the tumor regions, a new loss function is proposed to cope with the multiple class segmentation problem. Finally, a two-encoder based 3D U-Net segmentation network is presented to combine and refine three single prediction results to form the final segmentation result.

The proposed method has three major advantages. First, the proposed attention mechanism based fusion strategy can learn the complementary feature information across different modalities and extract the most useful features related to target regions. Second, the network can produce and take advantage of the context constraint information to help segment the brain tumor regions with the obscure contours. Third, the proposed multi-class segmentation loss function can utilize the hierarchical structure of the brain tumor to avoid the false prediction in the adjacent tumor regions. And it can be extended to other research fields.

Although our method gives good results, it did not leverage the feature maps of each modality, the performance of the proposed fusion method can be further improved by considering latent correlation between multi-modalities. In the following chapters, we will present the proposed works which take the multi-source correlation into account to help the brain tumor segmentation.





## Chapter 4

# Fusion based on Feature Correlation in Latent Space for Multi-modal Brain Tumor Segmentation

### Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>46</b>
<b>4.2</b>	<b>Related work</b>	<b>46</b>
<b>4.3</b>	<b>Methodology</b>	<b>48</b>
4.3.1	Network Architecture	48
4.3.2	Feature Correlation and Tri-attention Fusion Strategy	49
<b>4.4</b>	<b>Experimental Setup</b>	<b>52</b>
4.4.1	Data and Pre-processing	52
4.4.2	Implementation Details	52
4.4.3	Evaluation Metrics	54
<b>4.5</b>	<b>Experimental Results</b>	<b>54</b>
4.5.1	Quantitative Analysis	54
4.5.1.1	Ablation Study	54
4.5.1.2	Comparison with the State-of-the-art Methods	55
4.5.2	Qualitative Analysis	59
<b>4.6</b>	<b>Discussion</b>	<b>59</b>
4.6.1	Performance Analysis on Correlation Expression	59
4.6.2	Performance Analysis on Correlation Attention Module	60
4.6.3	Visualization of Feature Maps	62
<b>4.7</b>	<b>Conclusion</b>	<b>63</b>

---

1. **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “3D Medical Multi-modal Segmentation Network Guided by Multi-source Correlation Constraint”. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021, pp. 10243–10250
  2. **Tongxue Zhou**, Su Ruan, Pierre Vera, and Stéphane Canu. “A Tri-attention Fusion Guided Multi-modal Segmentation Network”. In: *Pattern Recognition* (2021), p. 108417
- 

## 4.1 Introduction

The proposed method in Chapter 3 doesn’t exploit the intrinsic relations of features in latent space. For example, it doesn’t take into account the correlation between modalities. Based on the fact that, there is strong correlation between multi MR modalities, since the same scene (the same patient) is observed by different modalities [29, 142], we propose a novel tri-attention fusion approach, including channel attention, spatial attention, and correlation of multimodal features, to guide 3D multi-modal brain tumor segmentation network. The conference version has been presented in the International Conference on Pattern Recognition (ICPR) [143]. The journal version has been presented in Pattern Recognition [144]. The main contributions in this chapter are concluded as follows:

- A novel correlation description block is introduced to discover the latent multi-source correlation between modalities.
- A correlation constraint using Kullback–Leibler divergence is proposed to aide the segmentation network to extract the correlated feature representation for a better segmentation.
- A tri-attention fusion strategy is proposed to re-weight the feature representation along modality-attention, spatial-attention and correlation-attention paths.
- A new 3D multi-modal brain tumor segmentation network guided by tri-attention fusion is proposed.

The rest of the chapter is organised as follows. Section 4.2 reviews the relevant prior works. Section 4.3 presents our proposed method. Section 4.4 describes the experimental setup. Section 4.5 presents the experimental results. 4.6 gives a further discussion about our method. Section 4.7 concludes our work.

## 4.2 Related work

A number of conventional brain tumor segmentation approaches haven been presented in recent years [29, 123, 125, 145]. However, the performance is limited due to the complex brain anatomy structure, different shape, texture of gliomas, and the low contrast of MR images (see Figure 4.1).

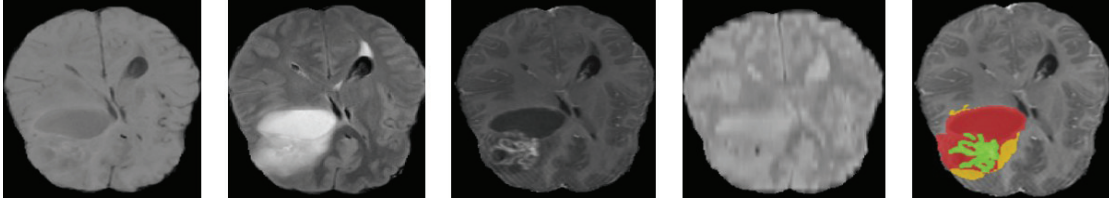


Figure 4.1 – An example from BraTS 2018 dataset[7]. The first four images from left to right show the MRI modalities: T1-weighted (T1), FLAIR, contrast enhanced T1-weighted (T1c), T2-weighted (T2) images, and the fifth image is the ground-truth labels created by experts. The color is used to distinguish the different tumor regions: red: necrotic and non-enhancing tumor, yellow: edema, green: enhancing tumor.

In recent years, various deep learning-based approaches have been successfully designed for brain tumor segmentation, such as CNN [93, 146], FCN [95] and U-Net [147, 99, 148, 58, 59]. Wang et al. [93] proposed to decompose the multi-class segmentation problem into a sequence of binary segmentation problems according to the sub-region hierarchy. Dolz et al. [146] presented an ensemble of deep CNNs to segment isointense infant brains in multi-modal MRI images, where a early fusion strategy is adopted to combine these different modalities. Chen et al. [58] proposed a dual-force training scheme to make use of multi-level information for more accurate segmentation. However, the above mentioned methods directly concatenate the four MRI modalities in the input space to obtain the segmentation. As we introduced in Chapter 2.4.1 [129], in the input-level fusion method, the complimentary information between different modalities can't be well exploited. However, the layer-level fusion method can learn the better latent feature representations to improve the segmentation. For example, Wei et al. [148] proposed a multi-model, multi-size and multi-view deep model to extract the useful features from different modalities for brain tumor segmentation. Zhang et al. [59] proposed a cross-modality deep feature learning framework for brain tumor segmentation, consisting of a cross-modality feature transition process and a cross-modality feature fusion process. It can learn the cross-modality feature representations using the knowledge transition. Tseng et al. [130] proposed a deep encoder-decoder network with cross-modality convolution layers to better exploit the multi-modalities.

The multi-encoder-based method applies separate encoders to extract individual feature representations, which can provide more room to exploit the complimentary information from different modalities, and can achieve the better segmentation results than the single-encode-based method [119]. Therefore, we propose a multi-encoder-based network architecture. In this chapter, we first proposed a dual attention based fusion block to selectively emphasize feature representations, which consists of a modality attention module and a spatial attention module (presented in the Chapter 3). The proposed fusion block uses the individual features obtained from encoders to derive a modality-wise and a spatial-wise weight map which quantify the relative importance of each modality's features and also of the different spatial locations across all the modalities. These fusion maps are then multiplied with the individual feature representations to obtain a fused

feature representation of the complementary multi-modality information. In this way, we can discover the most relevant characteristics to aide the segmentation.

Different methods have recently emerged to exploit the latent feature representations, such as multiple kernel learning [149], multiset canonical correlation analysis [150] and Variational Auto-Encoder (VAE) [151]. Moreover, discovering the latent correlation between modalities is essential to improve the segmentation. For multi-modal MR brain tumor segmentation, since the four MR modalities are from the same patient, there exists a strong correlation in the tumor regions between modalities [29]. To this end, we proposed a correlation attention module, which consists of a correlation description block and a KL divergence based correlation constraint. It can exploit and utilize the correlation between modalities to improve the segmentation performance. In the correlation attention module, based on the dual-attention fusion block, a correlation description block is first used to exploit the correlation between the spatial-attention feature representations. Then, a correlation constraint based on KL divergence is used to guide the segmentation network to learn the correlated features to enhance the segmentation result. The novelty of this method is capable of exploiting and utilizing the latent multi-source correlation to help the segmentation. The proposed method can be generalized to other applications.

### 4.3 Methodology

In this work, we aim to exploit the multi-source correlation between modalities and utilize the correlation to constrain the network to learn more effective features so as to improve the segmentation performance. The overview of the proposed network is described in Figure 4.2. U-Net is a neural network architecture widely used to medical image segmentation. The basic structure of a U-Net architecture consists of two paths. The encoder path is to extract feature representations at multiple different levels. The decoder path allows the network to project the discriminative features learnt by the encoder to the pixel space to get a dense prediction. To learn complementary features and cross-modal inter-dependencies from multi-modality MRIs, we applied the multi-encoder based U-Net framework. It takes 3D MRI modality as input in each encoder. Each encoder can produce a modality-specific feature representation. At the lowest level of the network, the tri-attention fusion block is used, which includes a dual-attention fusion block and a correlation attention module. The dual-attention fusion block can re-weight the feature representation along modality-wise and spatial-wise. The correlation attention module is to first exploit the latent multi-source correlation between the spatial-attention feature representations. Then, it uses a correlation based constraint to guide the network to learn the effective feature information. Finally, the fused feature representation is projected by decoder to the label space to obtain the segmentation result.

#### 4.3.1 Network Architecture

The proposed network is a multi-encoder one decoder based network (see Figure 4.2). The encoder includes a convolutional block, a res\_dil block (see Figure 3.3 in Chapter 3) followed by skip connection. All convolutions are  $3 \times 3 \times 3$ . Each decoder level begins

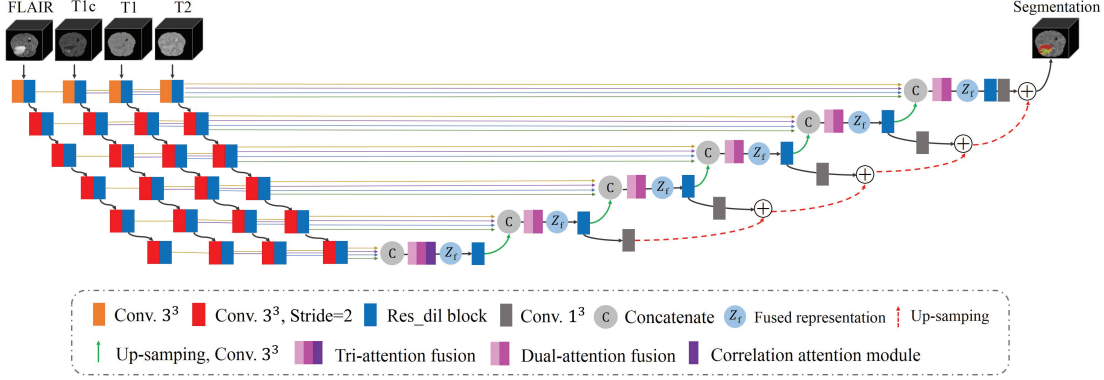


Figure 4.2 – Overview of our proposed segmentation network. The backbone is a multi-encoder based 3D U-Net, the separate encoders enable the network to extract the independent feature representations. The proposed dual-attention fusion block is to re-weight the feature representations along modality and space paths. The tri-attention fusion block consists of the dual-attention fusion and a correlation attention module.

with up-sampling layer followed by a convolution to adjust the number of features. Then the upsampled features are combined with the features from the corresponding level of the encoder part using concatenation. After the concatenation. Then the dual-attention fusion block is proposed to re-weight the feature representation along modality-wise and spatial-wise. Following that, the res\_dil block is used to increase the receptive field. In addition, we employ deep supervision [86] for the segmentation decoder by integrating segmentation results from different levels to form the final network output.

#### 4.3.2 Feature Correlation and Tri-attention Fusion Strategy

The purpose of fusion is to stand out the most important features from different source images to highlight regions that are greatly relevant to the target region. Since different MR modalities can identify different attributes of the target tumor. In addition, from the same MR modality, we can learn different information at different positions. To this end, we propose to use the attention mechanism (presented in Chapter 3) as the dual-attention fusion block, which consists of a modality attention module and a spatial attention module.

Based on the fact that, there is a strong correlation between multi MR modalities, since the same brain tumor region is observed by different modalities [29]. From Figure 4.3 presenting joint intensities of the MR images, we can observe a strong correlation (not always linear) in intensity distribution between each pair of modalities. Therefore, it's reasonable to assume that a strong correlation also exists in latent feature representation between modalities. To greatly leverage of this correlation, we proposed a correlation attention module and integrated it to the dual-attention fusion block to achieve a tri-attention fusion block. It's used to exploit and utilize the multi-source correlation between modalities, the architecture is depicted in Figure 4.4.

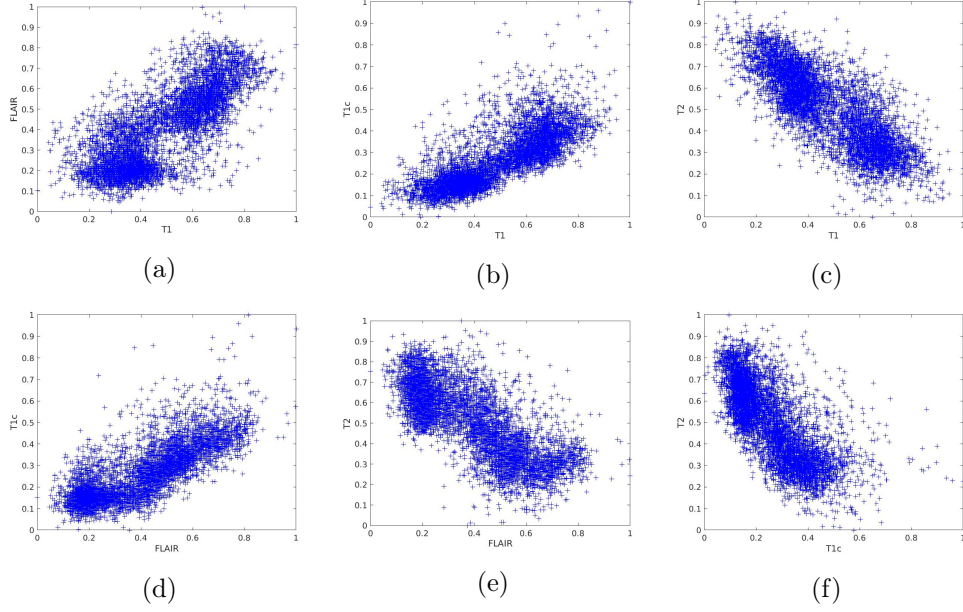


Figure 4.3 – Joint intensity distributions of MR images: (a) T1-FLAIR, (b) T1-T1c, (c) T1-T2, (d) FLAIR-T1c, (e) FLAIR-T2, (f) T1c-T2. The intensity of the first modality is read on abscissa axis and that of the second modality on the ordinate axis.

The input modality  $\{X_i, \dots, X_n\}$ , where  $n = 4$ , is first input to the independent encoders (with learning parameters  $\theta$  including the number of the filters and dropout rate) to learn the modality-specific representation  $Z_i$ . Then, a dual-attention fusion block is used. It takes the concatenation of the independent feature representations as input to produce the modality-weight and spatial-weight, respectively. And the two weights are multiplied with the input feature representation to obtain the modality-attention feature representation  $Z_{im}$  and spatial-attention feature representation  $Z_{is}$ , respectively. Finally, the learned fused feature representation is obtained by adding the modality-attention feature representation and spatial-attention feature representation.

The obtained spatial-attention feature representation  $Z_{is}$  is passed to the Correlation Description (CD) block consisting of two fully connected layers and LeakyReLU, it maps the spatial-attention feature representation  $Z_{is}$  to a set of independent parameters  $\Gamma_i = \{\alpha_i, \beta_i, \gamma_i\}$ ,  $i = 1, \dots, n$ . Finally, the correlated representation of  $i$  modality  $F_i$  can be obtained via correlation expression (Equation 4.1).

$$F_i = \alpha_i \odot Z_{is}^2 + \beta_i \odot Z_{is} + \gamma_i \quad (4.1)$$

It is noted that the nonlinear correlation expression we proposed in this work is specific to our work. However, the proposed correlation description block can be generally integrated to any multi-source correlation problem, and the specific correlation expression will depend on the application. In addition, we compare and discuss why the simplest linear correlation expression is not good for this work in Section 4.6.1.

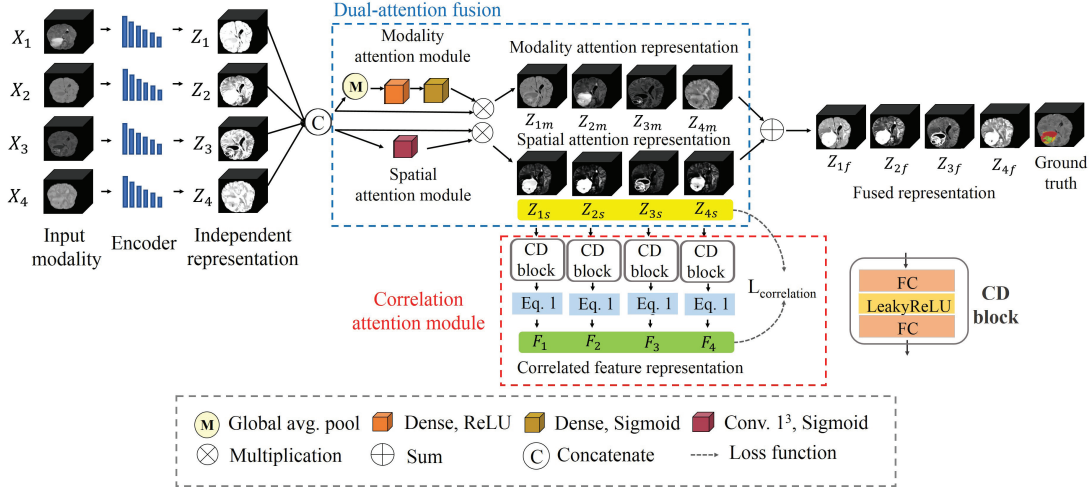


Figure 4.4 – Architecture of the tri-attention fusion strategy. The individual feature representations ( $Z_1, Z_2, Z_3, Z_4$ ) are first concatenated, then they are re-weighted by dual-attention fusion block along modality attention module and spatial attention module to achieve the modality attention representation  $Z_{im}$  and spatial attention representation  $Z_{is}$ . In addition, the correlation attention module is used to constrain the spatial-attention representations to learn segmentation-related representation. Finally, the  $Z_{im}$  and  $Z_{is}$  are added to obtain the fused feature representation  $Z_{if}$ .

To measure the divergence between the estimated correlated feature representation of  $i$  modality and the spatial-attention feature representation of  $j$  modality, there are some available f-divergences: Kullback–Leibler (KL) divergence, Jeffreys divergence, squared Hellinger divergence and exponential divergence. In our case, we propose to use a simple and widely used divergence, Kullback–Leibler divergence, to form the correlation loss function (Equation 4.2). It enables the segmentation network to learn the latent correlated features which are more relevant for segmentation. To make it clear, we take T1 modality ( $X_1$ ) and T1c modality ( $X_3$ ) as example. Since there exists a correlation between the two modalities, the spatial attention module is first used to obtain the two spatial-attention feature representations of T1 modality ( $Z_{1s}$ ) and T1c modality ( $Z_{3s}$ ). Then, the correlated feature representation ( $F_1$ ) of modality T1 can be obtained by CD block and Equation 4.1. Finally, the KL based correlation loss function is applied to constrain the two distributions ( $F_1$  and  $Z_{3s}$ ) to be as close as possible. It would be interesting to test in future other f-divergence functions, such as Hellinger distance, and compare the segmentation results.

$$L_{correlation} = \sum_{i=1}^n P(f_i) \log \frac{P(f_i)}{Q(g_j)} \quad (4.2)$$

where  $n$  is the number of modality,  $P(f_i)$  and  $Q(g_j)$  are probability distributions of spatial-attention feature representation of modality  $i$  and correlated feature representation of modality  $j$ , ( $i \neq j$ ), respectively.



From Figure 4.4, we can observe the characteristics of the target tumors in the four independent feature representations ( $Z_1, Z_2, Z_3, Z_4$ ) are not obvious. However, the modality attention module can stand out the different attributes of the modalities to provide complementary information. For example, the FLAIR modality ( $Z_{2m}$ ) highlights the whole tumor region and T1c modality ( $Z_{3m}$ ) stands out the tumor core region (red and green). In the spatial attention module, all the positions related to the target tumor regions are highlighted. In this way, we can discover the most relevant characteristics between modalities. Furthermore, the dual-attention fusion block, presented in Chapter 3, is completed by the correlation block to form a tri-attention fusion architecture. This architecture can be directly extended to any multi-modal (if existing a correlation relationship) fusion problem.

## 4.4 Experimental Setup

### 4.4.1 Data and Pre-processing

The dataset used in the experiments comes from BraTS 2018 dataset[7]. The training set includes 285 patients, each patient has four image modalities including T1, T1c, T2 and FLAIR. Following the challenge, four intra-tumor structures have been grouped into three mutually inclusive tumor regions: (a) WT which consists of all tumor tissues, (b) TC which consists of the ET, necrotic and non-enhancing tumor core, and (c) ET which is the enhancing tumor. The provided data have been pre-processed by organisers: co-registered to the same anatomical template, interpolated to the same resolution ( $1mm^3$ ) and skull-stripped. The ground-truth have been manually labeled by experts. We did additional pre-processing with a standard procedure. The N4ITK [139] method is used to correct the distortion of MRI data, and intensity normalization is applied to normalize each modality to a zero-mean, unit-variance space. To exploit the spatial contextual information of the image, we use 3D images, crop and resize them from  $155 \times 240 \times 240$  to  $128 \times 128 \times 128$ .

