HAL
open science

Image segmentation and extraction based on pixel communities

Thanh-Khoa Nguyen

To cite this version:

Thanh-Khoa Nguyen. Image segmentation and extraction based on pixel communities. Image Processing [eess.IV]. Université de La Rochelle, 2019. English. NNT : 2019LAROS035 . tel-03223157

HAL Id: tel-03223157

https://theses.hal.science/tel-03223157v1

Submitted on 10 May 2021

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.

UNIVERSITY OF LA ROCHELLE

DOCTORAL THESIS

Image Segmentation and Extraction Based on Pixel Communities

Author:
Thanh-Khoa NGUYEN

Supervisors:
Prof. Jean-Loup GUILLAUME
Assoc. Prof. Mickaël COUSTATY

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Laboratory of Informatics, Image and Interaction
Faculty of Science and Technology

December 20, 2019

Composition du jury :

M. COUSTATY Mickaël, Maître de conférences (Co-directeur de thèse)
L3i - La Rochelle Université, France

M. DANISCH Maximilien, Maître de conférences (Examinateur)
LIP6 - Sorbonne Université, France

Mme EGLIN Véronique, Professeure (Rapporteur)
LIRIS - INSA de Lyon, France

M. GUILLAUME Jean-Loup, Professeur (Directeur de thèse)
L3i - La Rochelle Université, France

M. REVEL Arnaurd, Professeur (Président)
L3i - La Rochelle Université, France

Mme LE-GRAND Bénédicte, Professeure (Rapporteur)
CRI - Université Paris 1, France

Mme VINCENT Nicole, Professeure (Examinateur)
LIPADE - Université Paris Descartes, France

Declaration of Authorship

I, Thanh-Khoa NGUYEN, declare that this thesis titled, "Image Segmentation and Extraction Based on Pixel Communities" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

RÉSUMÉ

La segmentation d'images est devenue une tâche indispensable largement utilisée dans plusieurs applications de traitement d'images, notamment la détection d'objets, le suivi d'objets, l'assistance automatique à la conduite et les systèmes de contrôle du trafic, *etc.* La littérature regorge d'algorithmes permettant de réaliser des tâches de segmentation d'images. Ces méthodes peuvent être divisées en groupes principaux en fonction des approches sous-jacentes, telles que la segmentation d'images basée sur les régions, la classification basée sur les caractéristiques de l'image, les approches basées sur les graphes et la segmentation d'images basée sur les réseaux de neurones.

Récemment, l'analyse de réseaux sociaux a proposé de nombreuses théories et méthodologies. En particulier, des techniques de segmentation d'images basées sur des algorithmes de détection de communautés ont été proposées et forment une famille d'approches visible dans la littérature. Dans cette thèse, nous proposons un nouveau cadre pour la segmentation d'images basée sur la détection de communautés. Si l'idée de base d'utiliser le domaine de l'analyse des réseaux sociaux dans la segmentation de l'image est tout à fait séduisante, la manière dont les algorithmes de détection de communautés peuvent être appliqués efficacement à la segmentation d'images est un sujet qui continue à interroger. L'apport de cette thèse est un effort pour construire de manière pertinente des meilleurs réseaux complexes en fonction de l'application, des méthodes utilisées pour la détection de communautés et pour proposer de nouvelles méthodes pour agréger les régions homogènes afin de produire de bonnes segmentations d'images.

Par ailleurs, nous proposons également un système de recherche d'images par le contenu (content-based image retrieval) utilisant les mêmes caractéristiques que celles obtenues par les processus de segmentation d'images. Le moteur de recherche d'images proposé fonctionne pour des images de scènes naturelles et permet de rechercher les images les plus similaires à l'image requête. Ce moteur de recherche d'images par le contenu repose sur l'utilisation des régions extraites comme mots visuels dans le modèle Bag-of-Visual-Words. Ceci permet de valider la généricité de notre approche de segmentation d'images à partir de réseaux complexes et son utilisation dans plusieurs domaines d'applications liés au traitement d'images et de vision par ordinateur.

Nos méthodes ont été testées sur plusieurs jeux de données et évaluées en utilisant différentes mesures classiques de la qualité d'une segmentation. Les méthodes proposées produisent des segmentation d'image dont la qualité est supérieure à l'état de l'art.

Mots clés: *Segmentation d'images, Segmentation d'images basée sur les graphes, Réseaux complexes, Modularité, Superpixels, Algorithme de Louvain, Détection de communautés, Segmentation d'images basée sur le consensus, Recherche d'images basée sur le contenu, Sacs de mots visuels, Extraction de caractéristiques.*

UNIVERSITY OF LA ROCHELLE

Abstract

Faculty of Science and Technology
Laboratory of Informatics, Image and Interaction

Doctor of Philosophy

Image Segmentation and Extraction Based on Pixel Communities

by Thanh-Khoa NGUYEN

Image segmentation has become an indispensable task that is widely employed in several image processing applications including object detection, object tracking, automatic driver assistance, and traffic control systems, *etc.* The literature abounds with algorithms for achieving image segmentation tasks. These methods can be divided into some main groups according to the underlying approaches, such as Region-based image segmentation, Feature-based clustering, Graph-based approaches and Artificial Neural Network-based image segmentation.

Recently, complex networks have mushroomed both theories and applications as a trend of developments. Hence, image segmentation techniques based on community detection algorithms have been proposed and have become an interesting discipline in the literature. In this thesis, we propose a novel framework for community detection based image segmentation. The idea that brings social networks analysis domain into image segmentation quite satisfies with most authors and harmony in those researches. However, how community detection algorithms can be applied in image segmentation efficiently is a topic that has challenged researchers for decades. The contribution of this thesis is an effort to construct best complex networks for applying community detection and proposal novel agglomerate methods in order to aggregate homogeneous regions producing good image segmentation results.

Besides, we also propose a content based image retrieval system using the same features than the ones obtained by the image segmentation processes. The proposed image search engine for real images can implement to search the closest similarity images with query image. This content based image retrieval relies on the incorporation of our extracted features into Bag-of-Visual-Words model. This is one of representative applications denoted that image segmentation benefits several image processing and computer visions applications.

Our methods have been tested on several data sets and evaluated by many well-known segmentation evaluation metrics. The proposed methods produce efficient image segmentation results compared to the state of the art.

Keywords: *Image segmentation, Graph-based image segmentation, Complex networks, Modularity, Superpixels, Louvain algorithm, Community detection, Consensus-based image segmentation, Content-based image retrieval, Bag-of-Visual-Words, Feature extraction.*

Acknowledgements

The PhD process is long difficult and complicated progress that requires intellectually stimulating and emotionally draining. Therefore, to fulfill this mission requires the support of many people. I would like to express my warmest gratitude and acknowledgement to the following people for their role in the completion of this thesis:

- My supervisors, Jean-Loup Guillaume and Mickaël Coustaty, for their superb guidance. I could not have asked for better support, advice, or enthusiasm, and working with them has been a great pleasure. Thank you for your patience through the many ups and downs over the years.
- Cao-De Tran and Jean-Marc Ogier, for their help. I could not have had a chance of study at Laboratory L3i, University of La Rochelle, France if without their recommendation and acceptation. Thank you very much for your help and trust in me.
- All of my friends in Laboratory L3i, thank you for making me laugh, for all of the research discussions, for listening to me, for letting me listen to you, for many coffee breaks, and for keeping me relatively sane.
- My family, thank you for your never-faltering encouragement, support, love and incredibly understanding despite my long absences.
- My wife, the strength of our relationship amazes me every day. Thank you for your devoted love, support and encouragement, whether near or far. Sharing a meal with you, I know that I'm home.

Last but not least, I am immensely grateful to Mekong 120 Ca Mau Project for their generous financial support, without which my studies would not have been possible. I would also like to thank the University of La Rochelle and Laboratory L3i for supporting my attendance at conferences.

Contents

Declaration of Authorship iii

Abstract vii

Acknowledgements ix

1 Introduction 1
- 1.1 Objectives . . . 1
- 1.2 Motivation . . . 2
- 1.3 Problem description . . . 4
- 1.4 Approach . . . 5
- 1.5 Contributions . . . 6
- 1.6 Outline . . . 6

2 Image Segmentation Algorithms 7
- 2.1 Introduction . . . 7
- 2.2 Taxonomy . . . 7
- 2.3 Image Segmentation Techniques . . . 10
  - 2.3.1 Region-based Image Segmentation . . . 10
  - 2.3.2 Feature-based Image Segmentation . . . 11
  - 2.3.3 Graph-based Image Segmentation . . . 11
  - 2.3.4 ANN-based Image Segmentation . . . 20

3 Complex Networks and Images 25
- 3.1 Introduction . . . 25
- 3.2 Complex Networks . . . 26
  - 3.2.1 Community Detection and Modularity . . . 26
  - 3.2.2 Louvain Algorithm . . . 28
- 3.3 Generating a complex network from an image . . . 29
  - 3.3.1 Pattern definition . . . 30
  - 3.3.2 Pattern evaluation . . . 32
- 3.4 Community Detection Based Image Segmentation . . . 33
  - 3.4.1 Aggregation of Regions Algorithm . . . 34
  - 3.4.2 Noise Removal . . . 35

4 Extraction of Precise Features 37
- 4.1 Introduction . . . 37
- 4.2 Color features . . . 38
  - 4.2.1 Mean and Standard Deviation of Color Approach . . . 38
  - 4.2.2 Reduced Histogram of Color Approach . . . 43

4.3 Texture features 46
4.3.1 Texture-based feature using HOG Approach 46
4.4 Combination of colors and textures 47
4.4.1 Mean and Standard deviation plus HOG feature 48
4.4.2 Balancing HOG and MeanSD Features 49
4.4.3 Merging all features in one 51
4.4.4 Merging and Adapting all Features in One 54
4.5 Conclusion 56

5 Extended Experimental Results for Image Segmentation 57
5.1 Introduction 57
5.2 Datasets 57
5.3 Evaluation metrics 58
5.4 Quantitative evaluation 59
5.4.1 Comparison of patterns and solutions 59
5.4.2 Comparison with the state of the art 61
5.5 Qualitative evaluation 62
5.6 Conclusion 66

6 A Consensus-Based Image Segmentation Algorithm 69
6.1 Proposed Approach 69
6.1.1 Complex Networks from an image 70
6.1.2 Louvain method and consensual clustering 70
6.2 Experimental Evaluation 72
6.2.1 Results 72
6.3 Conclusion 73

7 Using Extracted Features for Content Based Image Retrieval 77
7.1 Introduction 77
7.2 Bag of Visual Works Model 78
7.2.1 Keypoint Detection 79
7.2.2 Keypoint Descriptors 79
7.2.3 Building Vocabulary 80
7.3 Our CBIR Architecture 80
7.3.1 Training stage 80
7.3.2 Testing stage 82
7.4 Experimental Results 83
7.4.1 Dataset 83
7.4.2 Results 83
7.5 Conclusion 83