### 4.4.2 Implementation Details

Our network is implemented in Keras with a single Nvidia GPU Quadro P5000 (16GB). The models are optimized using the Adam optimizer(initial learning rate =  $5e-4$ ) with a decreasing learning rate factor 0.5 with patience of 10 epochs. To avoid over-fitting, early stopping is used when the validation loss isn't improved for 50 epoch. We randomly split the dataset into 80% training and 20% testing.

For segmentation, we use Dice loss to evaluate the overlap rate of prediction results and ground-truth.

$$L_{dice} = 1 - 2 \frac{\sum_{i=1}^C \sum_{j=1}^N p_{ij} g_{ij} + \epsilon}{\sum_{i=1}^C \sum_{j=1}^N (p_{ij} + g_{ij}) + \epsilon} \quad (4.3)$$

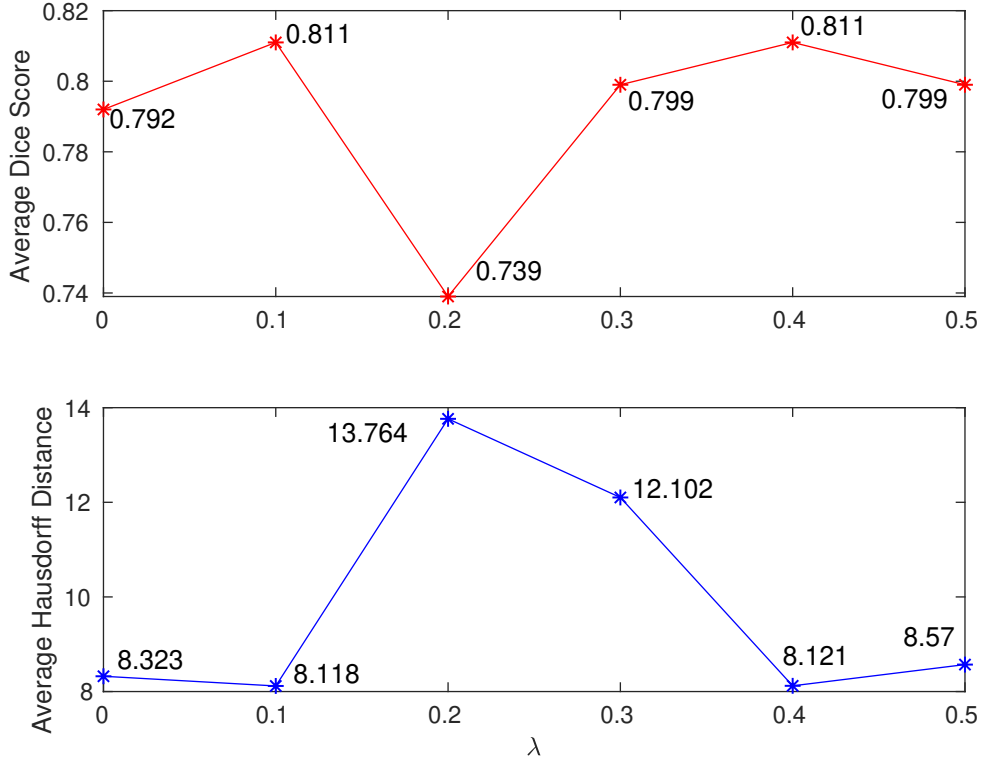


Figure 4.5 – Comparison of different weight coefficients in the loss function. Average DSC Score vs  $\lambda$  and Average HD vs  $\lambda$ .

where  $N$  is the set of examples,  $C$  is the set of the classes,  $p_{ij}$  is the probability that pixel  $i$  is of the tumor class  $j$ , the same is true for  $g_{ij}$ , and  $\epsilon$  is a small constant to avoid dividing by 0.

The network is trained by the overall loss function as follow:

$$L_{total} = L_{dice} + \lambda L_{correlation} \quad (4.4)$$

where  $\lambda$  is the trade-off parameter weighting the importance of each component. In this work, we used three most correlated MR pairs: T1-T1c, T1-T2, T2-FLAIR.

We did a grid search between  $[0, 0.5]$  to determine the optimal value for the weight coefficient  $\lambda$ , Figure 4.5 shows the comparison of average DSC Score and HD between different weight coefficients, we found that  $\lambda = 0.1$  can achieve the best segmentation results.

### 4.4.3 Evaluation Metrics

Two evaluation metrics, DSC and HD are used to evaluate the proposed method as we introduced in Chapter 3.4.3.

## 4.5 Experimental Results

We conduct a series of comparative experiments to demonstrate the effectiveness of our proposed method and compare it with other approaches. In Section 4.5.1.1, we first perform an ablation study to see the importance of our proposed components and demonstrate that adding the proposed components can enhance the segmentation performance. In Section 4.5.1.2, we compare our method with the state-of-the-art methods. In Section 4.5.2, the qualitative experiment results further demonstrate that our proposed method can achieve the promising segmentation results.

### 4.5.1 Quantitative Analysis

To prove the effectiveness of our network, we first did an ablation study to see the effectiveness of our proposed components, and then we compare our method with the state-of-the-art methods. All the results are obtained by online evaluation platform<sup>1</sup>.

#### 4.5.1.1 Ablation Study

To assess the performance of our method, and see the importance of the proposed components in our network, including dual attention fusion strategy and correlation attention module, we did an ablation study. Our network without the dual attention fusion and correlation attention module is denoted as baseline. From Table 4.1, we can observe the baseline method achieves DSC of 0.726, 0.867, 0.764 for enhancing tumor, whole tumor, tumor core, respectively. When the dual attention fusion strategy is applied to the network, we can see an increase of DSC, HD across all tumor regions with an average improvement of 0.76% and 6.44% compared to the baseline, respectively. The major reason is that the proposed fusion block can help to emphasize the most important representations from the different modalities across different positions in order to boost the segmentation result. In addition, another advantage of our method is using the correlation attention module in the lowest layer, which can constrain the encoders to discover the latent multi-source correlation representation between modalities and then guide the network to learn correlated representation to achieve a better segmentation. From the results, we can observe that with the assistance of correlation attention module, the network can achieve the best DSC of 0.75, 0.887 and 0.796, HD of 7.687, 8.306, 8.362 for enhancing tumor, whole tumor, tumor core, respectively with an average improvement of 3.18% and 8.75% relating to the baseline. Besides, we compare our method with the proposed method in Chapter 3 on BraTS 2017 training dataset, the results are presented in Table 4.2. It can be observed that the new proposed method can achieve similar

---

<sup>1</sup><https://ipp.cbica.upenn.edu/>

## CHAPTER 4. FUSION BASED ON FEATURE CORRELATION IN LATENT SPACE FOR MULTI-MODAL BRAIN TUMOR SEGMENTATION

average DSC across all the tumor regions while better average HD with an improvement of 16.94%. In addition, the new proposed method can be trained in end-to-end fashion, which is easier implemented than the proposed three-stage network in Chapter 3. Also, the proposed multi-source correlation can be extended to other applications such as helping recover the lost information in segmentation with missing modalities, which will be presented in next chapter. Moreover, we visualized the box plots of the DSC and HD for the three compared methods in Figure 4.6 and Figure 4.7. It can be observed that our proposed method not only has a higher accuracy but also a smaller standard deviation than the other two compared methods in the terms of DSC and HD. The results in Table 4.1, Figure 4.6 and Figure 4.7 demonstrate the effectiveness of each proposed component and our proposed network architecture can perform well on brain tumor segmentation.

Table 4.1 – Evaluation of our proposed method on BraTS 2018 training dataset, (1) Baseline (2) Baseline + Dual attention fusion (3) Baseline + Tri-attention fusion, ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC				HD			
	ET	WT	TC	Avg	ET	WT	TC	Avg
(1)	0.726	0.867	0.764	0.786	8.743	8.463	9.482	8.896
(2)	0.733*	0.879*	0.765	0.792	8.003*	<b>7.813*</b>	9.153	8.323
(3)	<b>0.75*</b>	<b>0.887</b>	<b>0.796*</b>	<b>0.811</b>	<b>7.687*</b>	8.306	<b>8.362*</b>	<b>8.118</b>

Table 4.2 – Evaluation of our proposed method on BraTS 2017 training dataset, (1) Proposed method in Chapter 3, (2) Proposed method in Chapter 4, ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC				HD			
	ET	WT	TC	Avg	ET	WT	TC	Avg
(1)	0.73	0.894	0.816	0.813	7.68	5.73	6.79	6.73
(2)	0.722	<b>0.896</b>	0.814	0.811	<b>5.58*</b>	<b>5.26</b>	<b>5.94*</b>	<b>5.59</b>

### 4.5.1.2 Comparison with the State-of-the-art Methods

We compare our proposed method with the state-of-the-art methods on BraTS 2018 online validation set, which contains 66 images of patients with hidden ground-truth. We first

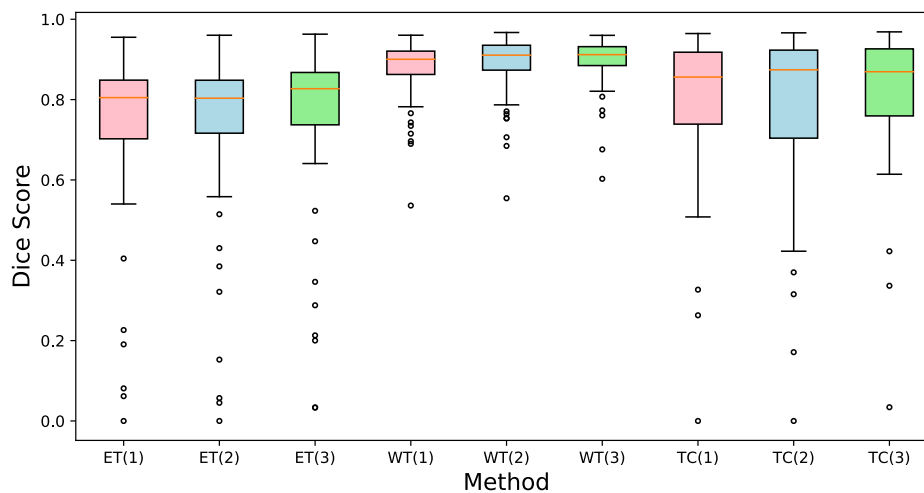


Figure 4.6 – Box plots of DSC for the three compared methods in Table 4.1 with regard to the three tumor regions: ET, WT and TC. Method (1) is shown in pink, method (2) in blue and method (3) in green.

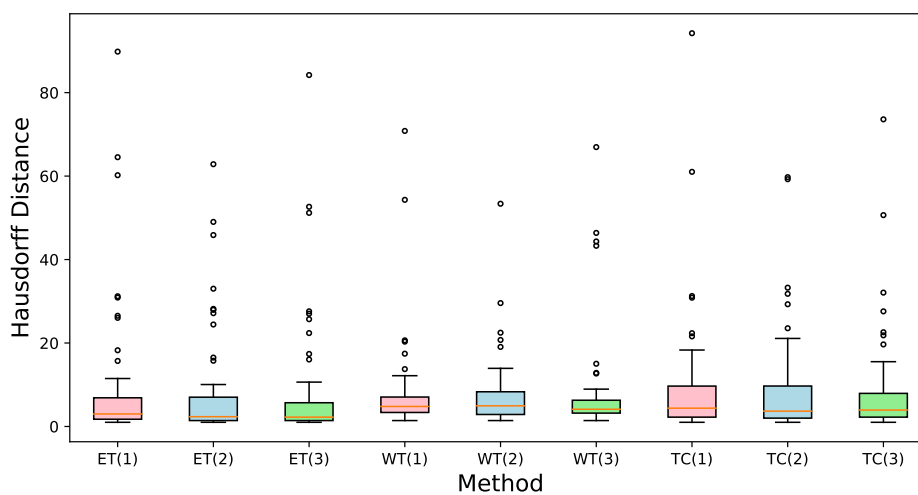


Figure 4.7 – Box plots of Hausdorff Distance for the three compared methods in Table 4.1 with regard to the three tumor regions: ET, WT and TC. Method (1) is shown in pink, method (2) in blue and method (3) in green.

predict the segmentation results on local machine and then submit them on the online evaluation platform<sup>2</sup> to obtain the evaluation results. Table 4.3 shows the comparison results. The experiment results of methods [95] and [94] are cited from [152]. We also did a computational complexity comparison between these state-of-the-art methods, including the data dimension, input size, number of network layers, number of initial convolution filter, data augmentation, post-processing, used GPU and training time, shown in Table 4.4.

(1) Zhao et al. [95] proposed to integrate FCNs and CRFs in a unified framework, where three segmentation models using 2D image patches and slices are trained in axial, coronal and sagittal views, respectively, and they are combined to segment brain tumors using a voting based fusion strategy.

(2) Kamnitsas et al. [94] introduced a dual pathway 3D convolutional neural network to incorporate both local and larger contextual information for brain tumor segmentation. In addition, they used a 3D fully connected CRF as the post-processing to remove the false positives.

(3) Hu et al. [153] proposed the Multi-level Up-sampling Network (MU-Net) for automated segmentation of brain tumors, where a novel Global Attention (GA) module is used to combine the low level feature maps obtained by the encoder and high level feature maps obtained by the decoder.

(4) Gates et al. [154] applied a multi-scale convolutional neural network based on the DeepMedic [94] to segment brain tumor.

(5) Tuan et al. [155] proposed using Bit-plane to generate a series of binary images by determining significant bits. Then, the first U-Net used the significant bits to segment the tumor boundary, and the other U-Net utilized the original images and images with least significant bits to predict the label of all pixel inside the boundary.

(6) Hu et al. [156] introduced the 3D residual Unet for brain tumor segmentation which consists of a context aggregation pathway and a localization pathway.

(7) Myronenko et al. [99] proposed a 3D MRI brain tumor segmentation using autoencoder regularization, where a variational autoencoder branch is added to reconstruct the input image itself in order to regularize the shared decoder and impose additional constraints on its layers.

From Table 4.3, we first observe that the U-Net based network [153, 155, 156, 99] can achieve better results than the CNN based network [95, 94, 154]. We explain that the skip connections in the U-Net can combine the high-level semantic feature maps from the decoder and corresponding low-level detailed feature maps from the encoder, which allows the network to learn more useful feature information to improve the segmentation. In addition, the best result in BraTS 2018 Challenge is from [99], which achieves 0.814, 0.904 and 0.859 in terms of DSC on enhancing tumor, whole tumor and tumor core regions, respectively. However, from Table 4.4, we can observe that it uses 32 initial convolution filters and a lot of memories (NVIDIA Tesla V100 32GB GPU is required) to train the model, which is computationally expensive. While our method used only 8 initial filters, a 16GB GPU is sufficient to conduct our experiments, and our network uses less training

---

<sup>2</sup><https://ipp.cbica.upenn.edu/>

CHAPTER 4. FUSION BASED ON FEATURE CORRELATION IN LATENT SPACE  
FOR MULTI-MODAL BRAIN TUMOR SEGMENTATION

Table 4.3 – Comparison of different methods on BraTS 2018 validation dataset, ET, WT, TC denote enhancing tumor, whole tumor, tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results, underline results denote the second best results.

Methods	DSC				HD			
	ET	WT	TC	Avg	ET	WT	TC	Avg
[95]	0.62	0.84	0.73	0.715	-	-	-	-
[94]	0.629	0.847	0.67	0.715	-	-	-	-
[153]	0.69	0.88	0.74	0.77	6.69	4.76	10.67	<u>7.373</u>
[154]	0.678	0.805	0.685	0.723	14.522	14.415	20.017	16.318
[155]	0.682	0.818	0.699	0.733	7.016	9.421	12.462	9.633
[156]	<u>0.719</u>	0.856	0.769	0.781	<u>5.5</u>	10.843	<u>9.985</u>	8.776
[99]	<b>0.814</b>	<b>0.904</b>	<b>0.859</b>	<b>0.859</b>	<b>3.804</b>	<b>4.483</b>	<b>8.278</b>	<b>5.521</b>
Proposed	0.688	<u>0.876</u>	<u>0.784</u>	<u>0.783</u>	6.900	<u>6.551</u>	10.199	7.883

Table 4.4 – Computational complexity comparison of different methods including data dimension, input size, number of layer, number of initial filter, data augmentation, post-processing, GPU and training time. ”-” indicates that the information is not provided in the published paper.

Methods	Dimension	Size	Layer	Filter	Aug	Post	GPU	Time
[95]	2D	$65 \times 65, 33 \times 33$	10	48	No	Yes	2G	288h
[94]	3D	$25 \times 25 \times 25, 19 \times 19 \times 19$	11	30	Yes	Yes	2G	72h
[153]	2D	$224 \times 224$	6	64	No	No	-	-
[154]	3D	$25 \times 25 \times 25, 19 \times 19 \times 19$	11	30	No	No	6G	96h
[155]	2D	$176 \times 176$	5	64	No	No	-	-
[156]	3D	$144 \times 144 \times 144$	5	16	No	No	-	-
[99]	3D	$160 \times 192 \times 128$	4	32	Yes	No	32G	48h
Proposed	3D	$128 \times 128 \times 128$	6	8	No	No	16G	40h

time. And from Table 4.3, it can be observed that our proposed method can yield the competitive results in terms of DSC and HD across all the tumor regions. The main advantage of our method is that it takes into account the multi-source correlation in brain MRI modalities to find those relevant features to obtain good segmentation. The proposed correlation attention module is a general one which can be applied to other multi-modal fusion applications if a correlation exists between them. Furthermore, compared with other methods, [156] has better DSC and HD on enhancing tumor, while our method uses smaller input size but one more layer, and finally achieves a better average DSC on all the tumor regions with an improvement of 3.84%, and it can also obtain an average improvement of 7.5% for HD.

#### 4.5.2 Qualitative Analysis

In order to evaluate the robustness of our model, we randomly select several examples on BraTS 2018 dataset and visualize the segmentation results in Figure 4.8, and the related DSC of whole tumor is presented. From Figure 4.8, we can observe that the segmentation results are gradually improved when the proposed strategies are integrated, these comparisons indicate that the effectiveness of the proposed strategies. In addition, with all the proposed strategies, our proposed method can achieve the best results.

### 4.6 Discussion

We discuss our method from the following aspects to further demonstrate the effectiveness of our method. First, we explore and compare the different correlation expressions in the correlation description block to determine which functional form provides the best fit in Section 4.6.1. Subsequently, we analyzed the performance of correlation attention module setting in different layer of network in Section 4.6.2. Finally, we visualize the feature maps of different approaches in Section 4.6.3 to demonstrate that the proposed fusion strategy can improve the segmentation.

#### 4.6.1 Performance Analysis on Correlation Expression

Table 4.5 compares the performance between linear (Equation 4.5) and nonlinear (Equation 4.1) correlation expression for segmenting brain tumor. As we can see, the nonlinear correlation expression exhibits clear advantages over the linear correlation expression across all the tumor regions. We explained that the capability is attributed to the complex nonlinear expression, which uses more parameters to fit a feature distribution, giving a better description for the feature distributions so as to guide the network to learn more correlated feature representations for segmentation. In addition, we visualized the box plots of the DSC and HD for the two compared expressions in Figure 4.9 and Figure 4.10. From the two box plots, we can obtain the consistent conclusion that the nonlinear correlation expression can achieve not only a higher accuracy but also a smaller standard deviation than the linear one in the terms of DSC and HD.



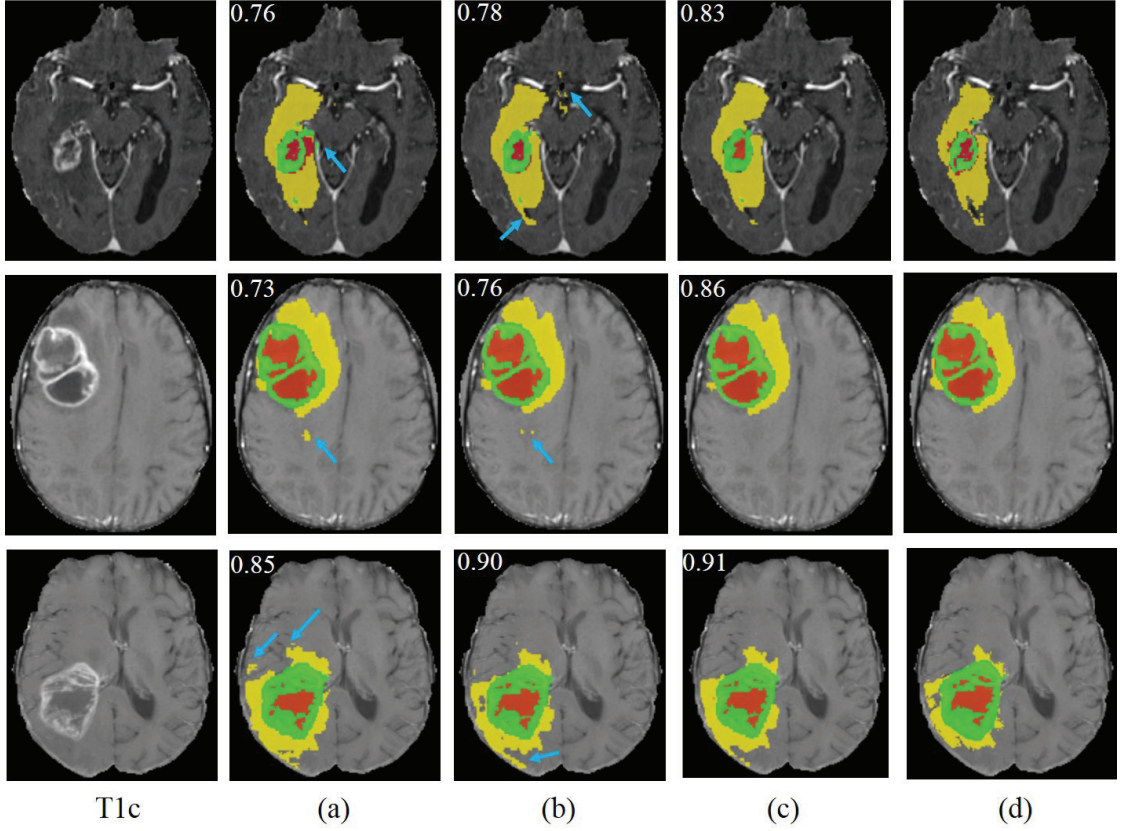


Figure 4.8 – Visualization of several segmentation results, and the related DSC of whole tumor is presented. (a) Baseline (b) Baseline with dual attention fusion (c) Baseline with tri-attention fusion (d) Ground-truth. Red: necrotic and non-enhancing tumor core; Yellow: edema; Green: enhancing tumor. Blue arrow emphasizes the mis-segmentation of the methods.

$$F_i = \alpha_i \odot Z_{is} + \gamma_i \quad (4.5)$$

#### 4.6.2 Performance Analysis on Correlation Attention Module

While experimenting with the network architectures, we have tested the addition of the correlation attention module in different layer of network. Table 4.6 shows the comparison results, (0) is our method without correlation attention module, which is used as a comparison baseline. As we can see, while setting the correlation attention module in the 4th and 6th layer can achieve the better segmentation results. Then we set the correlation attention module in both 4th and 6th layers ((7)), while the results aren't improved. Therefore, we choose to put it in the 6th layer. Then we tried to put the correlation attention module in more layers, while the correlation attention module in multi-shallower

Table 4.5 – Comparison of segmentation accuracy of different correlation expressions. Avg denotes the average results on the three tumor regions, bold results denote the best results, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC				HD			
	ET	WT	TC	Avg	ET	WT	TC	Avg
Linear	0.736	0.883	0.767	0.795	8.827	8.485	9.354	8.889
Nonlinear	<b>0.750</b>	<b>0.887</b>	<b>0.796*</b>	<b>0.811</b>	<b>7.687*</b>	<b>8.306</b>	<b>8.362*</b>	<b>8.118</b>

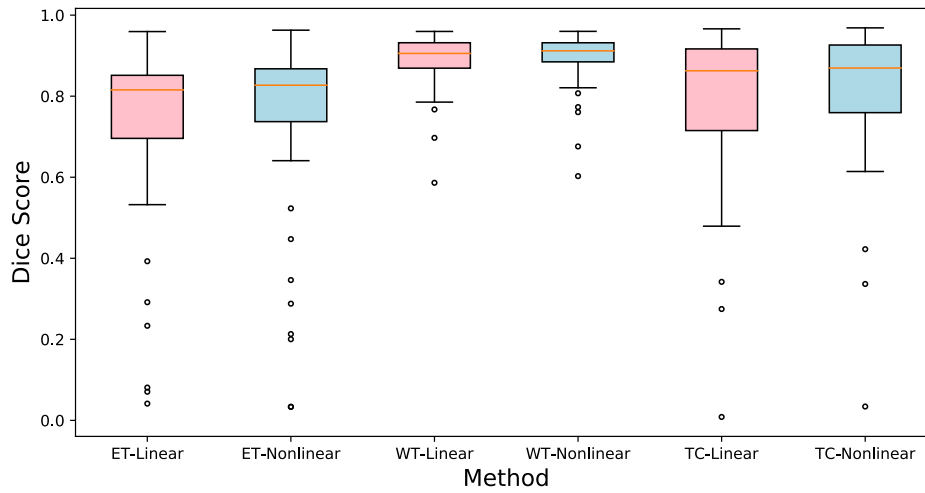


Figure 4.9 – Box plots of DSC for the two compared correlation expressions in Table 4.5 with regard to the three tumor regions: ET, WT and TC. Linear expression is shown in pink, Non-linear expression in blue.

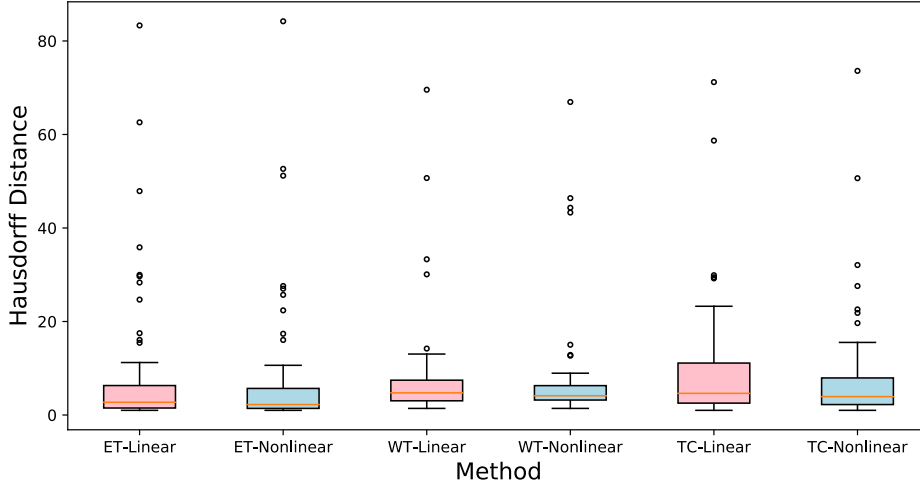


Figure 4.10 – Box plots of Hausdorff Distance for the two compared correlation expressions in Table 4.5 with regard to the three tumor regions: ET, WT and TC. Linear expression is shown in pink, Non-linear expression in blue.

layers ((8)-(12)) did not further improve the segmentation performance. We explained that since each layer represents different abstract feature representations of the inputs, where deeper levels provide more complex and abstract features, the correlation attention module can guide the most abstract feature distribution to satisfy the correlation relationship in order to improve the segmentation results.

### 4.6.3 Visualization of Feature Maps

In this section, we illustrate the advantage of our proposed correlation attention module by visualizing the feature representation maps. We select an example to show the feature representation maps of the first layer from four modalities in Figure 4.11. We denote our proposed method without any fusion strategy as baseline, the first column: input modality, the second column: baseline, third column: 'baseline + dual attention module', the fourth column: 'baseline + tri-attention module (added on the fused feature)', the fifth column: 'baseline + tri-attention module (added on the spatial attention feature)', and the sixth column: ground-truth. From Figure 4.11, we can observe that compared to the baseline, the attention mechanism (column: 3rd, 4th, 5th) allows to highlight feature representations related to brain tumor regions, especially when correlation is taken into account (column: 4th and 5th). In fact, the correlation attention module helps to enhance the fused modality-spatial feature representation for images with fewer information in the tumor region, such as T1 and T2 images.

To further investigate the contribution of the correlation attention module, we used it to guide the fused feature representations (column: 4th) and spatial-attention feature

## CHAPTER 4. FUSION BASED ON FEATURE CORRELATION IN LATENT SPACE FOR MULTI-MODAL BRAIN TUMOR SEGMENTATION

Table 4.6 – Comparison of segmentation accuracy of correlation attention module in different layer of the network. ET, WT, TC denote enhancing tumor, whole tumor and tumor core, respectively. Avg denotes the average results on the three tumor regions, bold results denote the best results.

	Methods						DSC				HD			
	1st	2nd	3rd	4th	5th	6th	ET	WT	TC	Avg	ET	WT	TC	Avg
(0)	–	–	–	–	–	–	0.733	0.879	0.765	0.792	8.003	7.813	9.153	8.323
(1)	✓	×	×	×	×	×	0.733	0.868	0.744	0.782	8.65	7.603	9.641	8.631
(2)	×	✓	×	×	×	×	0.74	0.878	0.761	0.793	7.978	8.168	9.404	8.517
(3)	×	×	✓	×	×	×	0.741	0.877	0.772	0.797	6.43	<b>6.994</b>	8.119	7.181
(4)	×	×	×	✓	×	×	<b>0.762</b>	0.886	0.776	0.808	<b>5.906</b>	7.516	<b>7.809</b>	<b>7.077</b>
(5)	×	×	×	×	✓	×	0.739	<b>0.889</b>	0.767	0.798	8.071	8.266	10.181	8.839
(6)	×	×	×	×	×	✓	0.75	0.887	<b>0.796</b>	<b>0.811</b>	7.687	8.306	8.362	8.118
(7)	×	×	×	✓	×	✓	0.754	0.886	0.778	0.806	7.677	8.206	9.18	8.354
(8)	×	×	×	×	✓	✓	0.754	0.887	0.785	0.809	7.674	7.643	8.696	8.004
(9)	×	×	×	✓	✓	✓	0.682	0.843	0.725	0.75	10.282	10.161	11.271	10.571
(10)	×	×	✓	✓	✓	✓	0.695	0.822	0.699	0.739	10.713	15.685	12.189	12.863
(11)	×	✓	✓	✓	✓	✓	0.702	0.848	0.713	0.754	9.516	9.449	10.64	9.868
(12)	✓	✓	✓	✓	✓	✓	0.536	0.724	0.406	0.555	17.102	30.667	21.359	23.043

representations (column: 5th), respectively. From Figure 4.11, we can observe that the correlation attention module added on the spatial-attention feature representations (column: 5th) can further stand out the interested tumor regions for segmentation, and the fuzzy contour becomes clear. We explained that the spatial attention module can help the network to extract the feature representations relating the tumor positions. In conclusion, the correlation attention module can constrain the network to emphasize the interested tumor region for segmentation, revealing that the addition of correlation attention module in the network encourages better segmentation results.

## 4.7 Conclusion

In this chapter, we proposed a tri-attention fusion guided 3D multi-modal brain tumor segmentation network. To take advantage of the complimentary information from different modalities, the multi-encoder based network is used to learn modality-specific feature representation. Considering the correlation between MR modalities can help the segmentation, a tri-attention fusion block is proposed, consisting of a modality attention module, a spatial attention module and a correlation attention module. The modality attention module is used to distinguish the contribution of each modality, and the spatial attention module is used to extract more useful spatial information to boost the segmentation result. Since there is a strong correlation between modalities, a correlation based constraint

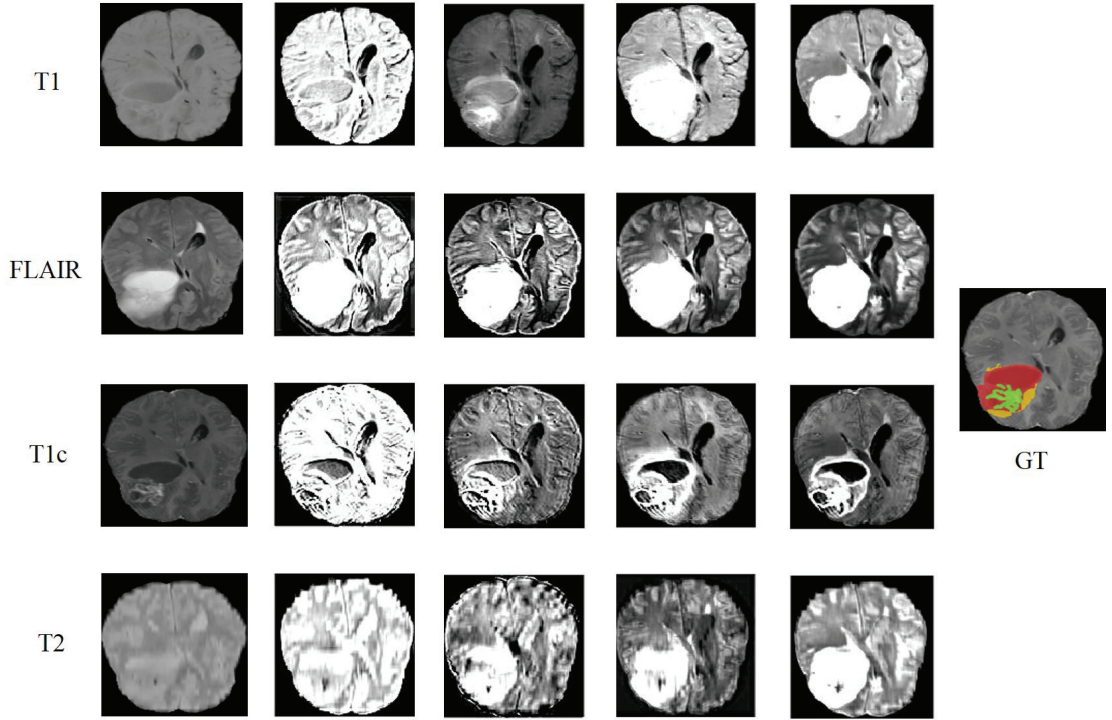


Figure 4.11 – Visualization of effectiveness of proposed correlation attention module. First column: input images, second column: baseline, third column: baseline + dual attention module, fourth column: baseline + tri-attention module (added on fused feature representation), fifth column: (Ours, added on spatial-attention feature representation), sixth column: ground-truth.

is introduced to guide the network to learn the most correlated feature representations to effectively fuse the four modalities. In conclusion, the proposed tri-attention fusion strategy utilize the complimentary information between modalities to encourage the network to learn more useful feature representation to enhance the segmentation result.

The advantages of our proposed network architecture are multiple. (1) The architecture are an end-to-end deep leaning approach and fully automatic without any user interventions. (2) The proposed correlation attention module can help the segmentation network to learn correlated feature representations to achieve very competitive results. (3) The proposed correlation attention module can be generalized to other multi-source image processing problem if some correlations exist between them. (4) The experiment results evaluated on the two metrics (DSC and HD) demonstrate that our proposed method can give the promising results for the segmentation of brain tumors and its sub-regions even small regions.

However, our work has some limitations that inspire future directions: (1) The work is only validated on multi-modal MR brain tumor images, in the future, we will valid our method in different multi-modal image datasets. (2) The proposed correlation description

block is a simple two-layer network, we intend to ameliorate the correlation description block to describe the correlation between multi modalities. It will be interesting to test in future other f-divergence functions, such as Hellinger distance. (3) It will be interesting to consider other correlation expression functions to improve the segmentation performance. (4) The proposed correlation module is applied on brain tumor segmentation, we plan to apply it to synthesize additional images to cope with the limited medical image dataset or deal with the missing modality segmentation issue. To deal with the brain tumor segmentation with missing modalities, we will present the proposed work in the next chapter.



## Chapter 5

# Multi-source Correlation Guided Brain Tumor Segmentation Network with Missing Modalities

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>68</b>
<b>5.2</b>	<b>Related Works</b>	<b>69</b>
<b>5.3</b>	<b>Methodology</b>	<b>71</b>
5.3.1	Motivation of the Proposed Method	71
5.3.2	Overview of the Method 1	71
5.3.3	Modeling the Multi-source Correlation	72
5.3.4	Overview of the Method 2	73
5.3.5	Synthesizing the Missing Modality	74
5.3.6	Brain Tumor Segmentation	77
5.3.7	Loss Functions	77
<b>5.4</b>	<b>Experimental Setup</b>	<b>78</b>
5.4.1	Dataset and Pre-processing	78
5.4.2	Implementation Details	78
5.4.3	Evaluation Metrics	78
<b>5.5</b>	<b>Experimental Results</b>	<b>78</b>
5.5.1	Quantitative Analysis	79
5.5.1.1	Ablation Study	79
5.5.1.2	Comparison with the State-of-the-art Methods	80
5.5.2	Qualitative Analysis	82
<b>5.6</b>	<b>Discussion and Conclusion</b>	<b>88</b>

---



1. **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Brain tumor segmentation with missing modalities via latent multi-source correlation representation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2020, pp. 533–541
  2. **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 4263–4274
  3. **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing MR modalities”. In: *Neurocomputing* 466 (2021), pp. 102–112
- 

## 5.1 Introduction

Different MRI modalities can highlight different sub-regions, which can provide the complimentary information to analyze the brain tumor. To make it clear, we visualize a case from BraTS 2018 dataset in Figure 5.1. It can be observed that the FLAIR highlights the whole tumor region, and T1c provides more information about tumor core. Intuitively, it’s possible to obtain the best segmentation result using full modalities. However, it’s common to have one or more missing modalities in clinical practice, due to motion artefacts, scan corruption and limited scan time. In this chapter, we will address the problem of brain tumor segmentation with incomplete modalities.

We propose to use the multi-source correlation to address the brain tumor segmentation with missing MR modalities. To this end, two methods are proposed. The conference version of the work has been presented in 2020 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) [142]. The journal versions have been presented in IEEE Transactions on Image Processing (TIP) [157] and Neurocomputing [158]. The main contributions of our work are:

- A new method based on attention mechanism with correlation representation is proposed to segment brain tumor in the case of missing MR modalities. In this method, the network is trained with complete modalities, and tested with missing modalities, in which the missing modalities are replaced by the most similar ones.
- To improve the first method, a novel feature-enhanced generation and multi-modality fusion based deep neural network is proposed. First, an auto-encoder based generator is introduced to generate the feature-enhanced missing modality. Then, a correlation constraint block is proposed to exploit the multi-source correlation on the new complete modalities. Finally, a multi-encoder based U-Net with the correlation constraint of multi-modalities is proposed to do the final segmentation.
- To exploit the multi-source correlation between MR modalities, we first designed a dedicated Correlation Parameter Estimation Module (CPEM) to learn the correlated

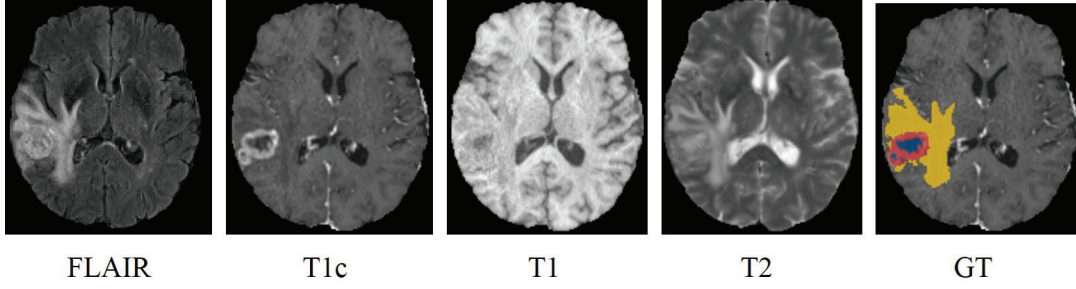


Figure 5.1 – A training sample from BraTS 2018 dataset. From left to right: FLAIR, contrast enhanced T1-weighted (T1c), T1-weighted (T1), T2-weighted (T2) images, and the ground-truth (GT). Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor.

weight parameters for each modality. And then we introduced a Linear Correlation Expression Module (LCEM) to form the correlated representation for each modality. Finally, a novel Correlation Constraint Loss (CCL) is employed to ensure that each modality satisfies the multi-source correlation with other modalities.

- The proposed correlation constraint block can guide, on the one hand, the generator to synthesize the feature-enhanced modality via satisfying the multi-source correlation with other available modalities; On the other hand, the following segmentation network to learn useful feature representations so as to achieve a better segmentation performance.
- The experimental results evaluated on the public multi-modal brain tumor segmentation dataset demonstrated the effectiveness of each proposed components, and the proposed methods can obtain the significant improvements compared with the baseline methods and state-of-the-art methods.

The rest of the chapter is organized as follows: Section 5.2 reviews the previous work, Section 5.3 offers an overview of this work and details of our network architecture. Section 5.4 describes experimental setup. Section 5.5 presents the experimental results. Section 5.6 concludes this work.

## 5.2 Related Works

Brain tumor segmentation in MRI plays a major role in the community of medical science, which has many applications in neurology such as quantitative analysis, operational planning, and functional imaging [159]. However, it's still a challenging task due to some limitations, such as the complex brain anatomy structure, various shapes, the texture of gliomas, and the low contrast of MR images [129, 94]. To address these issues, a multitude of brain tumor segmentation approaches have been proposed in the last decades. It can be generally categorized into two groups, conventional approach and deep learning based

approach. There exists some successful conventional approaches, such as Gaussian copula based Bayesian method [29], kernel feature selection [43], belief function based fusion strategy [123], random forests [124], conditional random fields [125] and support vector machines [126, 44]. Although these methods achieved good performance, they usually have small number of parameters that are insufficient to capture the complex features of brain tumor.

Recently, as a powerful alternative for feature learning, deep learning based approaches have attracted much attention in the field of brain tumor segmentation. For example, Chen et al. [147] proposed a novel deep convolutional symmetric neural network, which combines the symmetry prior knowledge into brain tumor segmentation. Ding et al. [160] proposed a multi-path adaptive fusion network to enhance entire feature hierarchy for multi-modal brain tumor segmentation. Zhou et al. [161] proposed a light-weight model, One-pass Multi-task Network (OM-Net) for brain tumor segmentation.