8 Conclusion and future directions 87
8.1 Conclusions 87
8.2 Future work 88

A List of Publications 91

Bibliography 93

List of Figures

1.1 Primary tasks of this thesis 2

2.1 Image segmentation techniques 8
2.2 Watershed Illustration Chien, Huang, and Chen, 2003 11
2.3 AlexNet Convolutional Neural Network architecture (Krizhevsky, Sutskever, and Hinton, 2012) 21
2.4 VGG-16 Convolutional Neural Network architecture (Simonyan and Zisserman, 2014) 21
2.5 Inception module with dimensionality reduction from the GoogLeNet architecture (Szegedy et al., 2014) 22
2.6 Residual block from the ResNet architecture (He et al., 2016) . 23
2.7 FCN can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation (Long, Shelhamer, and Darrell, 2015) 23

3.1 Networks of interest vary from the Internet to the World Wide Web to scientific collaborations networks. 26
3.2 Community structure in the social network of bottle-nose dolphins population extracted using the algorithm of Girvan and Newman (Girvan and Newman, 2002). The squares and circles denote the primary split of the network into two groups and the circles are further subdivided into four smaller group as shown (Newman, 2004). 27
3.3 Process of community detection for Louvain method (Blondel et al., 2008): four communities are identified after phase 1 (modularity optimization) and Louvain creates a network with four nodes and matching weights after phase 2 (community agglomeration). When this process is repeated during the second pass, two communities are identified and a two-nodes network is created. Finally no more merging can occur and the process stops. 29
3.4 Four typical patterns for creating Complex Networks 31
3.5 The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-ara method (Nguyen, Coustaty, and Guillaume, 2018a) 33
3.6 The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-ahr method (Nguyen, Coustaty, and Guillaume, 2018b) 33

3.7 The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-mhr method (Nguyen, Coustaty, and Guillaume, 2019) 34

4.1 Representation of Pattern A used in this chapter (Nguyen, Coustaty, and Guillaume, 2018a) 39
4.2 Vector representation for a region using RGB color space. . . . 40
4.3 Image segmentation results: *Top:* Original images. *Middle:* after the execution of the Louvain method. *Bottom:* after the merging of homogeneous regions using both Mean and Standard deviation feature and the ARA algorithm. 41
4.4 Image segmentation results: *Top:* Original images. *Middle:* after the execution of the Louvain method. *Bottom:* after the merging of homogeneous regions using both Mean and Standard deviation feature and the ARA algorithm. 42
4.5 Visual Segmentation results of some methods: *Top left:* Original image. *Top right:* Segmentation result of HoS method (Li and Wu, 2015). *Bottom left:* Our segmentation result using mean and standard deviation of color features *Bottom right:* Segmentation result of HOG method. (Li and Wu, 2015; Nguyen, Coustaty, and Guillaume, 2018a) 42
4.6 Creating 48-dimensional vector for each region 44
4.7 Image segmentation results: *Left:* Original images. *Middle:* after the execution of the Louvain method. *Right:* after the merging of homogeneous regions using both reduced histogram of colors and the ARA algorithm (Nguyen, Coustaty, and Guillaume, 2018b) 44
4.8 Visual Segmentation results of some methods: *Top left:* Original image. *Top right:* Segmentation result of HoS method (Li and Wu, 2015). *Bottom left :* Segmentation result of HOG method (Li and Wu, 2015). *Bottom right:* Our segmentation result by ARA and RH feature (Nguyen, Coustaty, and Guillaume, 2018b) 45
4.9 Creating 15-dimensional vector for each region 48
4.10 Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both HOGMeanSD feature and the ARA algorithm. *Fourth line:* Ground-truth. . . 49
4.11 Image segmentation results: *First:* Original images. *Second:* after the execution of the Louvain method. *Third:* after the merging of homogeneous regions using both HOGMeanSD feature and the ARA algorithm. *Fourth:* Ground-truth. 50
4.12 Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth. 51
4.13 Creating 27-dimensional vector for each region 51

4.14 Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth. 52
4.15 Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth. 53
4.16 Creating 63-dimensional vector for each region 54
4.17 Creating 45-dimensional vector for each region 55

5.1 Qualitative evaluation of results using Best pattern (Pattern A) on BSDS 300 dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth. 65
5.2 Qualitative evaluation of results using Best pattern (Pattern A) on BSDS 500 dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth. 66
5.3 Qualitative evaluation of results using Best pattern (Pattern A) on MSRC dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth. 67
5.4 Visual Segmentation results of some methods: *First column:* original image. *Second column:* Segmentation result of HOG method (Li and Wu, 2015). *Third column:* Segmentation result of HoS method. (Li and Wu, 2015) *Fourth column:* Our segmentation results 68
5.5 Qualitative comparison of segmentation results by HoS methods: *First column:* original image. *Second column:* Segmentation result of HoS method (Li and Wu, 2015). *Third column:* Our segmentation results 68

6.1 Evolution of PRI versus consensus threshold 73
6.2 Qualitative evaluation on image 12_19_s of MSRC dataset. From top to bottom and left to right: original image, ground-truth, best segmentation (threshold=0.844), most segmented image (threshold=0.999) 75

7.1 A general Bag of Visual Word Model 81
7.2 An illustration of our proposed CBIR Model 82

7.3 Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10) 84
7.4 Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10) 84
7.5 Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10) 85
7.6 Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10) 85

List of Tables

4.1 Comparative results of ARA algorithm plus MeanSD feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 43
4.2 Comparative results of ARA algorithm plus RH feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 46
4.3 Comparative results of ARA algorithm plus HOGMeanSD feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 50
4.4 Comparative results of ARA algorithm plus BHMF feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 52
4.5 Comparative results of ARA algorithm plus MFO feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 54
4.6 Comparative results of ARA algorithm plus MAFO feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set). 55

5.1 Comparative results of Lv-ahr algorithm using the three different metrics with the Quantized Histogram of Color features (RH) for regions merging (Nguyen, Coustaty, and Guillaume, 2018b). 60
5.2 Comparative results of Lv-ara algorithm using the three different metrics with the MeanSD features for regions merging (Nguyen, Coustaty, and Guillaume, 2018a). 60
5.3 Comparative results of Lv-mhr algorithm using the three different metrics with the HOGMeanSD features for regions merging (Nguyen, Coustaty, and Guillaume, 2019). 61
5.4 Comparative results of Lv-ara algorithm using the three different metrics with the MFO features for regions merging . . . 62
5.5 The results of Lv-ara algorithm using the three different metrics with the BHMF features for regions merging 62
5.6 The results of Lv-ara algorithm using the three different metrics with the MAFO features for regions merging 62
5.7 Quantitative comparisons of our methods using **pattern A** to others methods tested on BSDS300. 63
5.8 Quantitative comparisons of our methods using **Pattern A** to others methods tested on BSDS500. 63

6.1 Quantitative comparisons on BSDS500 and MSRCv2 datasets 74

7.1 Precision proportion of Top 1, Top 5 and Top 10 implemented on dataset including all classes. 86
7.2 Comparison of the proposed method with others. 86

List of Abbreviations

ANN	**A**rtificial **N**eural **N**etworks
ARA	**A**ggregation of **R**egions **A**lgorithm
BHMF	**B**alancing **H**OG and **M**eanSD **F**eatures
BoW	**B**ag **o**f **W**ords
BoVW	**B**ag **o**f **V**isual **W**ords
BRISK	**B**inary **R**obust **I**nvariant **S**calable **K**eypoints
BSDS300	**B**erkeley **S**egmentation **D**ata **S**et 300
BSDS500	**B**erkeley **S**egmentation **D**ata **S**et 500
CA	**C**ellular **A**utomaton
CBIR	**C**ontent-**B**ased **I**mage **R**etrieval
Co-HOG	**C**o-**o**ccurrence of **H**istogram of **O**riented **G**radients
CNN	**C**onvolutional **N**eural **N**etworks
DoG	**D**ifference **o**f **G**aussian
FAST	**F**eatures from **A**ccelerated **S**egment **T**est
FAST-ER	**FAST** **E**nhanced **R**epeatability
FCN	**F**ully **C**onvolutional **N**etwork
FREAK	**F**ast **Re**tin**a** **K**eypoint
GLOH	**G**radient **L**ocation and **O**rientation **H**istogram
HOG	**H**istogram of **O**riented **G**radients
HOGMeanSD	**H**istogram of **O**riented **G**radients, **Mean** and **S**tandard **D**eviation
HoS	**H**istogram **o**f **S**tates
HSV	**H**ue, **S**aturation, **V**alue
ILSVRC	**I**mageNet **L**arge **S**cale **V**isual **R**ecognition **C**ompetition
IR	**I**mage **R**etrieval
JEPD	**J**ointly **E**xhaustive and **P**airwise **D**isjoint set
LBP	**L**ocal **B**inary **P**attern
MeanSD	**Mean** and **S**tandard **D**eviation
MAFO	**M**erging and **A**dapting all **F**eature in **O**ne

MFO	**M**erging all **F**eature in **O**ne
MSER	**M**aximally **S**table **E**xtremal **R**egions
MSRC	**M**icro**s**oft **R**esearch **C**ambridge Object Recognition Image Database
ODS	**O**ptimal **D**ataset **S**cale
OIS	**O**ptimal **I**mage **S**cale
PRI	**P**robabilistic **R**and **I**ndex
VGG	**V**isual **G**eometry **G**roup
VI	**V**ariation of **I**nformation
RCC	**R**egion **C**onnection **C**alculus
ReLUs	**Re**ctified **L**inear **U**nit**s**
RGB	**R**ed, **G**reen, **B**lue
RH	**R**educed **H**istogram of Color Approach
SBIR	**S**emantic-**B**ased **I**mage **R**etrieval
SC	**S**egmentation **C**overing
SIFT	**S**cale-**I**nvariant **F**eature **T**ransform
SURF	**S**peeded **U**p **R**obust **F**eatures
SUSAN	**S**mallest **U**nivalue **S**egment **A**ssimilating **N**ucleus
TBIR	**T**ext-**B**ased **I**mage **R**etrieval

List of Symbols

θ	The orientation of gradient at a pixel	rad
α	The angle degree	$^{\circ}$

This thesis is dedicated to my parents,

I couldn't have done it without

Your genetic material.

Chapter 1

Introduction

1.1 Objectives

This thesis proposes new techniques for image segmentation using the complex network theory and more precisely community detection models. A second contribution presented in this manuscript deals with a content-based image retrieval system mixing the Bag-of-Visual-Words approach and the community detection models where visual words are communities. These tasks, illustrated in Figure 1.1, are described as follows:

For image segmentation methods, we address three sub-goals of a novel framework for image segmentation: (1) building a complex network from an image; (2) extracting precise features for aggregation processes; and (3) proposing an algorithm for agglomerating homogeneous regions.