The approaches mentioned above require the complete set of the modalities. However, the full modalities are not always available in clinical practice. It is highly desirable to design an automatic brain tumor segmentation approach to tackle with the missing modalities problem. Recently, many related studies have emerged. We generally classify them into three groups: (1) taking all possible combinations of the modalities into account and then training several models for each situation, while this method requires a large amount of data and it is time-consuming. (2) fusing the available modalities in a latent space to learn a shared feature representation, then project it to the segmentation space [162, 163, 164, 165]. This approach is more efficient than the first approach, since it doesn't need to learn a number of possible subsets of the multi-modalities. The first proposed method belongs to this category. However, it can't generate the missing modalities. (3) synthesizing missing modalities and then use the complete modalities to do the segmentation. This method can not only synthesize the missing modality but also obtain the segmentation results. GAN [101, 166, 167] is an effective approach for image synthesis. However, training a 3D GAN is highly unstable and difficult to converge. In addition, it is difficult to put the generation of a 3D image by GAN and the image segmentation in a same architecture. In the literature [168], these two tasks are performed separately. In the second proposed method, we include both tasks in the same framework that allows to optimize the generation and segmentation.

Currently, the approaches based on exploiting latent feature representation for missing modalities become prominent. The early network architecture designed for missing modalities is from HeMIS [162]. Independent feature maps are first extracted by independent convolutional network for each modality. Then, they are fused via computing the mean and variance for the final segmentation prediction. Similarly, Lau et al. [163] proposed a U-Net based network named Unified Representation Network (URN) to combine the independent features by calculating the mean to obtain the unified representation for the final segmentation. To further enhance the modality-invariance of latent representations, Chartsias et al. [164] proposed to minimize the L1 or L2 distance of features from different modalities. Since different MRI modalities have different intensity distributions, using arithmetic operations, such as mean and variance or simply encouraging the features

from different modalities to be close under L1 or L2 distance, could not guarantee the network can learn a shared latent representation. To this end, Chen et al. [169] introduced the feature disentanglement to tackle the missing data problem. Dorent et al. [165] proposed a Hetero-Modal Variational Encoder-Decoder (U-HVED) network to use multi-modal variational auto-encoders to embed all available modalities into a shared latent representation, and the experimental results demonstrated that it can outperform HeMIS. Furthermore, Shen et al. [170] designed a domain adaptation model to adapt feature maps from missing modalities to the one from full modalities. Hu et al. [171] employed the generalized knowledge distillation to transfer knowledge from a multi-modal segmentation network to a mono-modal one to achieve the brain tumor segmentation. Zhu et al. [172] proposed a cascade supplement module to first generate shared features for missing modalities and then use squeeze and excitation to fuse the generated features and real features to achieve the segmentation.

## 5.3 Methodology

### 5.3.1 Motivation of the Proposed Method

In clinical diagnosis and treatment planning, the patient undergoes multimodal MRI scans since each modality can provide the specific information. Our hypothesis is that there is some correlation among multimodal MRIs in the tumor regions. We take an example from BraTS 2018 dataset to present joint intensity distributions of the MR images. From Figure 4.3, we can observe that there is a strong correlation in intensity distribution between each pair of modalities. To this end, it's reasonable to assume that a strong correlation also exists in the latent representation between modalities [29]. Therefore, we introduce a Correlation Constraint (CC) block to discover the multi-source correlation between modalities. In general, the proposed CC block (similar to that in the Chapter 4) can be adapted to other computer vision fields. However, the proposed method will depend on the specific image modalities and the relationship among them.

### 5.3.2 Overview of the Method 1

The first proposed network is presented in Figure 5.2. It first takes the four modalities as inputs in each encoder. The independent encoders can not only learn modality-specific feature representation, but also can avoid the false-adaptation between modalities. To take into account the strong correlation between multi modalities, the Correlation Model (CM) block is developed to discover the correlation between modalities. Then, the correlation representations across modalities are fused via attention mechanism (presented in Chapter 3), named Fusion, to emphasize the most discriminative representation for segmentation. Finally, the fused latent representation is decoded to form the final segmentation result. At the same time, the four reconstruction decoders are applied to provide more supervision to improve the segmentation.

In this method, we use the complete modalities to train the network. Since we have four modalities, we can learn four correlations from the complete modalities. For the test,

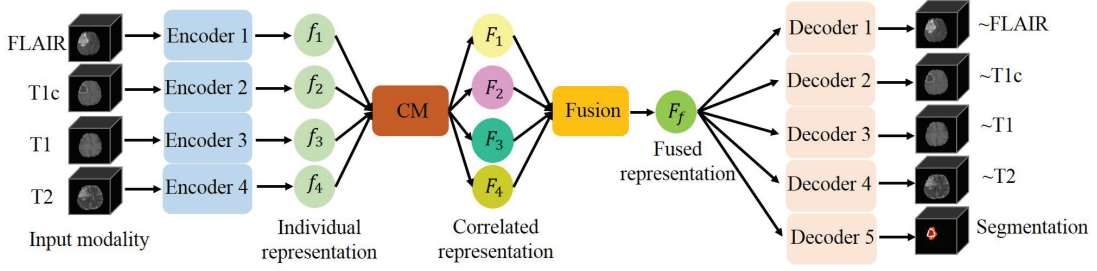


Figure 5.2 – A schematic overview of the proposed network. Each input modality is encoded by individual encoder to obtain the individual representation. The proposed Correlation Model (CM) block and Fusion block (Fusion) project the individual representations into a fused representation, which is finally decoded to form the reconstruction modalities and the segmentation result.

we replace the missing modalities by the most similar ones to always have four inputs for the trained model. In this way, the missing feature representations can be approximately recovered from the learned correlation expression with the available modalities.

### 5.3.3 Modeling the Multi-source Correlation

The proposed CM consists of two modules: Model Parameter Estimation Module (MPE Module) and Linear Correlation Expression Module (LCE Module). Each input modality  $\{X_i\}$ , where  $i = \{1, 2, 3, 4\}$ , is first input to the independent encoder to learn the modality-specific representation  $f_i(X_i|\theta_i)$ , where  $\theta_i$  denotes the parameters used in  $i$ th encoder, such as the number of filters, the kernel size of filter and the rate of dropout, which aid the encoder to obtain the most discriminative representation. Then, MPE Module, a network with two fully connected layers and Leaky ReLU, maps the modality-specific representation  $f_i(X_i|\theta_i)$  to a set of correlation parameters  $\Gamma_i = \{\alpha_i, \beta_i, \gamma_i, \delta_i\}$ , which is unique for each modality. Finally, the correlation representation  $F_i(X_i|\theta_i)$  can be obtained via LCE Module (Equation 5.1). It is noted that the number of latent multi-source correlation representation equals to the number of modalities, and the differences among the correlation representations are the weights in each correlation representation, which are related to the individual modalities.

$$F_i(X_i|\theta_i) = \alpha_i \odot f_j(X_j|\theta_j) + \beta_i \odot f_k(X_k|\theta_k) + \gamma_i \odot f_l(X_l|\theta_l) + \delta_i, (i \neq j \neq k \neq l) \quad (5.1)$$

where  $X$  is the input modality,  $i, j, k$  and  $l$  are the indexes of the modality, and  $i, j, k, l = \{1, 2, 3, 4\}$ ,  $\theta$  is the network parameters,  $f$  is the independent feature representation,  $F$  is the correlated feature representation,  $\alpha, \beta, \gamma$  and  $\delta$  are the correlation weight parameters.

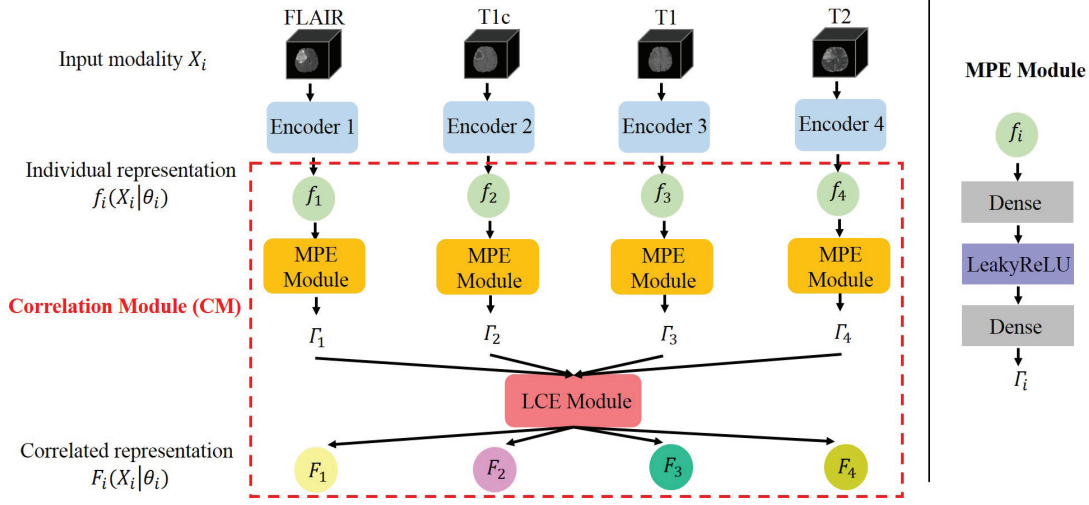


Figure 5.3 – Architecture of correlation model. MPE Module first maps the individual representation  $f_i(X_i|\theta_i)$  to a set of correlation parameters  $\Gamma_i$ , under these parameters, LCE Module transforms all the individual representations to form a latent multi-source correlation representation  $F_i(X_i|\theta_i)$ .

### 5.3.4 Overview of the Method 2

In the first method, we only replaced missing modalities by the existing ones, which leads to unsatisfactory results when more modalities are missing. Also, it did not take into account if the estimated correlated feature representation and the original feature representation are similar. Therefore, to improve the first method, we developed a new network to handle these problems. The overview of the proposed method 2 is depicted in Figure 5.4. First, the feature-enhanced generator takes the available modalities as inputs and generates a feature-enhanced modality  $X_5$ , which is an average of the missing modalities, to form the new complete modalities. At the endpoints of the encoders, the feature representations of each modality are extracted, while the feature representation of  $X_5$  is extracted by an independent encoder, which has the same architecture as the encoder of the feature-enhanced generator. Then, a Correlation Constraint (CC) block is used to exploit the multi-source correlation on the new complete set of the modalities. Finally, a segmentation decoder is applied to do the final segmentation. To train this network, we randomly zeros out any number of the modality to make them as missing modalities. Therefore, there are always four modalities as inputs, and five individual representations for the CC block. In this way, the network can adapt to the missing modalities and to maintain sufficient performance when modalities are missing during test time.

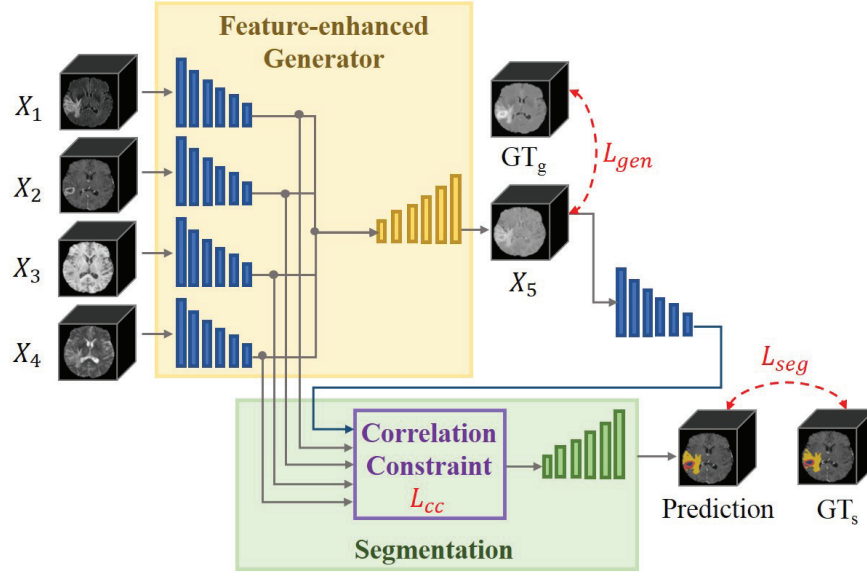


Figure 5.4 – An overview of the proposed network, consisting of a Feature-enhanced Generator (FeG), a Correlation Constraint (CC) block and a segmentation network. The feature-enhanced generator utilizes the available modalities to generate a feature-enhanced missing modality  $X_5$ . Then the complete modalities are input to the segmentation network for the final prediction. In addition, the correlation constraint block are used to guide both the feature-enhanced generator and segmentation network via exploiting the multi-source correlation.

### 5.3.5 Synthesizing the Missing Modality

There are different ways of synthesizing missing modalities. GAN [101] Generative Adversarial Network (GAN) has demonstrated to be a promising approach for image synthesis. However, mode collapse, non-convergence, instability, and highly sensibility to hyper-parameters make it difficult for training. In addition, compared to a 2D image, it is more difficult to generate a 3D image volume. And our goal is to segment brain tumor using the available modalities, not to synthesize. Therefore, we designed a special Feature-enhanced Generator, called FeG, which used an average image of the missing modalities as the ground-truth. It is noted that synthesizing the missignng modalities is only carried out in method 2 in order to improve the method 1. Here, we use the average image as the ground-truth for two reasons: (i) The averaging operation is simple to realize. And in this case, one decoder is sufficient to cope with any number of missing modalities. No additional decoders are required for each missing modality; (ii) Since the features from different modalities are correlated in one single patient, the averaging operation can extract the overall feature representations from multi-modalities, providing necessary feature representations for the following segmentation network. Specifically, we adopted a multi-encoder based network to generate the missing modalities. The architecture is shown in Figure 5.5. Each encoder starts with block 1 in the first level, which consists

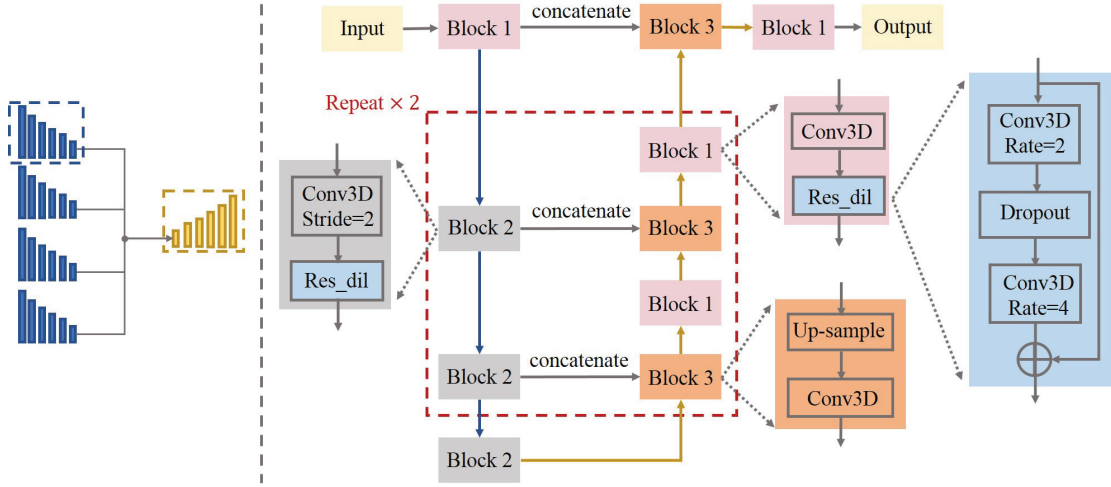


Figure 5.5 – Architecture of the Feature-enhanced Generator (FeG). Left: multi-encoder feature-enhanced generator shown in Figure 5.4; Right: we take one encoder and the decoder as an example. The left series of blocks connected by the blue arrows represent the encoder, and right series of blocks connected by the yellow arrows represent the decoder. In order to maintain the spatial information, we replace the pooling operation with the convolution block with stride=2. The res\_dil block used in both encoder and decoder is to increase the receptive field to capture more features, and the rate denotes the dilated rate.

of a 3D convolution layer and followed by a res\_dil block (see Figure 3.3 in Chapter 3). Then, block 2 is applied in the following levels, which consists of a convolutional block with  $stride = 2$  and a res\_dil block. The decoder begins with block 3, which consists of a up-sampling layer and a 3D convolution layer. Then the up-sampled features are concatenated with the features from the corresponding level of the encoders. Following the concatenation, block 1 is used to first adjust the number of features, and then enlarge the receptive field. It is noted that only the generated feature-enhanced modality needs to train an encoder to get the independent feature representation, the other encoders are directly taken from the feature-enhanced generator.

The generator and the segmentation network share the same encoders for the available modalities. There are three advantages of that: (i) The sharing operation can simplify the network architecture and reduce the training parameters. (ii) The feature-enhanced generator can learn the tumor related feature information from the segmentation network, making the interested regions obvious in the generated image volume. (iii) The segmentation network can utilize the generated feature-enhanced modality to improve the segmentation performance.

In the second method, the Correlation Constraint (CC) block (see Figure 5.6) has the similar architecture as CM in the first method (see Figure 5.3). Since a missing modality is generated, there are five individual representations as inputs for CC block, and the correlated representation can be obtained by Equation 5.2. Compared to CM of the



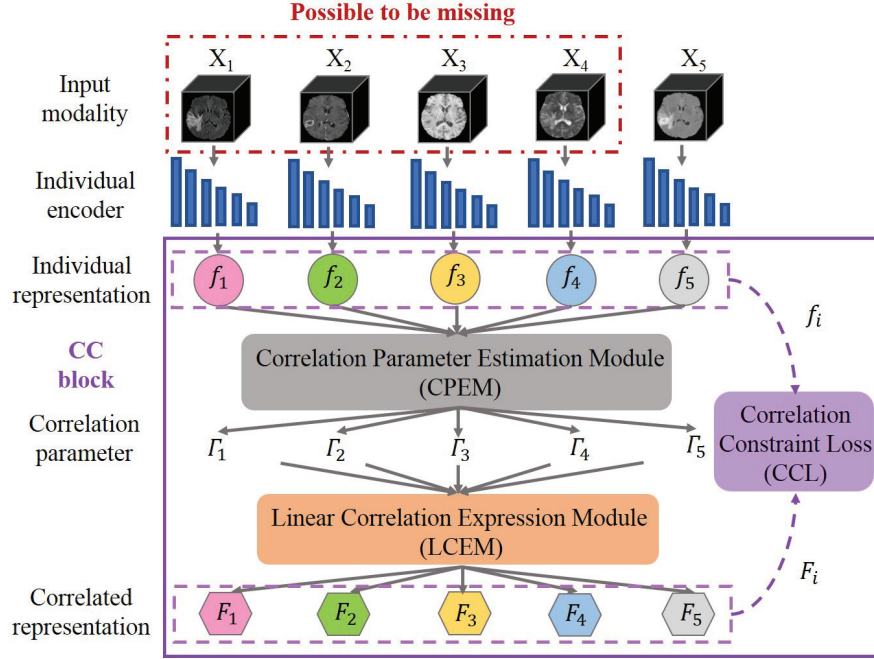


Figure 5.6 – Architecture of the Correlation Constraint (CC) block.  $X_1, X_2, X_3, X_4$  are the input modalities, which are possible to be missing.  $X_5$  is the generated average modality. CPEM first maps the individual representation  $f_i$  to a set of independent parameters  $\Gamma_i$ , under these parameters, LCEM transforms all the individual representations to form correlated representation  $F_i$ . In addition, the KL based CCL is employed to guide the whole training process.

method 1, there are two improvements in CC block. The first one is that, in CM of the method 1, each individual representation is fed separately to the MPE Module to learn the correlation parameters. To further enhance the correlation between modalities, in CC block, CPEM takes the concatenation of the five individual representations as input. The second one is that a Correlation Constraint Loss (CCL) of the method 2 is newly introduced to CC block, a Kullback–Leibler divergence based loss function (Equation 5.3). It can constrain the distributions between the estimated correlated feature representation and the original feature representation to be as close as possible. On the one hand, the CC block can provide the constraint information for the feature-enhanced generator. On the other hand, it can guide the segmentation network to learn the effective feature representations. Specifically, if the synthesized modality is not well generated, then the correlation between the synthesized modality and the available ones will be weak. Thus, through the loss function, the network iteratively generates an increasingly satisfactory feature modality with a high correlation with available modalities.

$$F_i(X_i|\theta_i) = \alpha_i \odot f_j(X_j|\theta_j) + \beta_i \odot f_k(X_k|\theta_k) + \gamma_i \odot f_l(X_l|\theta_l) + \delta_i \odot f_m(X_m|\theta_m) + \sigma_i, (i \neq j \neq k \neq l \neq m) \quad (5.2)$$

where  $X$  is the input modality,  $i, j, k, l$  and  $m$  are the indexes of the modality, and  $i, j, k, l, m = \{1, 2, 3, 4, 5\}$ ,  $\theta$  is the network parameters,  $f$  is the independent feature representation,  $F$  is the correlated feature representation,  $\alpha, \beta, \gamma, \delta$  and  $\sigma$  are the correlation weight parameters.

$$L_{cc} = \sum_{i=1}^n P(f_i) \log \frac{P(f_i)}{Q(g_i)} \quad (5.3)$$

where  $n$  is the number of modality,  $P(f_i)$  and  $Q(g_i)$  are probability distributions of original feature representation and correlated feature representation of modality  $i$ , respectively.