- **Building a complex network from an image** aims to investigate patterns to construct a graph (complex network) from an image that can increase Louvain algorithm performances.
- **Extracting precise features** addresses the correct choice of efficient features that produce accurate segmentation results, and support image retrieval system effectively.
- **Algorithm for agglomerating homogeneous regions** aims at merging consistent regions based on precise features. Noise removal processes are also used on this step thanks to embedded noise reduction function in the algorithm.

Besides, we also propose another approach to solve the image segmentation problem based on the general principle of consensual clustering (Strehl and Ghosh, 2002). The idea is to execute the Louvain method repeatedly k times and take the similarities between all the obtained partitions into account. The consensus clustering algorithm returns a hierarchy of clusterings that can be assimilated to a hierarchy of image segmentation.

For a content-based image retrieval system, we use our extracted features and the foundation of Bag-of-Words to build the "Image Search Engine" system to search images similar to a query image in an image database.

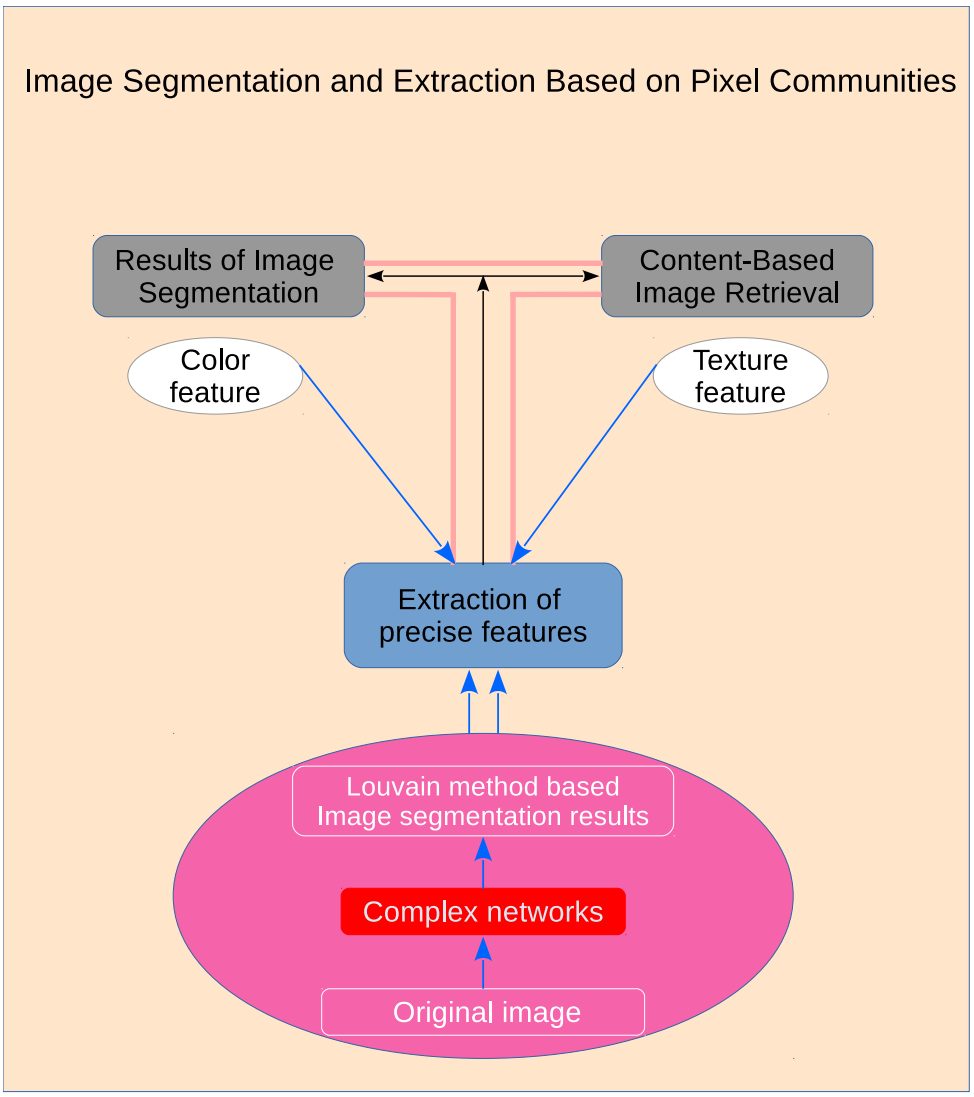


FIGURE 1.1: Primary tasks of this thesis

1.2 Motivation

Image segmentation has become an indispensable task that is widely employed in several image processing applications. The goal of image segmentation is not only to distinguish the interesting objects from the background, but also to identify them in an image. Particularly, image segmentation could benefit a variety of vision applications, for instance, object detection (Liu and Chen, 2008), object tracking (Zhou, Ong, and Ko, 2000), automatic driver assistance (Chen et al., 2008), and traffic control systems (Junwei and Shaokai, 2013), *etc.*

Recently, complex networks have mushroomed both theories and applications as a trend of developments. The study on complex networks has

become one of the most interesting topics during last decade (Abin, Mahdisoltani, and Beigy, 2011; Mourchid, El Hassouni, and Cherifi, 2017) and in particular the detection of communities in networks (Fortunato, 2010). Hence, image segmentation techniques based on community detection algorithms have been proposed and have become an interesting discipline in the literature (Youssef Mourchid, Mohammed El Hassouni and Hocine Cherifi, 2015; Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi, 2016; Abin, Mahdisoltani, and Beigy, 2011; Li and Wu, 2015; Mourchid, El Hassouni, and Cherifi, 2017; Linares et al., 2016; Li, 2013; Browet, Absil, and Van Dooren, 2011).

A community is a group of nodes with dense internal connections and sparse connections with members of other communities. The general idea of those techniques is to highlight the similarity between the modularity criterion in network analysis and the image segmentation process. In fact, the larger the modularity of a network is, the more accurate the detected communities, *i.e.* the objects in the image, are (Browet, Absil, and Van Dooren, 2011; Youssef Mourchid, Mohammed El Hassouni and Hocine Cherifi, 2015; Mourchid, El Hassouni, and Cherifi, 2017; Abin, Mahdisoltani, and Beigy, 2011). If the modeling of the image in a graph is well done then we can expect that a good partition in communities corresponds to a good segmentation of the image. The modularity of a partition is a scalar that measures the density of links inside communities as compared to links between communities, and its value falls into the interval [-0.5,1] (Newman and Girvan, 2004).

Applying the community detection methods for image segmentation has achieved much better results compared to other techniques subjectively (Youssef Mourchid, Mohammed El Hassouni and Hocine Cherifi, 2015). Among all the existing community detection algorithms, the Louvain method (Blondel et al., 2008) has received significant attention in the context of image segmentation (Li and Wu, 2015; Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi, 2016; Browet, Absil, and Van Dooren, 2011). To be inspired, we attempt to regard an image from the perspective of a complex network and using community detection algorithms to solve the image segmentation problem. The Community detection algorithm namely Louvain method has been considered to segment images.

However, it is still facing many problems such as how to generate an appropriate complex network that ensures community detection methods will perform efficiently. Another concern of researchers is that community detection based image segmentation leads to over-segmented results. To overcome this issue, many algorithms have been proposed to deal with image segmentation based on community detection, such as the ones described after.

Wenye Li (Li, 2013), and Youssef, *et al.* (Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi, 2016) attempt to apply community detection problems in complex networks to solve image segmentation problems

and investigate a new graph-based image segmentation as well as compare other methods.

The image segmentation approaches of Ahmad Ali Abin *et al.* (Abin, Mahdisoltani, and Beigy, 2011), and Oscar A. C. Linares *et al.* (Linares et al., 2016) are constructing weighted networks in which the small homogeneous regions *(super-pixels)* obtained by initial segmentation processes are nodes of the graph and the computed similarity distance between these regions are edge weights. One community detection method is applied to extract communities as segments.

Shijie Li, *et al.* (Li and Wu, 2015), and Youssef Mourchid, *et al.* (Mourchid, El Hassouni, and Cherifi, 2017) propose using super-pixel and features to solve the over-segmentation problem. Both strategies initialize with an over-segmented image segmentation in which each sub-segment represents a super-pixel. Then, they treat the over-segmentation issue in different ways. Shijie Li, *et al.* solve it by reconstructing the neighborhood system for each region *(super-pixel)* and the histogram of states *(HoS)* texture feature. Then, they estimate the distribution of the color feature for each region. The similarity matrix W is computed and adaptively updated based on color feature and histogram of states *(HoS)* texture feature. Youssef Mourchid, *et al.* approach the over-segmentation problem in a quite similar way but they compute coefficients to adaptively update the similarity matrix W based on color feature and histogram of oriented gradients *(HOG)* texture feature.

The question is that how we can achieve best results of image segmentation based on community detection in general and based on Louvain method in particular. This motivates our goal of segmenting image based on the idea that brings social networks analysis domain into image segmentation: building a good complex network, extracting precise features and proposing algorithms for agglomerating homogeneous regions.

1.3 Problem description

Maximizing modularity is a NP-hard problem and community detection algorithms are generally heuristics algorithms (mostly without guarantee) and not exact ones. Therefore, creating an appropriate graph (complex network) for any image segmentation based on community detection is a crucial process. Indeed, if we create an inadequate graph it will lead to poor result during the community detection phase. It means that the image segmentation result may get poor quality. On the other hand, if we create a dense graph, it might face to the NP-hardness of the problem. In this thesis, we consider that building an appropriate graph is one of the most important phases.

Furthermore, image segmentation based on community detection often leads to over-segmentation. In order to solve this problem, homogeneous

regions should be combined together whenever possible. The question is to determine which regions are homogeneous. It is a second part in this thesis.

A first sub-problem that should be solved in an image segmentation strategy is noise cancelling. Since we use of community detection algorithms in order to cluster pixels that belong to the same group of information and because, as we stated above, community detection algorithms are ordinarily heuristics algorithms, noise can remain in particular in the form of very small regions. In this thesis, we recommend applying a noise removal strategy to offer better results and obtain higher evaluation scores. The removing noise process is a crucial part of our algorithm because it merges the small regions that remain after Louvain process.

In addition, we propose an algorithm to agglomerate homogeneous regions generating image segmentation results. The majority of algorithm functions are merging regions and removing noise. Algorithm operates based on the precise features that is providing by the way we selected features in various methods. In addition, the noise reduction is embedded in this algorithm, so removing noise processes occurred silently and naturally.

Last but not least, we also propose an image retrieval system which utilizes our extracted features and image segmentation results based on the idea of Bag-of-Visual-Words.

1.4 Approach

We consider images from the perspective of a complex network and solve the image segmentation problem using community detection on graphs. In this approach, an image has been converted into a graph and considered as a complex network. The complex network is built by considering that each pixel is a vertex and edges weight measures the similarity of pairs of pixels.

The Louvain algorithm is applied to the obtained network. The result of this phase is a segmented image but that is over-segmented. Because over-segmented image segmentation produces poor results, we concern ourselves with the fact that over-segmented images in which sub-segments represent super-pixels or homogeneous regions is a challenge that has to be solved.

Our solution is therefore built on top of the Louvain method, using precise features so that they can overcome this drawback and produce more accurate results. Across several experiments and evaluations on many datasets, we obtain very encouraging results, which are in some cases state-of-the-art.

1.5 Contributions

The primary contributions of this dissertation include:

- Proposing methods for building an appropriate graph for image segmentation based community detection.
  - Proposing shapes of patterns to build the graph
  - Investigating and determine performance of these patterns
- Proposing precise features for considering the agglomeration of homogeneous regions
  - Color feature properties
  - Texture feature properties
  - Combination of color and texture features
- Proposing an algorithm for agglomerating homogeneous regions
  - Noise removing strategy
  - General algorithm to agglomerate based on the features
- Proposing a consensus-based image segmentation strategy.
- Proposing a content-based image retrieval system based on Bag-of-Visual-Words using our image segmentation results

1.6 Outline

The dissertation is organized as follows:

- Chapter 1: Introduction
- Chapter 2: Image Segmentation Algorithms
- Chapter 3: Complex Networks and Images
- Chapter 4: Extraction of Precise Features
- Chapter 5: Extended Experimental Results for Image Segmentation
- Chapter 6: A Consensus-Based Image Segmentation Algorithm
- Chapter 7: Using Extracted Features for Content Based Image Retrieval
- Chapter 8: Conclusion and future directions

Chapter 2

Image Segmentation Algorithms

2.1 Introduction

Image segmentation is one of the main steps in image processing. It divides an image into multiple regions in order to analyze them. Image segmentation is also used to distinguish different objects in the image. Rafael C. Gonzalez and Richard E. Woods wrote in their widely used book *Digital Image Processing* that *"Segmentation of nontrivial images is one of the most difficult tasks in image processing. Segmentation accuracy determines the success or failure of computerized analysis procedures."* (C Gonzalez and E Woods, 2002). For this reason, the literature abounds with algorithms for achieving various image segmentation tasks. The purpose of this Chapter is to review and discuss the state-of-the-art in image segmentation techniques. The research in image segmentation field involves a wide range of disciplines in a broad sense, including mathematics, psychology, and informatics, etc., some of which are far beyond the scope of this thesis. In this Chapter, we concentrate on the fundamental and well-known image segmentation techniques, and classify them into taxonomic group based on their characteristics.

2.2 Taxonomy

Image segmentation is considered as one of the fundamental approaches of the digital image processing. It is used widely in several aims of computer vision applications. The purpose of image segmentation is to partition an digital image into meaningful regions in order to analyze with respect to a particular application. In this perspective, several general purpose algorithm and techniques related to image segmentation have been proposed and developed for image segmentation goals. In this Section, we would like to categorize them based on image segmentation techniques into four groups: Region-based image segmentation, Feature-based image segmentation, Graph-based image segmentation and ANN-based image segmentation as it is illustrated in Figure 2.1.

- Region-based Image Segmentation: The region-based image segmentation is a method where neighboring pixels are scanned and added to

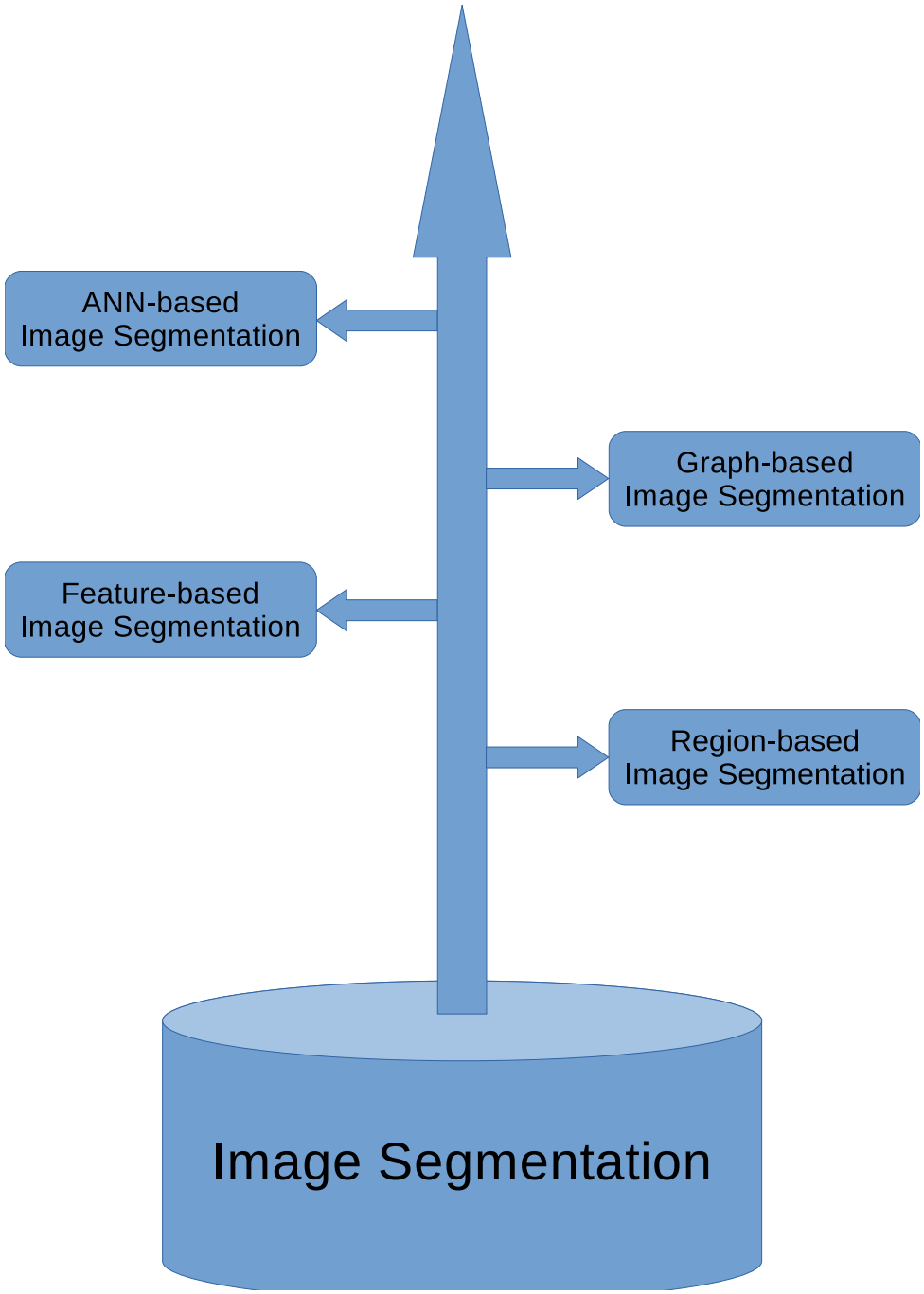


FIGURE 2.1: Image segmentation techniques

a region session if it satisfies a given homogeneity criteria. The region-based image segmentation depicts the splitting an image into standardized sub-regions of connected pixels through the application of homogeneity criteria. Each pixel in a region is similar with when it respects to some characteristics that repeated for each boundary pixel

in the considering region whose the distance similarity of pixels satisfy the homogeneity criteria. A pixel is assigned into an homogeneous region if it satisfies a given homogeneity criteria. Watershed segmentation (Beucher and Lantuéjoul, 1979), Split-and-merge (Horowitz and Pavlidis, 1974) and region-growing techniques (Ikonomatakis et al., 1997) are examples of such methods.

- Feature-based Image Segmentation: Feature is one of the most important attribute used in image processing. In image processing and computer vision, features are pieces of information without formal definition nor particular strict rule to detect them. They can be the result of processing image or computational task based on attributes of image properties. It exists a large variety of features used for image segmentation and other purposes, such as Pixel color, Sub-pixel Corner (Medioni and Yasumoto, 1987), Local Binary Pattern (LBP) (He and Wang, 1990), Shi-Tomasi (Shi and Tomasi, 1994), SUSAN (Smith and Brady, 1997), Scale-invariant feature transform (SIFT) (Lowe, 2004), Texton (Zhu et al., 2005), Histogram of oriented gradients (HOG) (Dalal and Triggs, 2005), Features from Accelerated Segment Test (FAST) (Rosten and Drummond, 2005), SURF (Bay, Tuytelaars, and Van Gool, 2006), FAST- ER (Rosten, Porter, and Drummond, 2010), just to name a few. In Feature-based clustering approaches, one or more particular features can be extracted. Then, the feature-based segmentation stage separates the regions on the image according to the edges. Finally, it is defined a specific distance metric system to aggregate the feature samples into homogeneous regions. The *Mean Shift* is one of the typical representations of Feature-based image segmentation techniques (Comaniciu and Meer, 2002).

- Graph-based Image Segmentation: These techniques are image segmentation methods where an image is regarded as a graph in which vertices represent individual pixels and weighted edges describe the similarity of neighboring pixels. This method is gaining popularity primarily thanks to its ability in reflecting global image properties. Moreover, with the development of complex network and the introduction of superpixel, many graph-based image segmentation approaches have been proposed. Application of modularity in community detection algorithms can be used on graph-based image segmentation, and can reduce the computational cost thereby providing fast computation of performances.

- ANN-based Image Segmentation: Artificial Neural Networks (ANN) have been applied in image segmentation, especially semantic and instance segmentation. ANN-based image segmentation replies on processing small areas of an image using an artificial neural network or a set of neural networks. After such processing the decision-making mechanism marks the areas of an image accordingly to the category recognized by the neural network. The Artificial Neural Networks consist of three typical layers: Input layer, Hidden layer, and Output layer.

An ANN with two or more hidden layers with each layer containing a large amount of units is called a Deep Neural Network which has spawned a new field of learning called Deep Learning.

2.3 Image Segmentation Techniques

In this Section, we briefly review some well-known image segmentation methods. Each algorithm has its own merits and demerits which will be described and set it into the four groups as we categorized in Section 2.2.

2.3.1 Region-based Image Segmentation

Watershed methods are one of the classical algorithms in the field of topography. Beucher and Lantuéjoul (Beucher and Lantuéjoul, 1979) were known as the first to propose a watershed algorithm based on a immersion analogy. *Watersheds* can also be determined through gray-tone skeleton (Meyer, 1989). Luc Vincent and Pierre Soille (Vincent and Soille, 1991) proposed an algorithm for computing watersheds in digital grayscale images based on immersion simulation. The concepts of watersheds is based on visualizing an image in three dimensions, two spatial coordinates and the gray level. In these approaches image segmentation problem is solved by regarding the gradient magnitude of images as topographic surfaces in order to find the watershed lines. The pixels in which is considered as three types of points: (1) points belonging to a regional minimum; (2) points at which a drop of water, if placed at the location of any of those points, would fall with certainty to a single minimum; and (3) points at which water would be equally likely of fall to more than one such minimum. For a particular regional minimum, the set of points satisfying condition (2) is called *catchment basins* or *watershed* of that minimum. The points satisfying condition (3) form crest lines on the topographic surface and are named *divide lines* or *watershed lines*. It is supposed that a hole is punched in each regional minimum and that the entire topography is flooded from below by letting water rise through the holes at a uniform rate. When the rising water in distinct catchment basins is about to merge, a dam is built to prevent the merging. The flooding will eventually reach a stage when only the tops of the dam are visible above the water line. These dam boundaries correspond to the divide lines of the watersheds. Therefore, they are the boundaries extracted by a watershed segmentation algorithm 2.2.