### 5.3.6 Brain Tumor Segmentation

The architecture of segmentation encoders and decoders in the two proposed methods are adopted from our previous work [119]. The multiple encoders are used to extract the independent feature representations for each modality. To emphasize the most important features from different modalities, a fusion block based on attention mechanism is introduced in each level of the network (presented in Chapter 3). It allows to selectively emphasize feature representations along spatial-wise and modality-wise. Moreover, the deep supervision [86] is employed by integrating the segmentation results from different levels to form the final network output.

### 5.3.7 Loss Functions

In the first method, the network is trained by the overall loss function:

$$L_{total} = L_{seg} + L_1 \quad (5.4)$$

where  $L_{seg}$  and  $L_1$  are designed for the segmentation network and reconstruction, respectively,  $L_1$  is the mean absolute loss.

For segmentation, we use Dice loss ( $L_{seg}$ ), which is defined in Chapter 4.3, to calculate the overlap rate of prediction results and ground-truth.

In the second method, the overall loss function is defined as follow:

$$L_{total} = L_{seg} + \lambda L_{gen} + \eta L_{cc} \quad (5.5)$$

where  $L_{seg}$ ,  $L_{gen}$  and  $L_{cc}$  are designed for the segmentation network, feature-enhanced generator and the correlation constraint block, respectively.  $\lambda$  and  $\eta$  are the trade-off parameters, where  $\lambda = 0.1$  and  $\eta = 0.1$  in our work.

For the feature-enhanced generator, we use Structural Similarity Index Metric (SSIM) as the loss function.

$$L_{gen} = 1 - \sum_{i=1}^N \frac{(2\mu_{i\hat{y}}\mu_{iy} + c_1)(2\sigma_{iy\hat{y}} + c_2)}{(\mu_{i\hat{y}}^2 + \mu_{iy}^2 + c_1)(\sigma_{iy\hat{y}}^2 + \sigma_{iy}^2 + c_2)} \quad (5.6)$$

where  $N$  is the set of all examples,  $y_i$  is the real image,  $\hat{y}_i$  is the generated image,  $\mu$  and  $\sigma$  are the mean and standard deviation of the image, and  $\sigma_{y\hat{y}}$  is the co-variance between

the real image  $y$  and the generated image  $\hat{y}$ ,  $c_1$  and  $c_2$  are to stabilize the division with weak denominator.

The correlation loss  $L_{cc}$  has been introduced in Section 5.3.5, which is defined in Equation 5.3.

## 5.4 Experimental Setup

### 5.4.1 Dataset and Pre-processing

To evaluate our proposed method, we used BraTS 2018 dataset [7]. It contains a training set including 285 training cases with ground-truth, and a validation set including 66 cases with hidden ground-truth. Each case has four image modalities including T1, T1c, T2 and FLAIR. Following the challenge, four intra-tumor structures (edema, enhancing tumor, necrotic and non-enhancing tumor core) have been grouped into three mutually inclusive tumor regions: (a) The WT, consisting of all tumor tissues. (b) The TC, consisting of the enhancing tumor, necrotic and non-enhancing tumor core. (c) The ET, the enhancing tumor.

The provided data have been pre-processed by organisers: co-registered to the same anatomical template, interpolated to the same resolution ( $1mm^3$ ) and skull-stripped. The ground-truth have been manually labeled by experts. We did additional pre-processing with a standard procedure. To exploit the spatial contextual information of the image, we used 3D image, we cropped and resized them from  $155 \times 240 \times 240$  to  $128 \times 128 \times 128$ . The N4ITK [139] method is used to do the bias field correction for MRI data, and intensity normalization is applied to normalize each modality to a zero-mean, unit-variance space.

### 5.4.2 Implementation Details

The proposed network is implemented using Keras with a single Nvidia Tesla V100 (32G). The model is trained using Nadam optimizer, the initial learning rate is 0.0005, it will reduce with a factor 0.5 with patience of 10 epochs for the first method and 5 epochs for the second method. To avoid over-fitting, early stopping is used if the validation loss is not improved over 10 epochs. We randomly split the dataset into 80% training and 20% testing. All the results are obtained by online evaluation platform<sup>1</sup>.

### 5.4.3 Evaluation Metrics

To obtain quantitative measurements of the segmentation accuracy, we used two commonly used evaluation metrics: DSC and HD as we introduced in Chapter 3.4.3

## 5.5 Experimental Results

We carry out a series of comparative experiments to demonstrate the effectiveness of our proposed methods. In Section 5.5.1.1, we illustrate the advantages of the proposed

---

<sup>1</sup><https://ipp.cbica.upenn.edu/>

components. In Section 5.5.1.2, we compare our methods with the state-of-the-art approaches. In Section 5.5.2, we conduct the qualitative experiment to further demonstrate that our proposed method can obtain the promising segmentation results.

### 5.5.1 Quantitative Analysis

#### 5.5.1.1 Ablation Study

**The first method** To prove the importance of the proposed components in our network, including reconstruction decoders and CM block, we conduct an ablation study on full modalities, and the results are shown in Table 5.1. We denote the original network without reconstruction decoders and CM as baseline. From Table 5.1, we can observe that the baseline method achieves average DSC and average HD of 77.4% and 8.3, respectively. When the reconstruction decoders are integrated to the network, we can see an increase in the terms of DSC on whole tumor and enhancing tumor. In addition, when the CM block is added, the average DSC and HD are further improved, which improves the baseline with 1.7% in the terms of average DSC and 14.5% in the terms of average HD.

Table 5.1 – Comparison of different strategies in our proposed method 1 on full modalities on BraTS 2018 training set,  $\uparrow$  denotes the improvement compared to the previous method, bold results show the best scores for each tumor region, AVG denotes the average results on the three target regions, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Methods	DSC (%)				HD			
	WT	TC	ET	AVG	WT	TC	ET	AVG
Baseline	87.8	76.5	67.8 $\uparrow$	77.4	7.8	9.2	8.0	8.3
Baseline + Reconstruction	87.9 $\uparrow$	76.2	68.1 $\uparrow$	77.4	8.5	9.7	8.3	8.8
Baseline + Reconstruction + CM (Ours)	<b>88.2<math>\uparrow</math></b>	78.6* $\uparrow$	69.4* $\uparrow$	78.7 $\uparrow$	6.7* $\uparrow$	7.6* $\uparrow$	<b>7.1<math>\uparrow</math></b>	7.1

**The second method** We also carry out an ablation study to investigate the relative contribution of the proposed components, including the Feature-enhanced Generator (FeG) and the CC block. The comparison results are presented in Table 5.2 and Table 5.3. The baseline denotes our method without FeG and CC block. From Table 5.2, we can observe that the baseline can obtain 79.6%, 67.1%, 49.4% and 65.4% in the terms of DSC on whole tumor, tumor core, enhancing tumor and the average result across all the situations, respectively. The proposed feature-enhanced generator can indeed improve the segmentation results compared to the baseline, with an improvement of 2.3%, 3.7%, 6.3% and 3.7% in terms of DSC on whole tumor, tumor core, enhancing tumor and the average result across all the situations, respectively. We explain that the proposed feature-enhanced generator can compensate the missing data and help the network to improve the segmentation performance.

It can also be observed that with the assistance of CC block, the segmentation results have significant improvements in all the cases. And statistically significant differences (an improvement of 15.5% compared to the 'Baseline + FeG') can be observed when only

## CHAPTER 5. MULTI-SOURCE CORRELATION GUIDED BRAIN TUMOR SEGMENTATION NETWORK WITH MISSING MODALITIES

FLAIR and T2 are available. The comparison results demonstrate that the multi-source correlation constraint block can guide the segmentation network to focus on the related target feature representations so as to improve the segmentation performance.

The similar comparison results in terms of HD can be observed in Table 5.3. It is noticed that, the average HD for 'Baseline', 'Baseline + FeG' and ours are 10.0, 8.7 and 7.1, respectively. We can observe an improvement of 13.0% by the FeG and 18.4% by the CC block, respectively, which indicates that the proposed additional generator and the correlation constrain block are really helpful for the segmentation. In conclusion, the comparison results demonstrate the effectiveness of the proposed components.

Table 5.2 – Comparison of different strategies in our proposed method 2 in terms of DSC (%) on BraTS 2018 dataset, • denotes the present modality and ◦ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Modality				Baseline				Baseline + FeG				Baseline + FeG + CC (Ours)			
F	T1	T1c	T2	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG
◦	◦	◦	•	76.5	50.3	21.4	49.4	78.5*	52.6*	25.5*	52.2	<b>80.4</b>	<b>59.5*</b>	<b>35.2*</b>	<b>58.4</b>
◦	◦	•	◦	65.6	78.1	70.5	71.4	68.4*	79.6*	71.6*	73.2	<b>72.0*</b>	<b>83.1*</b>	<b>75.0*</b>	<b>76.7</b>
◦	•	◦	◦	66.2	39.0	17.5	40.9	69.8*	47.5*	16.6	44.6	<b>74.9*</b>	<b>55.5*</b>	<b>29.2*</b>	<b>53.2</b>
•	◦	◦	◦	84.2	50.0	16.3	50.2	85.3*	53.6*	25.8*	54.9	<b>85.9*</b>	<b>64.6*</b>	<b>39.5*</b>	<b>63.3</b>
◦	◦	•	•	78.2	80.6	71.2	76.7	81.0*	83.3*	74.7*	79.7	<b>81.7</b>	<b>84.8*</b>	<b>75.5*</b>	<b>80.7</b>
◦	•	•	◦	70.0	79.0	71.0	73.3	74.3*	81.4*	73.9*	76.5	<b>76.2*</b>	<b>84.2*</b>	<b>75.8*</b>	<b>78.7</b>
•	•	◦	◦	84.4	53.8	21.1	53.1	85.8*	57.6*	28.5*	57.3	<b>86.5*</b>	<b>67.0*</b>	<b>43.2*</b>	<b>65.6</b>
◦	•	◦	•	77.9	51.5	25.6	51.7	80.8*	54.1*	27.8	54.2	<b>83.4*</b>	<b>62.6*</b>	<b>38.0*</b>	<b>61.3</b>
•	◦	◦	•	84.1	54.7	24.2	54.3	85.8*	56.2	30.0*	57.3	<b>86.5*</b>	<b>66.6*</b>	<b>45.5*</b>	<b>66.2</b>
•	◦	•	◦	84.7	80.8	72.4	79.3	85.5*	84.2*	75.9*	81.9	<b>86.2*</b>	<b>85.0*</b>	<b>77.1*</b>	<b>82.8</b>
•	•	•	◦	85.7	82.9	75.6	81.4	85.8	84.2	76.6*	82.2	<b>86.6*</b>	<b>85.6*</b>	<b>77.2*</b>	<b>83.1</b>
•	•	◦	•	85.5	58.4	32.0	58.6	86.2*	58.6	32.2	59.0	<b>86.8*</b>	<b>68.0*</b>	<b>45.6*</b>	<b>66.8</b>
•	◦	•	•	85.9	84.3	75.4	81.9	85.9	85.0*	76.5*	82.5	<b>86.4</b>	<b>86.0*</b>	<b>76.9*</b>	<b>83.1</b>
◦	•	•	•	80.9	81.9	74.3	79.0	82.1*	82.5	75.0	79.9	<b>82.9*</b>	<b>85.2*</b>	<b>76.2*</b>	<b>81.4</b>
•	•	•	•	84.7	81.4	72.1	79.4	85.9*	84.3*	76.7*	82.3	<b>86.6*</b>	<b>85.8*</b>	<b>76.9*</b>	<b>83.1</b>
Average				79.6	67.1	49.4	65.4	81.4	69.6	52.5	67.8	<b>82.9</b>	<b>74.9</b>	<b>59.1</b>	<b>72.3</b>

### 5.5.1.2 Comparison with the State-of-the-art Methods

**The first method** We first compare our proposed method with the state-of-the-art methods on full modalities on BraTS 2018 online validation sets. From the compared results presented in Table 5.4, we can observe that the plain U-Net has an unsatisfied performance on all the tumor regions. The high HD values illustrate that the method has the large segmentation errors on all the tumor regions, while other improved U-Net based methods have much better segmentation results. Secondly, we can observe that our proposed method achieves the second best average DSC and the second best HD on tumor core. Thirdly, it can be seen that the best method is from [99], which achieves 90.4%, 85.9% and 81.4% in terms of DSC on whole tumor, tumor core and enhancing tumor

## CHAPTER 5. MULTI-SOURCE CORRELATION GUIDED BRAIN TUMOR SEGMENTATION NETWORK WITH MISSING MODALITIES

Table 5.3 – Comparison of different strategies in our proposed method 2 in terms of HD on BraTS 2018 dataset, • denotes the present modality and ◦ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

				Baseline				Baseline + FeG				Baseline + FeG + CC (Ours)			
F	T1	T1c	T2	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG
◦	◦	◦	•	13.3	16.1	15.1	14.8	10.6	15.0	14.7	13.4	<b>9.0*</b>	<b>12.2*</b>	<b>11.9*</b>	<b>11.0</b>
◦	◦	•	◦	15.9	10.5	8.0	11.5	12.1*	8.9	6.2	9.1	<b>10.8*</b>	<b>6.5*</b>	<b>4.5*</b>	<b>7.3</b>
◦	•	◦	◦	15.7	23.9	22.5	20.7	14.3	18.3*	18.2*	16.9	<b>10.0*</b>	<b>15.2*</b>	<b>13.4*</b>	<b>12.9</b>
•	◦	◦	◦	5.9	13.5	13.3	10.9	6.2	13.3	12.2*	10.6	<b>5.4*</b>	<b>10.9*</b>	<b>9.7*</b>	<b>8.7</b>
◦	◦	•	•	7.6	6.1	4.8	6.2	7.0*	5.3	4.0*	5.4	<b>6.6</b>	<b>5.1</b>	<b>3.9*</b>	<b>5.2</b>
◦	•	•	◦	14.0	11.5	9.9	11.8	9.3*	7.9	5.0	7.4	<b>8.7</b>	<b>5.4*</b>	<b>4.2*</b>	<b>6.1</b>
•	•	◦	◦	6.3	12.7	12.3	10.4	5.7	12.8	12.0	10.2	<b>5.2*</b>	<b>10.4*</b>	<b>9.4*</b>	<b>8.3</b>
◦	•	◦	•	10.7	17.3	16.6	14.9	8.5*	14.7*	12.8*	12.0	<b>6.6*</b>	<b>10.3*</b>	<b>11.3*</b>	<b>9.4</b>
•	◦	◦	•	5.9	12.3	11.7	10.0	5.7	12.7	12.0	10.1	<b>5.2</b>	<b>10.2*</b>	<b>9.2*</b>	<b>8.2</b>
•	◦	•	◦	5.8	5.0	3.8	4.9	5.4	5.5	4.3	5.1	<b>5.0</b>	<b>4.6*</b>	<b>3.3*</b>	<b>4.3</b>
•	•	•	◦	6.3	5.4	3.5	5.1	5.3	5.6	4.1	5.0	<b>5.0</b>	<b>4.2*</b>	<b>3.0*</b>	<b>4.1</b>
•	•	◦	•	6.4	12.8	11.4	10.2	5.4	12.6	11.3	9.8	<b>5.2</b>	<b>9.0*</b>	<b>8.9*</b>	<b>7.7</b>
•	◦	•	•	6.9	4.6	3.6	5.0	5.5*	5.5	4.2	5.1	<b>5.2</b>	<b>4.1*</b>	<b>3.2*</b>	<b>4.2</b>
◦	•	•	•	9.7	8.1	6.8	8.2	<b>6.3*</b>	5.4	3.9*	5.2	6.5	<b>4.9*</b>	<b>3.5*</b>	<b>5.0</b>
•	•	•	•	<b>4.7</b>	5.7	4.2	4.9	5.3	5.6*	4.0*	5.0	5.2	<b>4.2*</b>	<b>3.0*</b>	<b>4.1</b>
Average				9.0	11.0	9.8	10.0	7.5	9.9	8.6	8.7	<b>6.6</b>	<b>7.8</b>	<b>6.8</b>	<b>7.1</b>

regions, respectively. However, it can only segment the brain tumors on full modalities, while our method is trained to do the segmentation when modalities are missing. In addition, the method [99] fed a very large patch size ( $160 \text{ voxels} \times 192 \text{ voxels} \times 128 \text{ voxels}$ ) into the network and used 32 initial convolution filters. Also, it required a 32GB GPU to train the model, which is computationally expensive (introduced in Chapter 4.5.1.2). In contrast, our method used only 8 initial filters and exhibited a reasonable performance. A 16GB GPU is sufficient to conduct our experiments.

Then, to demonstrate the effectiveness of the proposed CM block and evaluate the robustness of our proposed method on missing modalities, we compare it with WoCM, a specific case of our method without CM. As illustrated in Table 5.5, compared with WoCM, we can observe that the CM block improves the segmentation results on average DSC of 5.8%, 15.3%, 15.6% for whole, core and enhancing tumor, respectively. It demonstrates the importance of the proposed CM block.

**The second method** To demonstrate the advantages of our proposed method on missing modalities, we compare it with the state-of-the-art methods, which have been introduced in the related work. The comparison results are illustrated in Table 5.6. Since the method HeMIS didn't publish the available code, the reported results on HeMIS and U-HeMIS [162] are taken from the work in [165]. For all the tumor regions, our method achieves the best results in most of the cases. Compared to the current state-of-the-art method [165], our method can achieve the average DSC of 82.9%, 74.9% and 59.1% across

CHAPTER 5. MULTI-SOURCE CORRELATION GUIDED BRAIN TUMOR  
SEGMENTATION NETWORK WITH MISSING MODALITIES

Table 5.4 – Comparison of different methods on full modalities on BraTS 2018 validation set, bold results show the best scores, and underline results refer the second best results.

Methods	DSC (%)				HD			
	WT	TC	ET	AVG	WT	TC	ET	AVG
Ronneberger et al. [53]	27.6	9.0	3.7	13.4	53.2	134.1	187.9	125.1
Tuan et al. [155]	81.8	69.9	68.2	73.3	9.4	12.4	7.0	9.6
Hu et al. [153]	88.0	74.0	69.0	77.0	<u>4.7</u>	10.6	<u>6.6</u>	<u>7.3</u>
Myronenko et al. [99]	<b>90.4</b>	<b>85.9</b>	<b>81.4</b>	<b>85.9</b>	<b>4.4</b>	<b>8.2</b>	<b>3.8</b>	<b>5.5</b>
Ours (Method 1)	<u>87.1</u>	<u>78.3</u>	<u>70.8</u>	<u>78.7</u>	6.5	<u>9.9</u>	7.1	7.8

all the situations, which outperforms the best method by 3.5%, 17% and 18.2%. We explain that the proposed feature-enhanced generator compensates the missing modality with the correlation constraint block. It indicated the importance of the complete dataset, since it can provide the full information for the network to learn the effective features for segmentation. We also compared with the method trained on missing-one modality [172]. Our method can obtain much better results with a large margin, we can achieve the improvement of 6.7%, 4.4%, 7.4% and 12.6% in the terms of average DSC when T2, T1c, T1, Flair is missing, respectively. And for the full modalities, we can also outperform it by 6.3% in the terms of average DSC. In addition, we compare with our first proposed method, a large margin of improvement can be seen across all the missing cases, which demonstrates that the effectiveness of the improvements proposed in the second method.

From the comparison results, we can also obtain another observation. Missing FLAIR modality leads to a sharp decreasing on DSC for all the regions, since FLAIR is the principle modality for showing whole tumor. While missing T1c modality would have a severe decreasing on DSC for both tumor core and enhancing tumor, since T1c is the principle modality for showing tumor core and enhancing tumor regions. Missing T1 and T2 modalities would have a slight decreasing on DSC for all the regions. Furthermore, when only one modality is available, FLAIR modality can achieve the promising results. When two modalities are available, 'FLAIR + T1c' are the best combination, indicating that the importance of FLAIR and T1c for MR brain tumor segmentation. Our method can give better results in these cases than other methods.

### 5.5.2 Qualitative Analysis

**The first method** In order to evaluate the robustness of our model, we randomly select several examples from BraTS 2018 and visualize the segmentation results of different methods on full modalities in Figure 5.7. From Figure 5.7, we can observe that the segmentation results are gradually improved when the proposed strategies are integrated, these comparisons indicate that the effectiveness of the proposed strategies. In addition, with all the proposed strategies, our proposed method can achieve the best results.

The segmentation results on missing modalities are presented in Figure 5.8. As shown in Figure 5.8, we can observe that with the increasing number of missing modalities, the

CHAPTER 5. MULTI-SOURCE CORRELATION GUIDED BRAIN TUMOR  
SEGMENTATION NETWORK WITH MISSING MODALITIES

Table 5.5 – Comparison of different strategies in our proposed method 1 in the terms of DSC (%) on BraTS 2018 dataset,  $\circ$  denotes the missing modality and  $\bullet$  denotes the present modality, WoCM denotes our method without CM, bold results denote the best score, Average denotes the average results on one target region across all the situations, \* denotes the significant improvement evaluated via Wilcoxon signed-rank test ( $p < 0.05$ ).

Modalities				WT		TC		ET	
F	T1	T1c	T2	WoCM	Ours	WoCM	Ours	WoCM	Ours
$\circ$	$\circ$	$\circ$	$\bullet$	31.4	<b>33.0</b>	14.9	<b>15.9</b>	6.2	<b>7.2</b>
$\circ$	$\circ$	$\bullet$	$\circ$	29.7	<b>33.6</b>	49.3	<b>56.1</b>	50.0	<b>53.5</b>
$\circ$	$\bullet$	$\circ$	$\circ$	3.3	<b>5.6</b>	4.3	<b>6.3*</b>	4.5	<b>5.3*</b>
$\bullet$	$\circ$	$\circ$	$\circ$	71.4	<b>73.7*</b>	46.2	<b>48.6*</b>	5.0	<b>25.8*</b>
$\circ$	$\circ$	$\bullet$	$\bullet$	45.1	<b>48.3*</b>	48.1	<b>50.4*</b>	52.0	<b>52.4*</b>
$\circ$	$\bullet$	$\bullet$	$\circ$	11.4	<b>29.2*</b>	22.6	<b>55.0*</b>	24.8	<b>54.8*</b>
$\bullet$	$\bullet$	$\circ$	$\circ$	75.9	<b>80.4*</b>	47.4	<b>51.5*</b>	7.7	<b>10.2*</b>
$\circ$	$\bullet$	$\circ$	$\bullet$	31.6	<b>35.5*</b>	12.9	<b>14.3*</b>	2.5	<b>6.1*</b>
$\bullet$	$\circ$	$\circ$	$\bullet$	80.4	<b>81.3</b>	20.7	<b>25.2*</b>	9.3	<b>10.0*</b>
$\bullet$	$\circ$	$\bullet$	$\circ$	80.3	<b>81.5</b>	65.7	<b>73.4*</b>	62.7	<b>67.5*</b>
$\bullet$	$\bullet$	$\bullet$	$\circ$	81.1	<b>82.7</b>	71.7	<b>75.8*</b>	65.7	<b>68.4*</b>
$\bullet$	$\bullet$	$\circ$	$\bullet$	83.5	<b>85.4*</b>	41.3	<b>44.4*</b>	11.1	<b>12.9*</b>
$\bullet$	$\circ$	$\bullet$	$\bullet$	87.5	<b>87.9</b>	74.2	<b>77.5*</b>	65.4	<b>67.2*</b>
$\circ$	$\bullet$	$\bullet$	$\bullet$	46.9	<b>50.1*</b>	51.2	<b>52.1</b>	54.3	<b>54.8</b>
$\bullet$	$\bullet$	$\bullet$	$\bullet$	87.9	<b>88.2</b>	76.2	<b>78.6*</b>	68.1	<b>69.4*</b>
Average				56.5	<b>59.8</b>	43.1	<b>48.3</b>	32.6	<b>37.7</b>

segmentation results produced by our robust model just slightly degrade, rather than a sudden sharp degrading. In addition, only using FLAIR modality, the proposed method can generate a good segmentation of whole tumor. And with FLAIR and T1c modalities, it can yield a competitive segmentation result compared to the ground truth, T1 and T2 modalities can help to refine the boundary area of the tumor regions, which has the consistent conclusion with the quantitative results.

**The second method** To further demonstrate the performance of our proposed method, we randomly select an example on BraTS 2018 dataset and visualize the segmentation and generation results in Figure 5.9 and Figure 5.10, respectively. In Figure 5.9, the first row shows the four MR modalities. The last three rows show the segmentation results. Here, 'Ours' denotes the method 'Baseline + FeG + CC'. From Figure 5.9, we can observe that in each column, with the help of our proposed components, the segmentation results can be gradually improved. Especially when only FLAIR is available, the baseline method produces many false predictions on the tumor core region. When the FeG is applied,



Table 5.6 – Comparison of different methods in terms of DSC (%) on BraTS 2018 dataset, • denotes the present modality and ◦ denotes the missing one, bold results denotes the best scores. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively. AVG denotes the average results on the three target regions, Average denotes the average results on one target region across all the situations.

Modality			HeMIS [162]				U-HeMIS [162]				URN [163]				U-HVED [165]				[172]				The first method (Ours)				The second method (Ours)			
F	T1	T1c	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG	WT	TC	ET	AVG
◦	◦	•	38.6	19.5	0.0	19.4	79.2	50.0	23.3	50.8	77.5	43.6	20.3	47.1	80.9	54.1	30.8	55.3	-	-	-	-	33	15.9	7.2	18.7	80.4	59.5	35.2	58.4
◦	◦	◦	2.6	6.5	11.1	6.7	58.5	58.5	60.8	59.3	62.2	58.5	55.8	58.8	62.4	66.7	65.5	64.9	-	-	-	-	33.6	56.1	53.5	47.7	72.0	83.1	75.0	76.7
◦	◦	◦	0.0	0.0	0.0	0.0	54.3	37.9	12.4	34.9	50.4	34.2	19.1	34.6	52.4	37.2	13.7	34.4	-	-	-	-	5.6	6.3	5.3	5.7	74.9	55.5	29.2	53.3
◦	◦	◦	55.2	16.2	6.6	26.0	79.9	49.8	24.9	51.5	84.8	50.4	23.6	52.9	82.1	50.4	24.8	52.4	-	-	-	-	73.7	48.6	25.8	49.4	85.9	64.6	39.5	63.3
◦	◦	•	48.2	45.8	55.8	49.9	81.0	69.1	68.6	72.9	80.3	68.9	67.6	72.3	82.7	73.7	70.2	75.5	-	-	-	-	48.3	50.4	52.4	50.4	81.7	84.8	75.5	80.7
◦	•	◦	15.4	30.4	42.6	29.5	63.8	64.0	65.3	64.4	69.8	65.9	66.5	67.4	66.8	69.7	67.0	67.8	-	-	-	-	29.2	55.0	54.8	46.3	76.2	84.2	75.8	78.7
◦	◦	◦	71.1	11.9	1.2	28.1	83.9	56.7	29.0	56.5	85.5	52.6	25.3	54.5	84.3	55.3	24.2	54.6	-	-	-	-	80.4	51.5	10.2	47.4	86.5	67.0	43.2	65.6
◦	◦	◦	47.3	17.2	0.6	21.7	80.8	53.4	28.3	54.2	80.8	48.6	25.2	51.5	82.2	57.2	30.7	56.7	-	-	-	-	35.5	14.3	6.1	18.6	83.4	62.6	38.0	61.3
◦	◦	◦	74.8	17.7	0.8	31.1	86.0	58.7	28.0	57.6	86.3	50.7	25.2	54.1	87.5	59.7	34.6	60.6	-	-	-	-	81.3	25.2	10.0	38.8	86.5	66.5	45.5	66.2
◦	◦	◦	68.4	41.4	53.8	54.5	83.3	67.6	68.0	73.0	85.8	72.5	70.4	76.2	85.8	72.9	70.3	76.2	-	-	-	-	81.5	73.4	67.5	74.1	86.2	85.0	77.1	82.8
◦	◦	◦	70.2	48.8	60.9	60.0	85.1	70.7	69.9	75.2	85.6	72.0	71.0	76.2	86.2	74.2	71.1	77.2	86.1	78.2	69.3	77.9	82.7	75.8	68.4	75.6	86.6	85.6	77.2	83.1
◦	◦	◦	75.2	18.7	1.0	31.6	87.0	61.0	33.4	60.5	86.1	52.5	25.8	54.8	88.0	61.5	34.1	61.2	87.6	62.6	41.7	64.0	85.4	44.4	12.9	47.6	86.8	80.0	45.6	66.8
◦	◦	◦	75.6	54.9	60.5	63.7	87.0	72.2	69.7	76.3	86.5	72.2	69.8	76.2	88.6	75.6	71.2	78.5	87.1	77.8	67.4	77.4	87.9	77.5	67.2	77.5	86.4	86.0	76.9	83.1
◦	◦	◦	44.2	46.6	55.1	48.6	82.1	70.7	69.7	74.2	81.1	69.5	68.5	73.0	83.3	75.3	71.1	76.6	75.6	76.0	65.3	72.3	50.1	52.1	54.8	52.3	82.9	85.2	76.2	81.4
•	•	•	73.8	55.3	61.1	63.4	87.6	73.4	70.8	77.3	86.3	71.8	69.9	76.0	88.1	76.4	71.7	79.0	88.3	77.7	68.5	78.2	88.2	78.6	69.4	78.7	86.6	85.8	76.9	83.1
Average			50.7	28.7	27.4	35.6	78.6	59.7	48.1	62.1	79.3	58.9	46.9	61.7	80.1	64.0	50.0	64.7	-	-	-	-	59.8	48.3	37.7	48.6	82.9	74.9	59.1	72.3

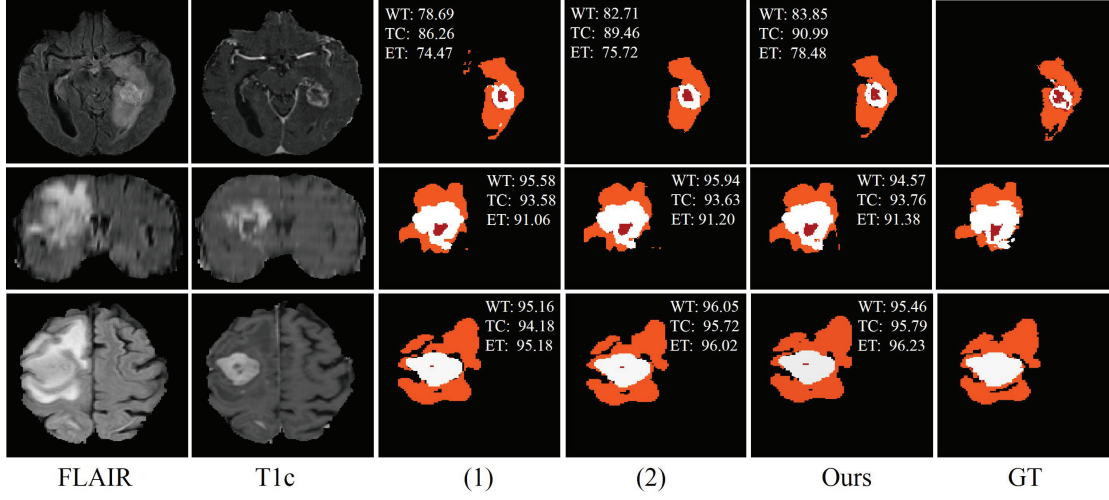


Figure 5.7 – Examples of the segmentation results on full modalities of the first proposed method. (1) denotes the baseline, (2) denotes baseline with reconstruction. Red: necrotic and non-enhancing tumor core; Orange: edema; White: enhancing tumor.

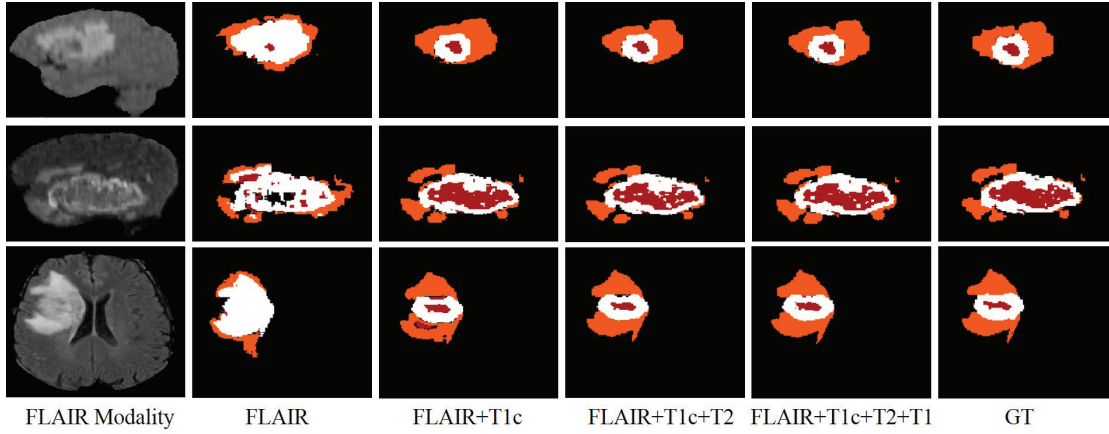


Figure 5.8 – Examples of the segmentation results on missing modalities of the first proposed method. Red: necrotic and non-enhancing tumor core; Orange: edema; White: enhancing tumor.

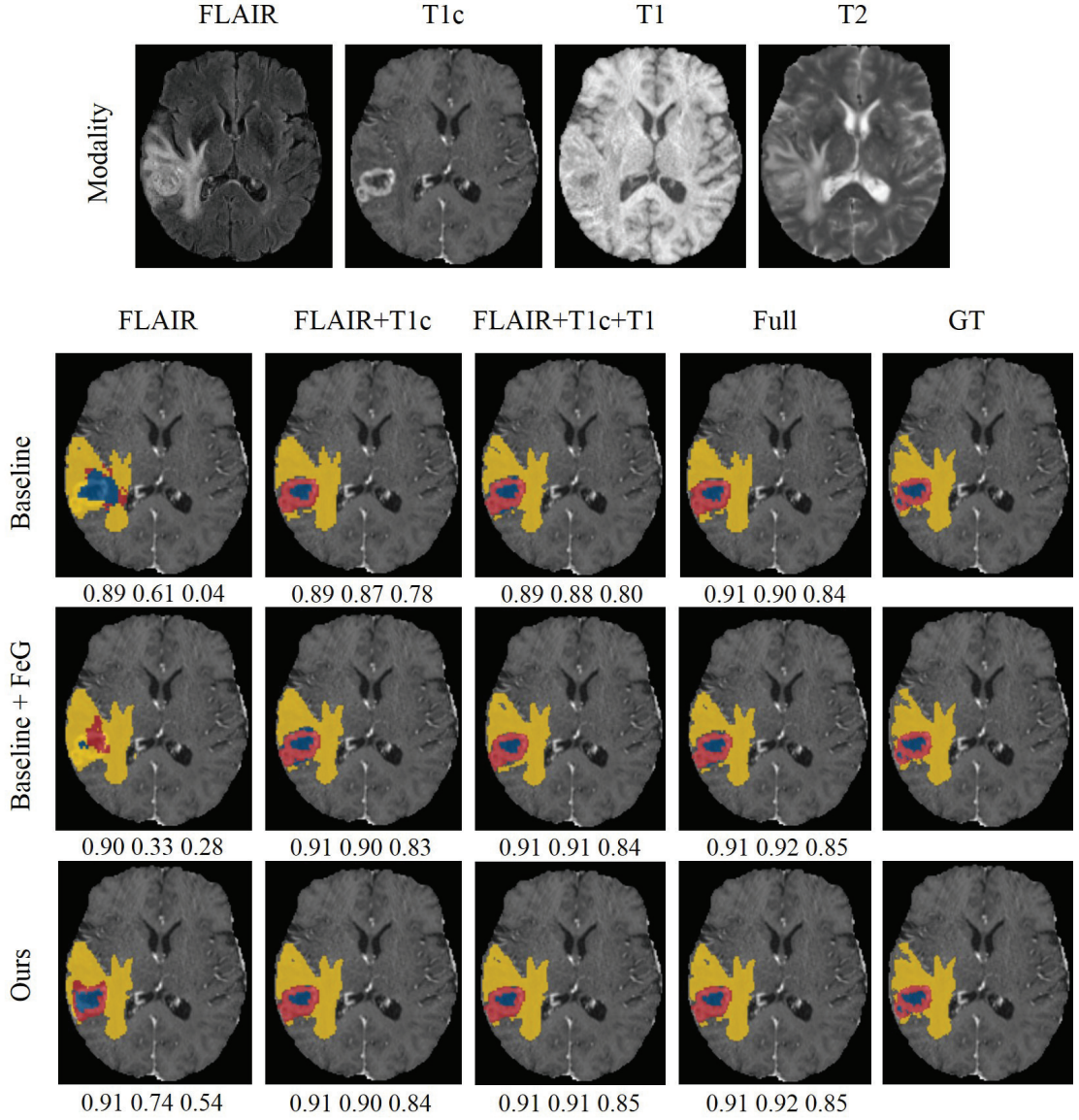


Figure 5.9 – Segmentation results of the second proposed method, the DSC of whole tumor, tumor core and enhancing tumor is denoted under each image. On the top, the first row shows the four MR modalities, FLAIR, T1c, T1 and T2. On the bottom, the three rows show the segmentation results of different methods. The first four columns show the different missing modalities situations. The last column shows the ground-truth segmentation. Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor.

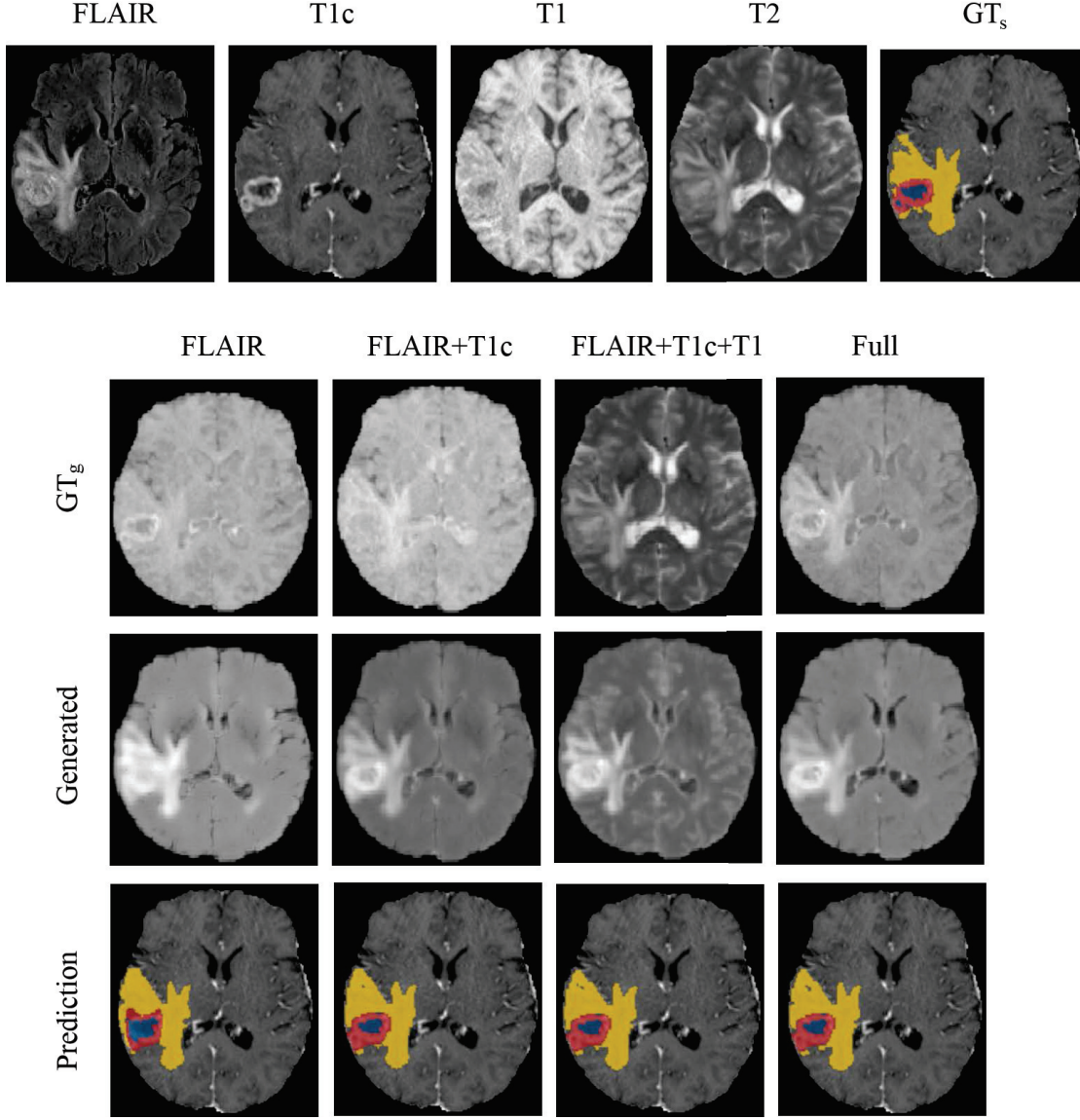


Figure 5.10 – Generation and segmentation results of the second proposed method. On the top, the first row shows the four MR modalities (FLAIR, T1c, T1 and T2) and the segmentation ground-truth. On the bottom, the four columns show the different missing modalities situations. The first row shows the ground-truth average modality. The second row shows the corresponding generated average modality. The last row shows the segmentation results of our method. Net&ncr is shown in blue, edema in yellow and enhancing tumor in red. Net refers non-enhancing tumor and ncr necrotic tumor.

these false predictions are corrected gradually. This improvement is mainly attributed to the new generated average modality, it can provide more rich information for the network to achieve more useful feature learning. Finally, with the help of the correlation constrain, we can obtain the best segmentation results. The reason for this improvement is that the effective feature representation guided by the correlation constrain can help the network to achieve the better performance. In each row, with the increase of the number of the modalities, the segmentation results become much better. It indicated that the full data is important to improve the segmentation accuracy.