One of the most notable segmentation methods is presented by Arbelaez *et al.* (Arbelaez et al., 2011). In this method, Arbelaez *et al.* propose a variant of watershed transform to generate a hierarchy of closed contours. The method combines multi-scale local cues and globalized cues via spectral clustering to create a contour map, called gPb-owt-ucm, and sets up a benchmark for edge detection and region segmentation research.

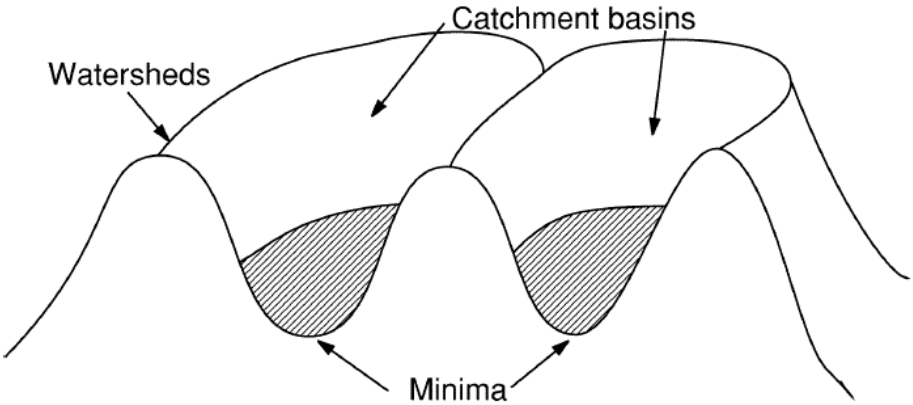


FIGURE 2.2: Watershed Illustration Chien, Huang, and Chen, 2003

2.3.2 Feature-based Image Segmentation

The *Mean Shift* based image segmentation was introduced in (Comaniciu and Meer, 2002). The *Mean Shift* is one of the most famous Feature space-based image segmentation techniques. The feature space consists of the (x, y) image location of each pixel and the pixel color in L*u*v* space or L*a*b*. The *Mean Shift* considers pixels both from the spatial and the color domains by concatenating the pixel color value and its spatial coordinates into a single vector. Then applying mean shift filtering in this domain yields a convergence point for each pixel. The image is segmented by grouping together all pixels whose convergence points are closer than a threshold h_s in the spatial domain and h_r in the range one, where h_s and h_r are parameters. In addition, the merging process can also be performed to force a constraint on minimum segment size. This method has usually good performance, and low time-consuming implementation tasks. However, it is sensitive to the bandwidth parameter h_s and h_r and often produces over-segmented results. The reason is that the mechanism for performing is based on the density estimation of the color feature for all the pixels. Some smooth changes in brightness and texture or the regularities of different colors will converge to different modes, even if they belong to the same segment visually.

2.3.3 Graph-based Image Segmentation

Considering image segmentation problem from the perspective of graph partition has interested several researchers. In this approach, the image is regarded as an undirected weighted graph in which each pixel is represented as a node in the graph and edge weights measure the similarity distance between nodes. The graph is clustered by optimizing any adequate criteria, *i.e. minimum cut, normalized cut or related variants.*

The *Normalized Cut* (Shi and Malik, 2000) criterion provides a way of integrating global image information into the grouping process. For this work, given an affinity matrix W in which each entry represents the similarity of two pixels, *Normalized Cut* tries to solve the problem using an algorithm based on the generalized eigenvectors problem of the linear system from equation 2.1.

$$(D - W)z = \lambda Dz \tag{2.1}$$

D is a diagonal matrix with its diagonal entry $D_{ij} = \sum_j W_{ij}$. Then K-means clustering is applied to obtain a segmentation into regions. Cour, *et al.* (Cour, Benezit, and Shi, 2005) approached Normalized Cuts with a variant namely Multiscale Normalized Cuts (NCuts). Sharon, *et al.* proposed an alternative to enhance the computational efficiency of Normalized Cuts. Besides, many image segmentation methods in this type were mentioned, for instance, Mumford and Shah proposed that the segmentation of an observed image u_0 is given by the minimization of functional:

$$F(u, C) = \int_{\Omega} (u - u_0)^2 dx + \mu \int_{\Omega \setminus C} |\bigtriangledown(u)|^2 dx + v|C| \tag{2.2}$$

where u is piecewise smooth in $\Omega \setminus C$ and μ, v are weighting parameters. In addition, several algorithms have been developed to minimize or simplify the energy by various approach strategies.

Another approach on image segmentation problem comes from the perspective of graph partition. In this approach, the image is regarded as an undirected weighted graph in which each pixel represented as a node in the graph and edge weights measure the similarity distance between nodes. Felzenszwalb and Huttenlocher (Felzenszwalb and Huttenlocher, 2004) attempt to partition image pixels into components. Constructing a graph in which pixel is represented as a node and edge weights measure dissimilarity between nodes (*e.g.* color differences), each node is initially placed in its own component. The internal difference of a component $Int(R)$ has been defined as the largest weight in the minimum spanning tree of R. Considering in non-decreasing order by weight of edges, each step of the algorithm merges components R_1 and R_2 connected by the current edge if weight of the edge is less than:

$$min(Int(R_1) + \tau(R_1), Int(R_2) + \tau(R_2)) \tag{2.3}$$

where $\tau(R) = k/|R|$, k is a scale parameter that can be used to set a preference for component size.

Ren and Malik (Ren and Malik, 2003) introduced superpixels in 2003 on the research of Learning a Classification Model for Segmentation. This research is very important in image segmentation discipline. Especially, with the development of complex networks theories and applications, the cooperation between superpixels and complex networks display as an absolutely

perfect partner-pair in graph-based image segmentation domain. Hence, image segmentation techniques based on community detection algorithms have been proposed and have become an interesting discipline in the literature. Complex networks analysis domain has been considered to segment images and has achieved outstanding results (Abin, Mahdisoltani, and Beigy, 2011; Li and Wu, 2015; Linares et al., 2016; Mourchid, El Hassouni, and Cherifi, 2017). The idea that community detection can be used for image segmentation offers a new perspective and prospect researches in image segmentation discipline.

Ahmad Ali Abin *et al.* (Abin, Mahdisoltani, and Beigy, 2011) proposed a new image segmentation method for color images which involves the ideas used for community detection in social networks. In the proposed method, an initial segmentation is applied to partition input image into small homogeneous regions. Then constructing a weighted networks in which the small homogeneous regions obtained initial segmentation processes are nodes of graph. Every pairs of vertices are connected through an edge and the similarity between these regions are computed through Bhattacharyya coefficient (Bhattacharyya, 1943). The RGB color space is used to compute the color histogram. Each color channel is uniformly quantized into l levels and then the histogram of each region is calculated in the feature space of $l \times l \times l = l^3$ bins. The similarity distance between two region R_i and R_j is the Bhattacharyya coefficient which is defined as 2.4:

$$B(R_i, R_j) = \sum_{u=1}^{l^3} \sqrt{H^u_{R_i}.H^u_{R_j}} \tag{2.4}$$

where H_{R_i} and H_{R_j} are the normalized histograms of region R_i and region R_j, respectively. The superscript u represents the uth element of them. Bhattacharyya coefficient is actually the cosine of the angle between two vectors $(\sqrt{H^1_{R_i}}, \cdots, \sqrt{H^{l^3}_{R_i}})^\intercal$ and $(\sqrt{H^1_{R_j}}, \cdots, \sqrt{H^{l^3}_{R_j}})^\intercal$. If two regions have similar contents, their histograms will be also similar, and hence the Bhattacharyya coefficient will be high. The higher Bhattacharyya coefficient, the higher similarity is. In this method, the weight of the edge that corresponds to two regions R_1 and R_2 is computed following the normalized Bhattacharyya similarity measure, defined as 2.5:

$$W(R_1, R_2) = \frac{exp(-\frac{1}{B(R_1,R_2)})}{\sum_{i \neq j \in R} exp(-\frac{1}{B(R_i,R_j)})} \tag{2.5}$$

where R is set of all regions.

After the creation of the network, one community detection method is applied to extract communities as segments. In this work, they are using the weighted Newman-Fast community detection algorithm (Newman, 2003) to extract the communities. Besides, the authors have suggested many existing

community detection algorithm, for instance, the algorithm of Girvan and Newman (Girvan and Newman, 2002) and the Louvain method (Blondel et al., 2008).

The results of the proposed method is compared to three well-known methods, JSEG (Deng and Manjunath, 2001), EDISON (Christoudias, Georgescu, and Meer, 2002) and MULTISCALE (Sumengen and Manjunath, 2005), and achieved much better results than others subjectively and in many cases. However, this method uses only the color histogram data as a measure of similarity, so it is an inadequate feature properties if we consider repetitive patterns of different colors in some homogeneous object. For example, in the case of zebra, cheetah and tiger.

Wenye Li (Li, 2013) proposed a graph-partition segmentation method based on a key notion from complex network analysis namely *modularity segmentation*. This approach exploits the alternative concept of modularity (Newman and Girvan, 2004) which has been successfully applied to community detection problems in complex networks (Easley and Kleinberg, 2010; Newman, 2010; Fortunato, 2010). The proposed method automatically determines the number of segments by optimizing a natural and theoretically justified criterion, eliminating the need for human intervention. In graph-partition based segmentation, each image is represented as an undirected graph $G = (V, E)$, where each vertex in $V = (v_1, v_2, ..., v_n)$ corresponds to an individual pixel, and each edge in E connects pairs of vertices. The weight on each edge, w_{ij}, is a non-negative value that measures the affinity between two vertices v_i and v_j. A higher affinity indicates a stronger relation between the associated pixels. Let $d_i = \sum_j w_{ij}$ denote the sum of affinities associated with vertex v_i, and $m = \frac{1}{2}\sum_i d_i = \frac{1}{2}\sum_{ij} w_{ij}$ denote the total sum of edge weights in the graph. Given a candidate division of vertices into disjoint groups, the modularity is defined to be the fraction of the affinities that fall within the given groups, minus the expected such fraction when the affinities are distributed randomly. The randomization is conducted by preserving the total affinity d_i of each vertex. Under this assumption, the expected affinity between two vertices v_i and v_j is $d_i d_j / 2m$, hence the corresponding modularity is $w_{ij} - d_i d_j / 2m$. Summing over all vertex pairs within the same group, the modularity, denoted Q, is defined:

$$Q = \frac{1}{2m} \sum_{ij} \left[w_{ij} - \frac{d_i d_j}{2m} \right] \delta(c_i, c_j) \tag{2.6}$$

where c_i denotes the group to which v_i belongs and $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise. An equivalent formulation can be given by defining s_{ik} to be 1 if vertex v_i belongs to group k and 0 otherwise. Then $\delta(c_i, c_j) = \sum_k s_{ik} s_{jk}$ and hence

$$Q = \frac{1}{2m} \sum_{ij} \sum_k [w_{ij} - \frac{d_i d_j}{2m}] s_{ik} s_{jk} = \frac{1}{2m} tr(S^\intercal B S) \tag{2.7}$$

where S is a matrix having elements s_{ik} and B is a modularity matrix with elements $b_{ij} = w_{ij} - d_i d_j / 2m$.

The modularity value Q always ranges in [-0.5,1]. All rows and columns of the modularity matrix B sum to zero. It means that the modularity value of unsegmented graph is always zero. The value Q is positive if the intra-group affinities exceed the expected affinities achieved at random. Thus seeking for a partition that maximizes the modularity automatically determines the appropriate number of segments to choose as well as their respective structure, and does not require the number of segments to be pre-specified.

Shijie Li, *et al.* (Li and Wu, 2015) proposed to use superpixels (Ren and Malik, 2003) and features to solve the over-segmentation problem. This strategy starts with a (over-segmented) result from an image segmentation process in which each sub-segment is represented as a superpixel, and treats the over segmentation issue by proposing a new texture feature from low level cues to capture the regularities for the visually coherent object and encode it into the similarity matrix. In addition, the similarity among regions of pixels is constructed in an adaptive manner in order to avoid over-segmentation. The main contribution of this research is the proposal of an efficient agglomerative segmentation algorithm incorporating the advantage of community detection and the inherent properties of images. A new feature is developed, histogram of state (HoS), together with an adaptive similarity matrix to avoid over-segmented problem. The authors use a publicly available code (Mori, 2005) to generate super-pixel initialization. The L*a*b* color space is chosen to achieve good segmentation performance because perceptually uniform color spaces, (for instance, L*u*v and L*a*b*), provide very good estimates of color difference (distance) between two color vectors. Perceptual differences between two colors can be estimated by considering each color as a point in a 3-dimensional space, which can be conveniently measured with the Euclidean distance. One neighborhood system has been constructed by considering the adjacent regions of this region to be its neighbors and store its neighboring regions using an adjacent list. The adjacent regions are regions that share at least one pixel with the current region. The method considers features for similarity based on color property and histogram of State (HoS) of each region. The color feature obtained by computing Mean Distance, used the mean of the pixel value for the two regions to approximate the Earth Mover's Distance 2.8, which neglects the difference of the co-variance matrices so that it is faster than computing Earth Mover's Distance, as formalized in Equation 2.9. To transform the above distribution distance into similarity measure, the authors used a Gaussian type radius basic function in Equation 2.10.

$$d_{EMD}\Big(N(\mu_1,\Sigma_1), N(\mu_2,\Sigma_2)\Big)^2 = (\mu_1-\mu_2)^\intercal(\mu_1-\mu_2) + Tr(\Sigma_1+\Sigma_2-2(\Sigma_1\Sigma_2)^{\frac{1}{2}}) \tag{2.8}$$

$$d_{MD}\Big(N(\mu_1,\Sigma_1), N(\mu_2,\Sigma_2)\Big)^2 = (\mu_1-\mu_2)^\intercal(\mu_1-\mu_2) \tag{2.9}$$

$$W_{ij}(color) = exp\Big(\frac{-dis(R_i, R_j)}{2\sigma^2}\Big) \tag{2.10}$$

where $dis(R_i, R_j)$ measures the distance between the pixel value distributions for regions R_i and R_j. Empirically, they set $\sigma = 0.08$ and use the Mean Distance Equation 2.9. The Histogram of States (HoS) are constructed for each region following the three steps below:

- Computing the gradient magnitude and 360 degree orientation information for each pixel in the image, and eliminating magnitude without any informative texture property by thresholding the magnitude information;
- Using a 5×5 sliding window, and forming a histogram of the oriented gradient by distributing 360 degree orientation into 8 bins. Then, each pixel is represented by a 8-dimensional binary orientation vector with each dimension indicating whether this orientation exists in the current sliding window. In this sense, the texture information for each pixel belongs to one of the $2^8 = 256$ states;
- Constructing a 256 dimensional histogram of states vector, each dimension counts the number of such state in the segment, then normalize it by the total number of pixels in the segment.

Cosine similarity measure 2.11 is used to measure the HoS texture feature that is represented by a 256-dimensional vector for each region, *i.e.*, for each region R_i and R_j, the HoS texture feature are h_i and h_j, $(h_i, h_j \in R^{256})$, respectively.

$$W_{ij}(texture) = cos(h_i, h_j) = \frac{h_i^{\mathsf{T}} h_j}{\|h_i\|.\|h_j\|} \tag{2.11}$$

The adaptive similarity matrix is constructed during each iteration by recomputing the similarity between regions according to Equation 2.10 and 2.11. The reason for this is that during the merging process, the regions keep expanding, and the similarity measure computed from the previous iteration might not be appropriate for the current iteration. Maintaining an adaptive similarity matrix and reevaluating the similarity between current regions offer an effective strategy to overcome the problem of the non-uniformly distributed color or texture, which should be grouped into the same segment from the perspective of human vision system. Empirically, the color feature and HoS texture feature have been combined as Equation 2.12.

$$W_{ij} = \alpha \times \sqrt{W_{ij}(texture) \times W_{ij}(color)} + (1 - \alpha) \times W_{ij}(color) \tag{2.12}$$

where $\alpha \in [0, 1]$ is a balancing parameter.

Verdoja and Grangetto (Verdoja and Grangetto, 2015) analyzed the benefits of using superpixels on a simple merging regions graph-based approach. A weighted undirected graph whose nodes are initialized with superpixels is built, and proper metrics to drive the regions merging are proposed

in this research. It is noticed that superpixels can efficiently boost merging based segmentation by reducing the computational cost without impacting on the segmentation performance. The proposed method provides accurate segmentation results both in terms of visual and objective metrics. Starting from an image I with over-segmented partition L^m composed of m regions, an undirected weighted graph $G^m = \{L^m, W^m\}$ is constructed over the superpixel set L^m, and W^m is defined as Equation 2.13.

$$W^m = \{w^m_{ij}\}, \forall i \neq j \mid l^m_i, l^m_j \in L^m \wedge A(l^m_i, l^m_j) = 1 \tag{2.13}$$

where A is adjacency function. Note that G^m is an undirected graph, so we have $w^m_{ij} = w^m_{ji}$; the weights represent the distance (or dissimilarity measure) between pair of regions $w^m_{ij} = \delta(l^m_i, l^m_j)$. At each iteration, the algorithm picks the pair of labels $l^k_p, l^k_q \in L^k$ having $w^k_{pq} = min\{W^k\}$ and merge them; *i.e.* it generates a new partition $L^{k-1} = L^k - \{l^k_q\}$ having all the pixels $x \in l^k_p \cup l^k_q$ assigned to the label l^{k-1}_p. L^{k-1} contains now just $k-1$ segments. Then, edges and corresponding weights need to be updated as well. W^{k-1} is generated according to the following formula 2.14:

$$w^{k-1}_{ij} = \begin{cases} \delta(l^{k-1}_p, l^{k-1}_j) & if \quad i = p \vee i = q \\ w^k_{ij} & otherwise \end{cases} \tag{2.14}$$

Note that w^k_{pq} is not included in W^{k-1} since it does not exist anymore. When $k = 2$ the algorithm stops and returns the full dendrogram $D = \{L^m, \ldots, L^2\}$ that can be cut to obtain the desired number of regions. In this research, the CIELAB color space and the standard CIEDE2000 color difference (Sharma, Wu, and Dalal, 2005) have been chosen. The possible functions that can be used to compute the distance δ are going to be described in detail below. Given two regions l_1 and l_2, we compute the mean values of the L*a*b* components $M_1 = (\mu_{L*,1}, \mu_{a*,1}, \mu_{b*,1})$ and $M_2 = (\mu_{L*,2}, \mu_{a*,2}, \mu_{b*,2})$, and the distance between the two labels is defined by Equation 2.15.

$$\delta_C(l_i, l_j) = \Delta E_{00}(M_i, M_j) \tag{2.15}$$

where ΔE_{00} is the CIEDE2000 color difference.

Another solution to compute the distance between two labels as the minimum of their relative Mahalanobis distances has been proposed (Mahalanobis, 1936). Given a set of n_1 pixels $l_1 = \{x_i = (x_{L*,i}, x_{a*,i}, x_{b*,i})\}^{n_1}_{i=1}$, the authors can estimate their mean $M_1 = (\mu_{L*}, \mu_{a*}, \mu_{b*})$ and co-variance as Equation 2.16

$$C_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - M_1)(x_i - M_1)^\intercal \tag{2.16}$$

Then they compute the Mahalanobis distance of any other set of n_2 pixels $l_2 = \{y_i = (y_{L*,i}, y_{a*,i}, y_{b*,i})\}^{n_2}_{i=1}$ from the estimated distribution of l_1 as

Equation 2.17

$$\Delta M(l_1, l_2) = \frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - M_1)^\intercal C_1^{-1} (y_i - M_1) \tag{2.17}$$