To demonstrate the effectiveness of the proposed FeG, we also visualize the generated results in Figure 5.10. The first row presents the complete four MR modalities and the ground-truth. The second row shows the ground-truth average modality. The third row shows the generated average modality. The last row presents the corresponding segmentation results. We can observe that the target tumor regions are really obvious in the generated image. We explain that as the feature-enhanced generator shared the same encoders with the segmentation part, it makes the generator able to learn the related segmentation features. In addition, thanks to the generated modality, the segmentation network can utilize more data information to achieve the promising results when one or more modalities are missing.

## 5.6 Discussion and Conclusion

In this chapter, we propose two multi-source correlation based deep neural networks for brain tumor segmentation with missing MR modalities. The first proposed method replace the missing modalities by available ones, which can't retrieve the lost information when more modalities are missing. Moreover, there is not any constrain to guarantee the similarity between the correlated feature representation and original feature representation. To handle this issue and improve the segmentation results of the first method, we propose the method 2 which generates a feature-enhanced missing modality under the multi-source correlation condition. In addition, a KL is applied to guarantee the similarity between the estimated correlated feature representation and the original feature representation. We suppose that the correlation is linear. The multi-source correlation, on the one hand, can constrain the generator to generate the feature-enhanced missing modality; and on the other hand, can help the segmentation network to learn the correlated feature information to improve the segmentation performance. Finally, a segmentation decoder is used to achieve the brain tumor segmentation.

Even if most existing synthesis approaches are based on GAN. Otherwise, it is difficult to use the generation of a 3D image by GAN and to integrate it to the image segmentation architecture. In the literature, these two tasks are performed separately. In our work, we include both tasks in the same framework that allows to optimize both the generation and the segmentation. In addition, the training of GAN is challenging, such as mode collapse, non-convergence, instability, and highly sensibility to hyper-parameters. Another reason that drives us not to use GAN is that, our principle goal is to utilize the multi-source correlation to segment the brain tumor when individual modalities are missing, not to

generate a perfect missing modality. Therefore, we design a specific multi-encoder based network to generate the missing modalities.

To investigate the importance of the proposed components of our network, several comparison experiments are implemented with regard to the feature-enhanced generator and the correlation constraint block. The experimental results demonstrate the proposed feature-enhanced generator enable the network to generate a relevant feature-enhanced missing modality, and the correlation constraint block can aide the network to achieve the promising segmentation results. In addition, the comparison results on BraTS 2018 dataset show that our method can outperform the state-of-the-art methods despite missing modalities.

However, there are some limitations of the proposed method. (i) The proposed method is only validated on multi MR modalities. (ii) There are still some rooms to improve the proposed correlation constraint block. (iii) The proposed segmentation network is evaluated on one public brain tumor dataset.

To overcome the above limitations, in our future work, we would like to expand our method to other multi-modal segmentation problems, such as CT and MRI. And it would be interesting to ameliorate the proposed correlation constraint block and to improve the segmentation performance. In addition, we plan to validate our method on other public or private multi-modal segmentation datasets to investigate the robustness of our method.



## Chapter 6

# General Conclusions and Perspectives

### Contents

<b>6.1</b>	<b>Conclusions . . . . .</b>	<b>91</b>
<b>6.2</b>	<b>Perspectives . . . . .</b>	<b>92</b>

### 6.1 Conclusions

In this thesis, we presented the original deep learning methods for multi-modal brain tumor segmentation. We first gave a general introduction about the biomedical background in Chapter 1, including the MRI, brain tumor and brain tumor segmentation challenges. Then, the technical state-of-the-arts are presented in Chapter 2, including the principle of deep learning, multi-modal medical segmentation methods, data preparation and multi-modal fusion strategies. The main contributions are introduced from Chapter 3 to Chapter 5. They are summarized as follows:

- In Chapter 3, we propose a novel three-stage network based on context constraint and attention mechanism for multi-modal brain tumor segmentation. In the first stage, to decrease the influence of the fuzzy contour in the brain tumors, we used an initial segmentation network to produce a context constraint for each tumor region. In the second stage, based on the constraint, we applied a multi-encoder based network to obtain three single tumor region segmentations. In addition, the attention mechanism is introduced to achieve the fusion of different modalities. Furthermore, considering the location relationship of the tumor regions, we proposed a new loss function to cope with the multi-class segmentation problem. In the third stage, a two-encoder based 3D U-Net is presented to combine and refine the three single prediction results to form the final segmentation result. Experimental results evaluated on BraTS 2018 dataset demonstrated the effectiveness of our proposed method.



- In Chapter 4, to further improve the brain tumor segmentation, we propose to exploit the multi-source correlation between modalities in latent space and propose a tri-attention fusion guided 3D multi-modal brain tumor segmentation network. The proposed tri-attention fusion consists of the modality attention modules, the spatial attention module and the correlation attention module. Since there is a strong correlation between modalities, the correlation attention module is introduced to guide the network to learn the most correlated feature representations for segmentation. In conclusion, it utilized the complimentary information between modalities to encourage the network to learn more useful feature representation to improve the segmentation result. Extensive experimental results demonstrate that our method can achieve the promising results and outperform the state-of-the-art methods.
- In Chapter 5, we address the problem of brain tumor segmentation with missing data. We first propose a multi-source correlation based deep neural network, while in this method, we replace the missing modalities by available ones, which can't achieve very satisfactory results when more modalities are missing. To this end, we ameliorate the network framework by proposing a novel feature-enhanced generation and multi-modality fusion based deep neural network. First, an auto-encoder based generator is introduced to generate the feature-enhanced missing modality. Then, a correlation constraint block is proposed to exploit the multi-source correlation between modalities to make the network extracting principally the correlated latent features. Finally, a multi-encoder based U-Net with the correlation constraint of multi-modalities is proposed to achieve the final segmentation. The experimental results demonstrated the effectiveness of the proposed components, and the proposed method can obtain the significant improvements compared with the baseline methods and state-of-the-art methods.

In summary, in this thesis, we proposed several deep learning methods to address the multi-modal fusion problem and multi-modal brain tumor segmentation problem. The proposed fusion methods can exploit latent correlation between modalities and fuse the complementary information to improve the segmentation performance. The proposed segmentation methods can achieve promising results with complete modalities as well as missing modalities.

## 6.2 Perspectives

Regarding the limitations of the presented works in this thesis, the future works can be divided into two parts. One is to further improve the performance of the proposed methods. The other one is to tackle the new challenges in multi-modal medical image segmentation.

For the first aspect, the future work can be summarized as follows:

- The presented deep learning methods can be further improved by larger clinical datasets and also by the recent state-of-the-art network architectures. We would

like to valid our proposed methods on other medical modalities such as PET and CT, and also in other multi-modal image datasets, not limited in medical image field.

- In Chapter 3, the proposed brain tumor segmentation network consists of three stages, we plan to reduce the number of training stages and integrate them to an end-to-end training fashion. Besides, we would apply our proposed method on some other medical segmentation tasks such as organ segmentation.
- In Chapter 4, the proposed correlation description block is a simple two-layer network, we intend to explore other network architectures to describe the correlation between multi modalities. In this work, we choose a simple and widely used f-divergence function, Kullback–Leibler divergence. It will be interesting to test other f-divergence functions, such as Hellinger distance. In addition, a nonlinear correlation expression is applied in this work, in the future, we would consider other correlation expressions to better describe the correlation in the latent space.
- In Chapter 5, we designed a special Feature-enhanced generator to extract the overall feature representations from multi-modalities, where the average image of the missing modalities is used as the ground truth. There are some rooms can be improved. We intend to modify the network to generate all the missing modalities instead of the average one. In addition, it will be interesting to test GAN to generate the missing modalities and compare the segmentation results with our proposed method.

For the second aspect, the future work can be summarized as follows:

- We would like to extend our proposed multi-source correlation block to other clinical applications such as tumor recurrence location prediction. The proposed multi-source correlation block might be introduced to exploit the time correlation among the longitude images in order to predict the future tumor recurrence location.
- For multi-modality fusion, to enrich the feature representations, some state-of-the-art methods such as Transformer, the multi-head attention module can be used to further fuse the multi-modalities to improve the segmentation. For the latent feature representations, some other methods such as multiple kernel learning, manifold learning and graph neural networks can be developed to exploit the latent feature space to benefit the segmentation.
- For the segmentation with missing data, conditional GAN can be expected to supervise the generation of the missing modalities via the condition factor. In addition, transfer learning could be a solution to solve the segmentation with missing data by transferring the knowledge learned from the complete modalities to the missing ones.



# List of publications

## Journal Papers

- **Tongxue Zhou**, Su Ruan, and Stéphane Canu. “A review: Deep learning for medical image segmentation using multi-modality fusion”. In: *Array* (2019), p. 100004
- **Tongxue Zhou**, Stéphane Canu, and Su Ruan. “Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism”. In: *International Journal of Imaging Systems and Technology* 31.1 (2021), pp. 16–27
- **Tongxue Zhou**, Stéphane Canu, and Su Ruan. “Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation”. In: *Computerized Medical Imaging and Graphics* 86 (2020), p. 101811
- **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 4263–4274
- **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing MR modalities”. In: *Neurocomputing* 466 (2021), pp. 102–112
- **Tongxue Zhou**, Su Ruan, Pierre Vera, and Stéphane Canu. “A Tri-attention Fusion Guided Multi-modal Segmentation Network”. In: *Pattern Recognition* (2021), p. 108417

## Conference Papers

- **Tongxue Zhou**, Su Ruan, Haigen Hu, and Stéphane Canu. “Deep Learning Model Integrating Dilated Convolution and Deep Supervision for Brain Tumor Segmentation in Multi-parametric MRI”. in: *International Workshop on Machine Learning in Medical Imaging (MLMI)*. Springer. 2019, pp. 574–582
- **Tongxue Zhou**, Su Ruan, Yu Guo, and Stéphane Canu. “A Multi-Modality Fusion Network Based on Attention Mechanism for Brain Tumor Segmentation”. In: *2020*

- IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 377–380
- **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Brain tumor segmentation with missing modalities via latent multi-source correlation representation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2020, pp. 533–541
  - **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “3D Medical Multimodal Segmentation Network Guided by Multi-source Correlation Constraint”. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021, pp. 10243–10250
  - **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “A Dual Supervision Guided Attentional Network for Multimodal MR Brain Tumor Segmentation”. In: *International Conference on Medical Image and Computer-Aided Diagnosis (MICAD)*. 2021
  - Haigen Hu, Leizhao Shen, **Tongxue Zhou**, Pierre Decazes, Pierre Vera, and Su Ruan. “Lymphoma segmentation in PET images based on multi-view and Conv3D fusion strategy”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 1197–1200

# Bibliography

- [1] *The Physics of MRI and How We Use It to Reveal the Mysteries of the Mind*. URL: <https://kids.frontiersin.org/articles/10.3389/frym.2019.00023> (visited on 07/02/2021).
- [2] *MRI scan*. URL: <https://www.nhs.uk/conditions/mri-scan/> (visited on 07/02/2021).
- [3] *The Anatomy of MRI*. URL: <https://www.northwestradiology.com/the-anatomy-of-mri/> (visited on 07/02/2021).
- [4] *Imaging of the Visual System with MRI*. URL: <https://slideplayer.com/slide/3893127/> (visited on 07/02/2021).
- [5] Arsène Ella, David A Barrière, Hans Adriaensen, David N Palmer, Tracy R Melzer, Nadia L Mitchell, and Matthieu Keller. “The development of brain magnetic resonance approaches in large animal models for preclinical research”. In: *Animal Frontiers* 9.3 (2019), pp. 44–51.
- [6] *Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics*. URL: <https://case.edu/med/neurology/NR/MRIBasics.htm> (visited on 07/02/2021).
- [7] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge”. In: *arXiv preprint arXiv:1811.02629* (2018).
- [8] *MRI sequences*. URL: <https://www.slideshare.net/DrTusharPatil/mri-sequences> (visited on 07/02/2021).
- [9] *Perfusion MRI*. URL: <https://www.independentimaging.com/different-types-of-mris/> (visited on 07/12/2021).
- [10] *Perfusion MRI*. URL: [https://en.wikipedia.org/wiki/Perfusion\\_MRI](https://en.wikipedia.org/wiki/Perfusion_MRI) (visited on 07/12/2021).

- [11] Steffen E Petersen, Nay Aung, Mihir M Sanghvi, Filip Zemrak, Kenneth Fung, Jose Miguel Paiva, Jane M Francis, Mohammed Y Khanji, Elena Lukaschuk, Aaron M Lee, et al. “Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort”. In: *Journal of Cardiovascular Magnetic Resonance* 19.1 (2017), pp. 1–19.
- [12] Gerwin P Schmidt, Maximilian F Reiser, and Andrea Baur-Melnyk. “Whole-body imaging of the musculoskeletal system: the value of MR imaging”. In: *Skeletal radiology* 36.12 (2007), pp. 1109–1119.
- [13] Robert E Watson. “Lessons learned from MRI safety events”. In: *Current Radiology Reports* 3.10 (2015), pp. 1–7.
- [14] Girish Katti, Syeda Arshiya Ara, and Ayesha Shireen. “Magnetic resonance imaging (MRI)—A review”. In: *International journal of dental clinics* 3.1 (2011), pp. 65–70.
- [15] Ali Işın, Cem Direkçöğlü, and Melike Şah. “Review of MRI-based brain tumor image segmentation using deep learning methods”. In: *Procedia Computer Science* 102 (2016), pp. 317–324.
- [16] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. “The multimodal brain tumor image segmentation benchmark (BRATS)”. In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.
- [17] *Brain Tumor: Symptoms and Signs*. URL: <https://www.cancer.net/cancer-types/brain-tumor/symptoms-and-signs> (visited on 07/02/2021).
- [18] *Brain Tumor (Cancer)*. URL: <https://brainmadesimple.com/brain-tumor-cancer/> (visited on 07/23/2021).
- [19] *Brain Cancer: Natural Strategies to Help the Fight*. URL: <https://drjockers.com/brain-cancer/> (visited on 07/12/2021).
- [20] *Brain Tumor: Risk Factors*. URL: <https://www.cancer.net/cancer-types/brain-tumor/risk-factors> (visited on 07/23/2021).
- [21] *Brain Tumor: Diagnosis*. URL: <https://www.cancer.net/cancer-types/brain-tumor/diagnosis> (visited on 07/23/2021).
- [22] *Brain Tumor: Types of Treatment*. URL: <https://www.cancer.net/cancer-types/brain-tumor/types-treatment> (visited on 07/06/2021).
- [23] Sweet Ping Ng, Carlos E Cardenas, Hesham Elhalawani, Courtney Pollard III, Bahar Elgohari, Penny Fang, Mohamed Meheissen, Nandita Guha-Thakurta, Houda Bahig, Jason M Johnson, et al. “Comparison of tumor delineation using dual energy computed tomography versus magnetic resonance imaging in head and neck cancer re-irradiation cases”. In: *Physics and imaging in radiation oncology* 14 (2020), pp. 1–5.

- [24] Gisele C Pereira, Melanie Traugher, and Raymond F Muzic. “The role of imaging in radiation therapy planning: past, present, and future”. In: *BioMed research international* 2014 (2014).
- [25] Abdelmajid Bousselham, Omar Bouattane, Mohamed Youssfi, and Abdelhadi Raihani. “Towards reinforced brain tumor segmentation on MRI images based on temperature changes on pathologic area”. In: *International journal of biomedical imaging* 2019 (2019).
- [26] Mina Ghaffari, Arcot Sowmya, and Ruth Oliver. “Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges”. In: *IEEE reviews in biomedical engineering* 13 (2019), pp. 156–168.
- [27] Linmin Pei, Lasitha Vidyaratne, Md Monibor Rahman, and Khan M Iftekharuddin. “Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images”. In: *Scientific Reports* 10.1 (2020), pp. 1–11.
- [28] Didier Dubois and Henri Prade. “Combination of fuzzy information in the framework of possibility theory”. In: *Data fusion in robotics and machine intelligence* 12 (1992), pp. 481–505.
- [29] Jerome Lapuyade-Lahorgue, Jing-Hao Xue, and Su Ruan. “Segmenting multi-source images using hidden Markov fields with copula-based multivariate statistical distributions”. In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3187–3195.
- [30] Qinkun Xiao, Minying Qin, Peng Guo, and Yidan Zhao. “Multimodal fusion based on LSTM and a couple conditional hidden Markov model for chinese sign language recognition”. In: *IEEE Access* 7 (2019), pp. 112258–112268.
- [31] Weibei Dou, Su Ruan, Yanping Chen, Daniel Bloyet, and Jean-Marc Constans. “A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images”. In: *Image and vision Computing* 25.2 (2007), pp. 164–171.
- [32] Sudeb Das and Malay Kumar Kundu. “A neuro-fuzzy approach for medical image fusion”. In: *IEEE transactions on biomedical engineering* 60.12 (2013), pp. 3347–3353.
- [33] Pagavathigounder Balasubramaniam and VP Ananthi. “Image fusion using intuitionistic fuzzy sets”. In: *Information Fusion* 20 (2014), pp. 21–30.
- [34] Li-Wei Ko, Yi-Chen Lu, Humberto Bustince, Yu-Cheng Chang, Yang Chang, Javier Fernandez, Yu-Kai Wang, Jose Antonio Sanz, Gracaliz Pereira Dimuro, and Chin-Teng Lin. “Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface”. In: *IEEE Computational Intelligence Magazine* 14.1 (2019), pp. 96–106.
- [35] Min Wu, Wanjuan Su, Luefeng Chen, Witold Pedrycz, and Kaoru Hirota. “Two-stage fuzzy fusion based-convolution neural network for dynamic emotion recognition”. In: *IEEE Transactions on Affective Computing* (2020).



- [36] Philippe Smets. “The combination of evidence in the transferable belief model”. In: *IEEE Transactions on pattern analysis and machine intelligence* 12.5 (1990), pp. 447–458.
- [37] Chunfeng Lian, Su Ruan, Thierry Dencœux, Hua Li, and Pierre Vera. “Joint Tumor Segmentation in PET-CT Images Using Co-Clustering and Fusion Based on Belief Functions”. In: *IEEE Transactions on Image Processing* 28.2 (2019), pp. 755–766.
- [38] Zhunga Liu, Xuxia Zhang, Jiawei Niu, and Jean Dezert. “Combination of classifiers with different frames of discernment based on belief functions”. In: *IEEE Transactions on Fuzzy Systems* (2020).
- [39] Hind Laghmara, Thomas Laurain, Christophe Cudel, and Jean-Philippe Lauffenburger. “Heterogeneous sensor data fusion for multiple object association using belief functions”. In: *Information Fusion* 57 (2020), pp. 44–58.
- [40] Amelio Vazquez-Reina, Michael Gelbart, Daniel Huang, Jeff Lichtman, Eric Miller, and Hanspeter Pfister. “Segmentation fusion for connectomics”. In: *2011 International Conference on Computer Vision*. IEEE. 2011, pp. 177–184.
- [41] Nitish Srivastava and Ruslan R Salakhutdinov. “Multimodal learning with deep boltzmann machines”. In: *Advances in neural information processing systems*. 2012, pp. 2222–2230.
- [42] Hongmin Cai, Ragini Verma, Yangming Ou, Seung-koo Lee, Elias R Melhem, and Christos Davatzikos. “Probabilistic segmentation of brain tumors based on multi-modality magnetic resonance images”. In: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. 2007, pp. 600–603.
- [43] Nan Zhang, Su Ruan, Stéphane Lebonvallet, Qingmin Liao, and Yuemin Zhu. “Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation”. In: *Computer Vision and Image Understanding* 115.2 (2011), pp. 256–269.
- [44] S Krishnakumar and K Manivannan. “Effective segmentation and classification of brain tumor using rough K means algorithm and multi kernel SVM in MR images”. In: *Journal of Ambient Intelligence and Humanized Computing* 12.6 (2021), pp. 6751–6760.
- [45] Ling Huang, Thierry Dencœux, David Tonnelet, Pierre Decazes, and Su Ruan. “Deep PET/CT fusion with Dempster-Shafer theory for lymphoma segmentation”. In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2021, pp. 30–39.
- [46] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.
- [47] Matthew D Zeiler and Rob Fergus. “Visualizing and understanding convolutional networks”. In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

- [48] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
- [49] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.
- [50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Identity mappings in deep residual networks”. In: *European conference on computer vision*. Springer. 2016, pp. 630–645.
- [51] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.
- [52] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.
- [53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [54] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *CVPR09*. 2009.
- [55] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. “The multimodal brain tumor image segmentation benchmark (BRATS)”. In: *IEEE transactions on medical imaging* 34.10 (2015), pp. 1993–2024.
- [56] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. “Brain tumor segmentation using convolutional neural networks in MRI images”. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1240–1251.
- [57] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. “Brain tumor segmentation with deep neural networks”. In: *Medical image analysis* 35 (2017), pp. 18–31.
- [58] Shengcong Chen, Changxing Ding, and Minfeng Liu. “Dual-force convolutional neural networks for accurate brain tumor segmentation”. In: *Pattern Recognition* 88 (2019), pp. 90–100.
- [59] Dingwen Zhang, Guohai Huang, Qiang Zhang, Jungong Han, Junwei Han, and Yizhou Yu. “Cross-modality deep feature learning for brain tumor segmentation”. In: *Pattern Recognition* 110 (2021), p. 107562.
- [60] Alexander Kalinovsky and Vassili Kovalev. “Lung image Ssgmentation using deep learning methods and convolutional neural networks”. In: (2016).