Since ΔM is non-symmetric, *i.e.* $\Delta M(l_1, l_2) \neq \Delta M(l_2, l_1)$, they compute the distance between two labels as the minimum of their relative Mahalanobis distances obtaining the following symmetric metric as Equation 2.18

$$\delta_M(l_i, l_j) = \{\Delta M(l_i, l_j), \Delta M(l_j, l_i)\} \tag{2.18}$$

Since merging homogeneous regions is much more important than the heterogeneous ones, in particular without crossing object boundaries, the authors also investigate a local Mahalanobis metric that aims at detecting image segment whose adjacent border look very different. This border variation consists in evaluating the Mahalanobis distance just for the pixels near the border between two adjacent image segments. More precisely, they defined b_{ij} the portion of common border between two adjacent image segments and a subset of pixels whose location is across the two adjacent regions $c_{ij} = \{x \in I | r_1 < d(x, b_{ij}) < r_2\}$, where d is the Euclidean spatial distance, and r_1 and r_2 are proper ranges. Then, the authors introduced function $B(l_i, l_j)$ that returns two new set of pixels $l'_i = l_i \cap c_{ij}$ and $l'_j = l_j \cap c_{ij}$ that represent the pixels of l_i and l_j respectively which are located close to the common border. The distance metric is defined as Equation 2.19

$$\delta_B(l_i, l_j) = min\{\Delta M(l'_i, l'_j), \Delta M(l'_j, l'_i)\} \tag{2.19}$$

where l'_i and l'_j are the two outputs of $B(l_i, l_j)$. They investigated a fourth metric based on the color histogram distance. One possible solution to measure histogram difference is the Bhattacharyya distance (Bhattacharyya, 1943), which is the general case of the Mahalanobis distance. Given two histograms h_1 and h_2 composed each by B bins, the Bhattacharyya distance is defined as Equation 2.20

$$\Delta H(h_1, h_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{h_1}\bar{h_2}B^2}} \sum_{i=1}^{B} \sqrt{h_1(i).h_2(i)}} \tag{2.20}$$

where $h(i)$ is the number of pixels in the bin i, while $\bar{h} = \frac{1}{B}\sum_{i=1}^{B} h(i)$. ΔH is computed on three channels L*a*b* independently and then the maximum value of the three is used as dissimilarity measure; this is chosen over other possibilities, like taking the mean of the three distances, as it yields higher discriminating power in finding differences just on one of the channels. In conclusion, the last dissimilarity measure between two regions l_i and l_j having respectively histograms $H_i = \{h_{L*,i}, h_{a*,i}, h_{b*,i}\}$ and $H_i =$

$\{h_{L*,j}, h_{a*,j}, h_{b*,j}\}$ is defined as Equation 2.21

$$\delta_H(l_i, l_j) = max \begin{Bmatrix} \Delta H(h_{L*,i}, h_{L*,j}) \\ \Delta H(h_{a*,i}, h_{a*,j}) \\ \Delta H(h_{b*,i}, h_{b*,j}) \end{Bmatrix} \tag{2.21}$$

The overall time required by the algorithm is linear to the size of the image, and lower complexity than both merging techniques that work on pixels and superpixels.

Youssef, *et al.* (Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi, 2016) attempted to apply community detection problems in complex networks to solve image segmentation problems and investigate a new graph-based image segmentation as well as compare to other methods. In this research, authors attempt image segmentation methods based on community detection approach like Infomap algorithm, Louvain algorithm, Fast multi-scale detection of communities based on Local Criteria (FMD), Multi-scale detection of communities using stability optimization (MD) and Stability optimization based on the Louvain method. These studies point out the potential perspective and prospect of community detection based image segmentation domain.

The image segmentation approach of Oscar A. C. Linares *et al.* (Linares et al., 2016) consist in constructing weighted networks in which the small homogeneous regions *(super-pixels)* obtained by initial segmentation processes are nodes of graph, and the edge weighs are similarity distance between these regions computed by Euclidean distance. In this strategy, a 3-dimensional descriptor formed by the components of the CIELAB color space (the mean CIELAB value of all pixels belonging to a super-pixel). Links between superpixels were created if the similarity distance was below or equal a certain threshold. One community detection method is applied to extract communities as segments based on the maximization of the modularity measure (Clauset, E J Newman, and Moore, 2005). It is improved by using a greedy optimization procedure for the algorithm proposed by Newman (Newman, 2003).

Youssef Mourchid, *et al.* (Mourchid, El Hassouni, and Cherifi, 2017) have also approached the over-segmentation problem based on superpixels associated with color and texture features. Their approach is similar to the Shijie Li's method (Li and Wu, 2015). However, they compute coefficients to adaptively update the similarity matrix W based on color feature and histogram of oriented gradients *(HOG)* texture feature. The segmentation results present a noticeable improvement in terms of accuracy and PRI metric score.

Finally, Ting Liu, *et al.* (Liu, Seyedhosseini, and Tasdizen, 2016) introduced a method for Image segmentation using hierarchical merge tree which achieved state-of-the-art region accuracy. Considering a graph, in which each

node corresponds to a superpixel, and an edge is defined if two nodes share a boundary to each other. Starting with the initial over-segmentation and trying to find a final segmentation, which is essential for the merging of initial superpixels, can be considered as combining nodes and removing edges between them. This process can be done in a iterative fashion, and corresponding edges are updated when a pair of neighboring nodes are combined in the graph. In this research, the authors reformulate the hierarchical merge tree as a constrained conditional model with global optimal solutions defined. They also developed an efficient inference algorithm instead of using the greedy tree model. It is an iterative approach to diversify merge tree generation, and to improve results via segmentation accumulation.

All in all, image segmentation methods that rely on graphs have a high cost computation in general, including community detection based image segmentation techniques. This last family is using superpixels (sub-regions) or over-segmented image segmentation results as initial inputs. Although these methods generally produced good image segmentation results as we stated above, they are also facing many problems such as high time complexity and over-segmented image segmentation results. However, there is still some space in the literature to answer the question of how to effectively generate complex network that ensures good performance of community detection. They are also some room to deal with over-segmented results by using community detection based image approaches.

2.3.4 ANN-based Image Segmentation

Recently, Artificial neural networks (ANN) have been risen as a phenomenon that fascinates several researchers in the literature. They can be used on many tasks including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis. Especially, in image segmentation problem, deep Convolutional neural networks (CNN) have made such significant contributions, and have become widely known standards such as: AlexNet, VGG-16, ResNet, GoogLeNet and FCN.

AlexNet is the name of a convolutional neural network, designed by Alex Krizhevsky, and published with Ilya Sutskever and Geoffrey Hinton (Krizhevsky, Sutskever, and Hinton, 2012). It was the pioneering deep CNN that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 with a TOP-5 test accuracy of 84.6% while the closest competitor, which made use of traditional techniques instead of deep architectures, achieved a 73% accuracy in the same challenge. The architecture of AlexNet is relatively simple, consisted of five convolutional layers, max-pooling ones, Rectified Linear Units (ReLUs) as non-linearities, three fully connected layers, and dropout. It is illustrated in Figure 2.3

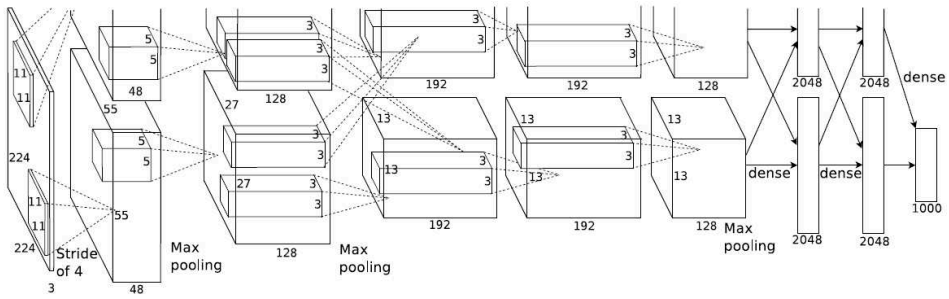


FIGURE 2.3: AlexNet Convolutional Neural Network architecture (Krizhevsky, Sutskever, and Hinton, 2012)

VGG is a CNN model introduced by the Visual Geometry Group (VGG) from University of Oxford. They proposed several models and configurations of deep CNNs (Simonyan and Zisserman, 2014), one of them was submitted to the ILSVRC-2013. It is also known as VGG-16 because it is composed by 16 weight layers. It has been famous thanks to its achievement of 92.7% TOP-5 test accuracy. This model has less parameters and more nonlinearities in between, therefore, making the decision function more discriminative and easier to train than its predecessors. It is illustrated in Figure 2.4

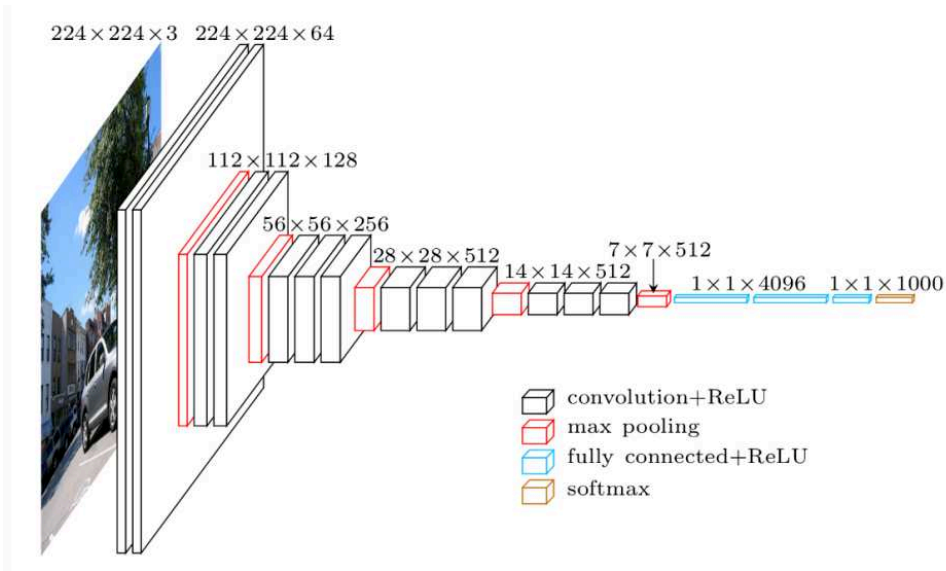


FIGURE 2.4: VGG-16 Convolutional Neural Network architecture (Simonyan and Zisserman, 2014)
Source https://www.jeremyjordan.me/convnet-architectures/

GoogLeNet is a CNN introduced by Szegedy *et al.* (Szegedy et al., 2014) which won the ILSVRC-2014 with a TOP-5 test accuracy of 93.3%. The architecture of GoogLeNet is composed of 22 layers and a newly introduced

building block called *inception* module, shown in Figure 2.5. This new approach gives more proof that CNN layers could be stacked in more ways than a typical sequential manner. In fact, those modules containing a Network in Network layer, a pooling operation, a large-sized convolution layer, and small-sized convolution layer. They are computed in parallel and followed by 1×1 convolution operations to reduce dimensionality. Because those modules support this CNN puts special consideration on memory and computational cost by dramatically reducing the number of parameters and operations.