- [61] Johnatan Carvalho Souza, João Otávio Bandeira Diniz, Jonnison Lima Ferreira, Giovanni Lucca França da Silva, Aristofanes Correa Silva, and Anselmo Cardoso de Paiva. “An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks”. In: *Computer methods and programs in biomedicine* 177 (2019), pp. 285–296.
- [62] Qinhua Hu, Luis Fabricio de F Souza, Gabriel Bandeira Holanda, Shara SA Alves, Francisco Hercules dos S Silva, Tao Han, and Pedro P Reboucas Filho. “An effective approach for CT lung segmentation using mask region-based convolutional neural networks”. In: *Artificial intelligence in medicine* 103 (2020), p. 101792.
- [63] Min Fu, Wenming Wu, Xiafei Hong, Qiuhua Liu, Jialin Jiang, Yaobin Ou, Yupei Zhao, and Xinqi Gong. “Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images”. In: *BMC systems biology* 12.4 (2018), p. 56.
- [64] Yunze Man, Yangsibo Huang, Junyi Feng, Xi Li, and Fei Wu. “Deep q learning driven ct pancreas segmentation with geometry-aware u-net”. In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1971–1980.
- [65] Chuansheng Zheng, Xianbo Deng, Qing Fu, Qiang Zhou, Jiawei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang. “Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label”. In: *medRxiv* (2020).
- [66] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. “Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2015, pp. 556–564.
- [67] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng. “Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images”. In: *Thirty-first AAAI conference on artificial intelligence*. 2017.
- [68] Yi Wang, Haoran Dou, Xiaowei Hu, Lei Zhu, Xin Yang, Ming Xu, Jing Qin, Pheng-Ann Heng, Tianfu Wang, and Dong Ni. “Deep attentive features for prostate segmentation in 3d transrectal ultrasound”. In: *IEEE transactions on medical imaging* 38.12 (2019), pp. 2768–2778.
- [69] Quande Liu, Qi Dou, Lequan Yu, and Pheng Ann Heng. “MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data”. In: *IEEE transactions on medical imaging* 39.9 (2020), pp. 2713–2724.
- [70] Xiangrong Zhou, Ryosuke Takayama, Song Wang, Takeshi Hara, and Hiroshi Fujita. “Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method”. In: *Medical physics* 44.10 (2017), pp. 5221–5233.

- [71] Roger Trullo, Caroline Petitjean, Dong Nie, Dinggang Shen, and Su Ruan. “Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 21–29.
- [72] Yan Wang, Yuyin Zhou, Wei Shen, Seyoun Park, Elliot K Fishman, and Alan L Yuille. “Abdominal multi-organ segmentation with organ-attention networks and statistical fusion”. In: *Medical image analysis* 55 (2019), pp. 88–102.
- [73] Yang Lei, Tonghe Wang, Sibao Tian, Xue Dong, Ashesh B Jani, David Schuster, Walter J Curran, Pretesh Patel, Tian Liu, and Xiaofeng Yang. “Male pelvic multi-organ segmentation aided by CBCT-based synthetic MRI”. In: *Physics in Medicine & Biology* 65.3 (2020), p. 035013.
- [74] Gaurav Bhatnagar, QM Jonathan Wu, and Zheng Liu. “A new contrast based multimodal medical image fusion framework”. In: *Neurocomputing* 157 (2015), pp. 143–152.
- [75] Zhe Guo, Xiang Li, Heng Huang, Ning Guo, and Quanzheng Li. “Deep Learning-Based Image Segmentation on Multimodal Medical Imaging”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 3.2 (2019), pp. 162–169.
- [76] *The Anatomy of MRI*. URL: <https://www.nia.nih.gov/health/biomarkers-dementia-detection-and-research/> (visited on 09/27/2021).
- [77] Y LeCun, Y Bengio, and G Hinton. “Deep learning. nature 521 (7553): 436”. In: *Google Scholar* (2015).
- [78] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. “Greedy layer-wise training of deep networks”. In: *Advances in neural information processing systems*. 2007, pp. 153–160.
- [79] Ruslan Salakhutdinov and Geoffrey Hinton. “Deep boltzmann machines”. In: *Artificial intelligence and statistics*. 2009, pp. 448–455.
- [80] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. “Backpropagation applied to handwritten zip code recognition”. In: *Neural computation* 1.4 (1989), pp. 541–551.
- [81] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- [82] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.

- [83] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, et al. “MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans”. In: *Computational intelligence and neuroscience* 2015 (2015), p. 1.
- [84] Ivana Išgum, Manon JNL Benders, Brian Avants, M Jorge Cardoso, Serena J Counsell, Elda Fischi Gomez, Laura Gui, Petra S Hüppi, Karina J Kersbergen, Antonios Makropoulos, et al. “Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge”. In: *Medical image analysis* 20.1 (2015), pp. 135–151.
- [85] Li Wang, Dong Nie, Guannan Li, Élodie Puybareau, Jose Dolz, Qian Zhang, Fan Wang, Jing Xia, Zhengwang Wu, Jiawei Chen, et al. “Benchmark on Automatic 6-month-old Infant Brain Segmentation Algorithms: The iSeg-2017 Challenge”. In: *IEEE transactions on medical imaging* (2019).
- [86] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. “Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge”. In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 287–297.
- [87] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. “No new-net”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 234–244.
- [88] Yao Qin, Konstantinos Kamnitsas, Siddharth Ancha, Jay Nanavati, Garrison Cottrell, Antonio Criminisi, and Aditya Nori. “Autofocus layer for semantic segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 603–611.
- [89] Jose Dolz, Karthik Gopinath, Jing Yuan, Herve Lombaert, Christian Desrosiers, and Ismail Ben Ayed. “HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation”. In: *IEEE transactions on medical imaging* (2018).
- [90] Shaoguo Cui, Lei Mao, Jingfeng Jiang, Chang Liu, and Shuyu Xiong. “Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network”. In: *Journal of healthcare engineering* 2018 (2018).
- [91] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. “IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet”. In: *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*. Springer. 2018, pp. 130–143.
- [92] Dong Nie, Li Wang, Yaozong Gao, and Dinggang Sken. “Fully convolutional networks for multi-modality isointense infant brain image segmentation”. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2016, pp. 1342–1345.

- [93] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. “Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks”. In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 178–190.
- [94] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation”. In: *Medical image analysis* 36 (2017), pp. 61–78.
- [95] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan. “A deep learning model integrating FCNNs and CRFs for brain tumor segmentation”. In: *Medical image analysis* 43 (2018), pp. 98–111.
- [96] Pawel Mlynarski, Hervé Delingette, Antonio Criminisi, and Nicholas Ayache. “3D convolutional neural networks for tumor segmentation using long-range 2D context”. In: *Computerized Medical Imaging and Graphics* 73 (2019), pp. 60–72.
- [97] Konstantinos Kamnitsas, Wenjia Bai, Enzo Ferrante, Steven McDonagh, Matthew Sinclair, Nick Pawlowski, Martin Rajchl, Matthew Lee, Bernhard Kainz, Daniel Rueckert, et al. “Ensembles of multiple models and architectures for robust brain tumour segmentation”. In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 450–462.
- [98] Luis Perez and Jason Wang. “The effectiveness of data augmentation in image classification using deep learning”. In: *arXiv preprint arXiv:1712.04621* (2017).
- [99] Andriy Myronenko. “3D MRI brain tumor segmentation using autoencoder regularization”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 311–320.
- [100] Albert Clèrigues, Sergi Valverde, Jose Bernal, Jordi Freixenet, Arnau Oliver, and Xavier Lladó. “SUNet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal MRI”. In: *arXiv preprint arXiv:1810.13304* (2018).
- [101] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets”. In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.
- [102] Hao-Yu Yang and Junlin Yang. “Automatic Brain Tumor Segmentation with Contour Aware Residual Network and Adversarial Training”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 267–278.
- [103] Yuankai Huo, Zhoubing Xu, Shunxing Bao, Camilo Bermudez, Hyeonsoo Moon, Prasanna Parvathaneni, Tamara K Moyo, Michael R Savona, Albert Assad, Richard G Abramson, et al. “Splenomegaly Segmentation on Multi-modal MRI using Deep Convolutional Networks”. In: *IEEE transactions on medical imaging* (2018).
- [104] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. “Image-to-image translation with conditional adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.

- [105] Lele Chen, Yue Wu, Adora M DSouza, Anas Z Abidin, Axel Wismüller, and Chenliang Xu. “MRI tumor segmentation with densely connected 3D CNN”. In: *Medical Imaging 2018: Image Processing*. Vol. 10574. International Society for Optics and Photonics. 2018, 105741F.
- [106] Lior Rokach. “Ensemble-based classifiers”. In: *Artificial Intelligence Review* 33.1-2 (2010), pp. 1–39.
- [107] Mehmet Aygün, Yusuf Hüseyin Şahin, and Gözde Ünal. “Multi modal convolutional neural networks for brain tumor segmentation”. In: *arXiv preprint arXiv:1809.06191* (2018).
- [108] Chenhong Zhou, Changxing Ding, Zhentai Lu, Xinchao Wang, and Dacheng Tao. “One-pass multi-task convolutional neural networks for efficient brain tumor segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 637–645.
- [109] Peter D Chang et al. “Fully convolutional neural networks with hyperlocal features for brain tumor segmentation”. In: *Proceedings MICCAI-BRATS Workshop*. 2016, pp. 4–9.
- [110] Zeyu Jiang, Changxing Ding, Minfeng Liu, and Dacheng Tao. “Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task”. In: *International MICCAI Brainlesion Workshop*. Springer. 2019, pp. 231–241.
- [111] Fabian Isensee, Paul F. Jäger, Peter M. Full, Philipp Vollmuth, and Klaus H. Maier-Hein. “nnU-Net for Brain Tumor Segmentation”. In: *Lecture Notes in Computer Science* (2021), 118–132. ISSN: 1611-3349. DOI: 10.1007/978-3-030-72087-2\_11.
- [112] Nedelcho Georgiev and Asen Asenov. “Automatic Segmentation of Lumbar Spine MRI Using Ensemble of 2D Algorithms”. In: *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*. Springer. 2018, pp. 154–162.
- [113] Annegreet van Opbroek, Fedde van der Lijn, and Marleen de Bruijne. “Automated brain-tissue segmentation by multi-feature SVM classification”. In: *Proceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS’13)*. 2013.
- [114] Toan Duc Bui, Jitae Shin, and Taesup Moon. “3D densely convolutional networks for volumetric segmentation”. In: *arXiv preprint arXiv:1709.03199* (2017).
- [115] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.
- [116] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. “A survey of MRI-based medical image analysis for brain tumor studies”. In: *Physics in Medicine & Biology* 58.13 (2013), R97.
- [117] Antonios Drevelegas. *Imaging of brain tumors with histological correlations*. Springer Science & Business Media, 2010.

- [118] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. “Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 421–429.
- [119] **Tongxue Zhou**, Su Ruan, Yu Guo, and Stéphane Canu. “A Multi-Modality Fusion Network Based on Attention Mechanism for Brain Tumor Segmentation”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 377–380.
- [120] **Tongxue Zhou**, Su Ruan, Haigen Hu, and Stéphane Canu. “Deep Learning Model Integrating Dilated Convolution and Deep Supervision for Brain Tumor Segmentation in Multi-parametric MRI”. In: *International Workshop on Machine Learning in Medical Imaging (MLMI)*. Springer. 2019, pp. 574–582.
- [121] **Tongxue Zhou**, Stéphane Canu, and Su Ruan. “Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation”. In: *Computerized Medical Imaging and Graphics* 86 (2020), p. 101811.
- [122] **Tongxue Zhou**, Stéphane Canu, and Su Ruan. “Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism”. In: *International Journal of Imaging Systems and Technology* 31.1 (2021), pp. 16–27.
- [123] Chunfeng Lian, Su Ruan, Thierry Denœux, Hua Li, and Pierre Vera. “Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions”. In: *IEEE Transactions on Image Processing* 28.2 (2018), pp. 755–766.
- [124] Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, Cagatay Demiralp, Jamie Shotton, Owen M Thomas, Tilak Das, Raj Jena, and Stephen J Price. “Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2012, pp. 369–376.
- [125] Yuntao Yu, Pierre Decazes, Jérôme Lapuyade-Lahorgue, Isabelle Gardin, Pierre Vera, and Su Ruan. “Semi-automatic lymphoma detection and segmentation using fully conditional random fields”. In: *Computerized Medical Imaging and Graphics* 70 (2018), pp. 1–7.
- [126] Stefan Bauer, Lutz-P Nolte, and Mauricio Reyes. “Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2011, pp. 354–361.
- [127] DP Onoma, Su Ruan, Sébastien Thureau, Lamyaa Nkhali, Romain Modzelewski, GA Monnehan, Pierre Vera, and Isabelle Gardin. “Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm”. In: *Computerized Medical Imaging and Graphics* 38.8 (2014), pp. 753–763.



- [128] Yu Wang, Changsheng Li, Ting Zhu, and Jingyang Zhang. “Multimodal brain tumor image segmentation using WRN-PPNet”. In: *Computerized Medical Imaging and Graphics* 75 (2019), pp. 56–65.
- [129] **Tongxue Zhou**, Su Ruan, and Stéphane Canu. “A review: Deep learning for medical image segmentation using multi-modality fusion”. In: *Array* (2019), p. 100004.
- [130] Kuan-Lun Tseng, Yen-Liang Lin, Winston Hsu, and Chung-Yang Huang. “Joint sequence learning and cross-modality convolution for 3d biomedical segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6393–6400.
- [131] Abhinav Valada, Gabriel L Oliveira, Thomas Brox, and Wolfram Burgard. “Deep multispectral semantic scene understanding of forested environments using multi-modal fusion”. In: *International Symposium on Experimental Robotics*. Springer. 2016, pp. 465–477.
- [132] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-excitation networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.
- [133] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. “Pyramid attention network for semantic segmentation”. In: *arXiv preprint arXiv:1805.10180* (2018).
- [134] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. “Attention u-net: Learning where to look for the pancreas”. In: *arXiv preprint arXiv:1804.03999* (2018).
- [135] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. “‘Squeeze & excite’guided few-shot segmentation of volumetric images”. In: *Medical image analysis* 59 (2020), p. 101587.
- [136] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. “Dual attention network for scene segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 3146–3154.
- [137] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. “Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features”. In: *Scientific data* 4 (2017), p. 170117.
- [138] Fisher Yu and Vladlen Koltun. “Multi-scale context aggregation by dilated convolutions”. In: *arXiv preprint arXiv:1511.07122* (2015).
- [139] Brian B Avants, Nick Tustison, and Gang Song. “Advanced normalization tools (ANTS)”. In: *Insight j* 2 (2009), pp. 1–35.
- [140] Sérgio Pereira, Victor Alves, and Carlos A Silva. “Adaptive feature recombination and recalibration for semantic segmentation: application to brain tumor segmentation in MRI”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 706–714.

- [141] Andrew Jesson and Tal Arbel. “Brain tumor segmentation using a 3D FCN with multi-scale loss”. In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 392–402.
- [142] **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Brain tumor segmentation with missing modalities via latent multi-source correlation representation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2020, pp. 533–541.
- [143] **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “3D Medical Multi-modal Segmentation Network Guided by Multi-source Correlation Constraint”. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021, pp. 10243–10250.
- [144] **Tongxue Zhou**, Su Ruan, Pierre Vera, and Stéphane Canu. “A Tri-attention Fusion Guided Multi-modal Segmentation Network”. In: *Pattern Recognition (2021)*, p. 108417.
- [145] Asieh Khosravanian, Mohammad Rahmanimanesh, Parviz Keshavarzi, and Saeed Mozaffari. “Fast level set method for glioma brain tumor segmentation based on Superpixel fuzzy clustering and lattice Boltzmann method”. In: *Computer Methods and Programs in Biomedicine* 198 (2021), p. 105809.
- [146] Jose Dolz, Christian Desrosiers, Li Wang, Jing Yuan, Dinggang Shen, and Ismail Ben Ayed. “Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation”. In: *Computerized Medical Imaging and Graphics* 79 (2020), p. 101660.
- [147] Hao Chen, Zhiguang Qin, Yi Ding, Lan Tian, and Zhen Qin. “Brain tumor segmentation with deep convolutional symmetric neural network”. In: *Neurocomputing* 392 (2020), pp. 305–313.
- [148] Jie Wei, Yong Xia, and Yanning Zhang. “M3Net: A multi-model, multi-size, and multi-view deep neural network for brain magnetic resonance image segmentation”. In: *Pattern Recognition* 91 (2019), pp. 366–378.
- [149] Sergio Sanchez-Martinez, Nicolas Duchateau, Tamas Erdei, Alan G Fraser, Bart H Bijnens, and Gemma Piella. “Characterization of myocardial motion patterns by unsupervised multiple kernel learning”. In: *Medical image analysis* 35 (2017), pp. 70–82.
- [150] Nicolle M Correa, Tom Eichele, Tülay Adalı, Yi-Ou Li, and Vince D Calhoun. “Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRF”. In: *Neuroimage* 50.4 (2010), pp. 1438–1445.
- [151] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).
- [152] Mohamed Akil, Rachida Saouli, Rostom Kachouri, et al. “Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy”. In: *Medical Image Analysis* (2020), p. 101692.

- [153] Yan Hu, Xiang Liu, Xin Wen, Chen Niu, and Yong Xia. “Brain Tumor Segmentation on Multimodal MR Imaging Using Multi-level Upsampling in Decoder”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 168–177.
- [154] Evan Gates, J Gregory Pauloski, Dawid Schellingerhout, and David Fuentes. “Glioma segmentation and a simple accurate model for overall survival prediction”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 476–484.
- [155] Tran Anh Tuan et al. “Brain Tumor Segmentation Using Bit-plane and UNET”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 466–475.
- [156] Xiaobin Hu, Hongwei Li, Yu Zhao, Chao Dong, Bjoern H Menze, and Marie Piraud. “Hierarchical multi-class segmentation of glioma images using networks with multi-level activation function”. In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 116–127.
- [157] **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 4263–4274.
- [158] **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing MR modalities”. In: *Neurocomputing* 466 (2021), pp. 102–112.
- [159] Anjali Wadhwa, Anuj Bhardwaj, and Vivek Singh Verma. “A review on brain tumor segmentation of MRI images”. In: *Magnetic resonance imaging* 61 (2019), pp. 247–259.
- [160] Yi Ding, Linpeng Gong, Mingfeng Zhang, Chang Li, and Zhiguang Qin. “A multi-path adaptive fusion network for multimodal brain tumor segmentation”. In: *Neurocomputing* 412 (2020), pp. 19–30.
- [161] Chenhong Zhou, Changxing Ding, Xinchao Wang, Zhentai Lu, and Dacheng Tao. “One-pass multi-task networks with cross-task guided attention for brain tumor segmentation”. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 4516–4529.
- [162] Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. “Hemis: Hetero-modal image segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 469–477.
- [163] Kenneth Lau, Jonas Adler, and Jens Sjölund. “A unified representation network for segmentation with missing modalities”. In: *arXiv preprint arXiv:1908.06683* (2019).
- [164] Agisilaos Chatsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A Tsafaris. “Multimodal MR synthesis via modality-invariant latent representation”. In: *IEEE transactions on medical imaging* 37.3 (2017), pp. 803–814.

- [165] Reuben Dorent, Samuel Joutard, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. “Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation”. In: *arXiv preprint arXiv:1907.11150* (2019).
- [166] Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. “Medical image synthesis with deep convolutional adversarial networks”. In: *IEEE Transactions on Biomedical Engineering* 65.12 (2018), pp. 2720–2730.
- [167] Biting Yu, Luping Zhou, Lei Wang, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat. “Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis”. In: *IEEE transactions on medical imaging* 38.7 (2019), pp. 1750–1762.
- [168] Yan Xia, Le Zhang, Nishant Ravikumar, Rahman Attar, Stefan K Piechnik, Stefan Neubauer, Steffen E Petersen, and Alejandro F Frangi. “Recovering from missing data in population imaging—Cardiac MR image imputation via conditional generative adversarial nets”. In: *Medical Image Analysis* 67 (2020), p. 101812.
- [169] Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. “Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 447–456.
- [170] Yan Shen and Mingchen Gao. “Brain Tumor Segmentation on MRI with Missing Modalities”. In: *International Conference on Information Processing in Medical Imaging*. Springer. 2019, pp. 417–428.
- [171] Minhao Hu, Matthis Maillard, Ya Zhang, Tommaso Ciceri, Giammarco La Barbera, Isabelle Bloch, and Pietro Gori. “Knowledge distillation from multi-modal to mono-modal segmentation networks”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 772–781.
- [172] Yian Zhu, Shaoyu Wang, Runlong Lin, Yun Hu, and Qiang Chen. “Brain Tumor Segmentation for Missing Modalities by Supplementing Missing Features”. In: *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. IEEE. 2021, pp. 652–656.
- [173] **Tongxue Zhou**, Stéphane Canu, Pierre Vera, and Su Ruan. “A Dual Supervision Guided Attentional Network for Multimodal MR Brain Tumor Segmentation”. In: *International Conferenc on Medical Image and Computer-Aided Diagnosis (MICAD)*. 2021.
- [174] Haigen Hu, Leizhao Shen, **Tongxue Zhou**, Pierre Decazes, Pierre Vera, and Su Ruan. “Lymphoma segmentation in PET images based on multi-view and Conv3D fusion strategy”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 1197–1200.