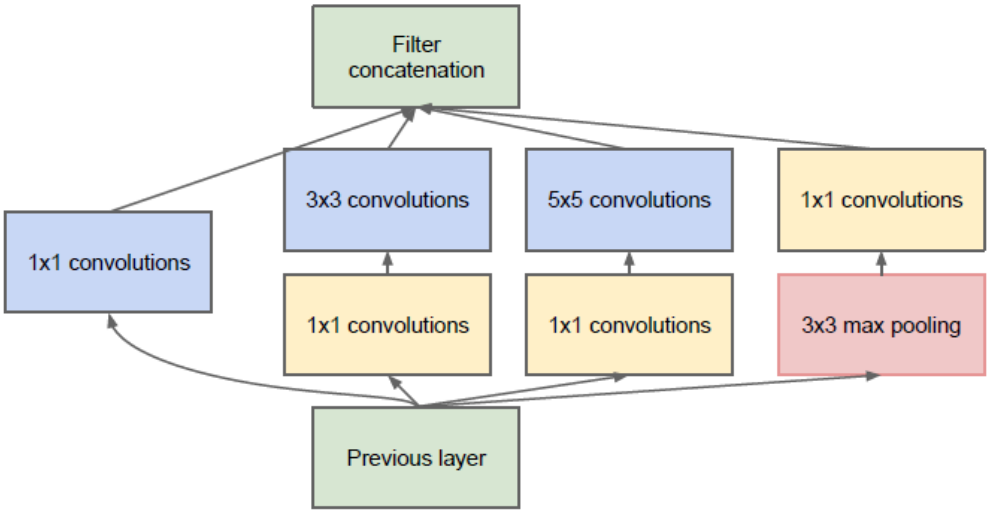


FIGURE 2.5: Inception module with dimensionality reduction from the GoogLeNet architecture (Szegedy et al., 2014)

Microsoft's ResNet (He et al., 2016) is a highlighted remarkable CNN thanks to won ILSVRC-2016 with 96.4% accuracy. This network is well-known because of its depth and the introduction of residual block. It is composed by 152 layers and the residual blocks address the problem of training a really deep architecture by introducing identity skip connections and then layers can copy their inputs to the next layer. The residual block is shown in Figure 2.6. The intuitive idea of this approach is that it ensures that the next layer can learn differential new things from what the input has already encoded. It is noticed that both the output of the previous layer and its unchanged input will be provided next layer. Furthermore, this kind of connections help overcoming the vanishing gradients problem.

FCN is a CNN introduced by Long *et al.* (Long, Shelhamer, and Darrell, 2015) for semantic segmentation. Fully Convolutional Network (FCN) is built only from locally connected layers, such as convolution, pooling and upsampling. No dense layer is used in this kind of architecture. This reduces the number of parameters and computation cost. Also, the network can work regardless of the original image size, without requiring any fixed

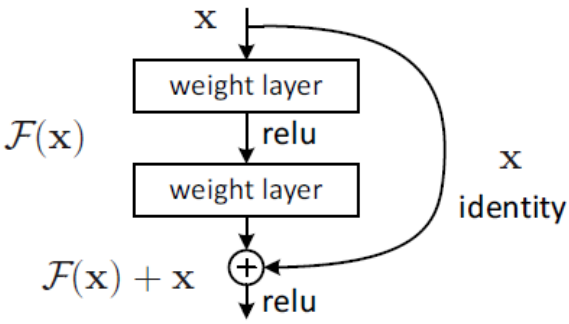


FIGURE 2.6: Residual block from the ResNet architecture (He et al., 2016)

number of units at any stage, given that all connections are local. To obtain a segmentation map (output), segmentation networks usually have 2 parts : Downsampling path and Upsampling path. The downsampling path is used to extract and interpret the context (what), while the upsampling path is used to enable precise localization (where). Furthermore, to fully recover the fine-grained spatial information lost in the pooling or downsampling layers, we often use skip connections. A skip connection is a connection that bypasses at least one layer. Here, it is often used to transfer local information by concatenating or summing feature maps from the downsampling path with feature maps from the upsampling path. Merging features from various resolution levels helps combining context information with spatial information.

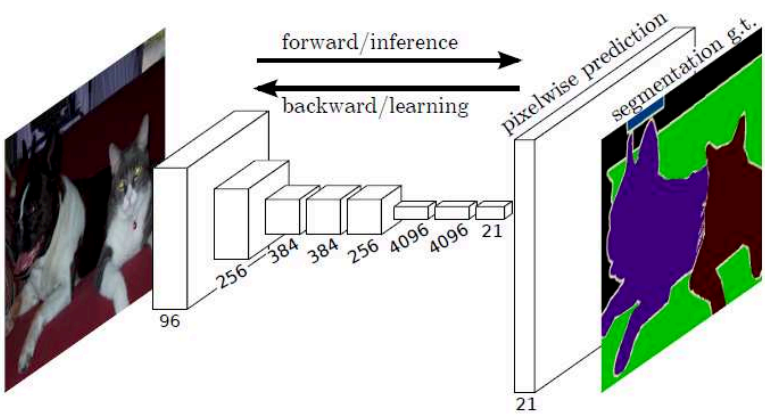


FIGURE 2.7: FCN can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation (Long, Shelhamer, and Darrell, 2015)

In short, modern image segmentation techniques are powered by deep learning technology. Many computer vision tasks require intelligent segmentation of an image, to understand what is in the image and enable easier

analysis of each part. The ANN-based image segmentation techniques can reach to semantic segmentation or instance segmentation. Nowadays, image segmentation techniques use models of deep learning for computer vision to understand exactly which real-world object is represented by each pixel of an image that we used to believe that it is at a level unimaginable only a decade ago.

Chapter 3

Complex Networks and Images

3.1 Introduction

The city of Königsberg in Prussia (now Kaliningrad, Russia) was set on both sides of the Pregel River, and included two large islands - Kneiphof and Lomse - which were connected to each other and to the two mainland portions of the city, by seven bridges. The question arose as to whether it was possible to devise a walk through the city that would cross each of those bridges once and only once, or not. The Seven Bridges of Königsberg becomes a historically notable problem in mathematics. Its negative resolution by Leonhard Euler in 1736 (Euler, 1736) laid the foundations of what we know today as graph theory or network science. Since then, this branch of discrete mathematical properties have led to a deep understanding of many complex systems.

In the 20th century graphs have also become extremely useful to represent a wide variety of systems in different areas. Indeed, biological, social, technological, and information networks can be modeled as graphs, and graph analysis has become crucial to understand the features of these systems. For instance, social network analysis started in the 1930's and has become one of the most important topics in sociology (Wasserman and Faust, 1994; Scott, 2000). In computer science, graph theory has been applied to analyze the World Wide web (Albert, Jeong, and Barabási, 1999; Barabasi and Albert, 1999) among many other systems.

In recent times, the computer revolution has provided scholars with a huge amount of data and computational resources to process and analyze these data - the Big Data era. The size of real networks one can potentially handle has also grown considerably, reaching billions of vertices for web-graphs or online social networks. The need to deal with such a large number of units has produced a deep change in the way graphs are approached (Albert and Barabási, 2002; Dorogovtsev and Mendes, 2003; E. J. Newman, 2003; Pastor-Satorras and Vespignani, 2004; Boccaletti et al., 2006; Barrat, Barthlemy, and Vespignani, 2008). Complex network is one of the typical approaches that will be described in Section 3.2.

3.2 Complex Networks

A complex network is a graph (network) whose topological structure cannot be trivially described. It comprises properties that emerge as a consequence of global topological organization of the system. Complex network structures describe various systems of high technological and intellectual importance, such as the Internet, the World Wide Web, financial, social, neural and communication networks. One property that has attracted particular attention is the community structure of these networks. Communities, also called clusters or modules, are groups of vertices which probably share common properties and (or) play similar roles within a network. A more graph theoretic definition states that communities are groups of nodes with dense internal connections and sparse connections with nodes of other communities.

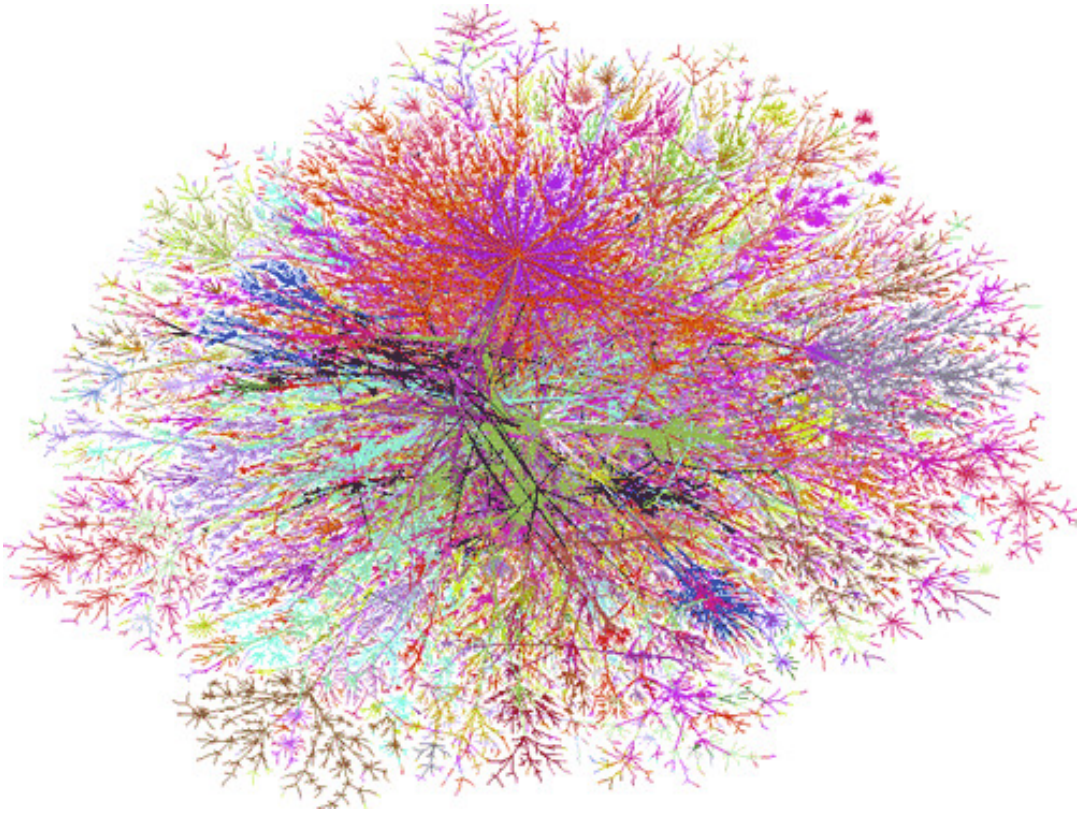


FIGURE 3.1: Networks of interest vary from the Internet to the World Wide Web to scientific collaborations networks. *(Source: https://newmedialab.cuny.edu/project/complex-networks/)*

3.2.1 Community Detection and Modularity

The problem of community detection is usually defined as finding the best partition (or covering) of a network into communities of densely connected

nodes, with the nodes belonging to different communities being only sparsely connected. Community identification is essential in networks, since they may represent some functional modules in the network. For example, a community in an online social network (where users are nodes and friendship connections are edges) may represent a group of friends that share common hobbies; a community in a scientific papers citation network (where articles are nodes and citations to other articles are edges) may represent a certain research domain.

To unfold the interconnection of the nodes in a network, community detection algorithms aim to find a partition of the network so that every part represents a given community. Several algorithms have been proposed to find good partitions in term of a reasonably fast way. These algorithms can be divided into some main types such as, divisive algorithms that detect inter-community links and remove them from the network, agglomerative algorithms that merge similar or close nodes and more generally optimization methods that are based on the maximization of an objective function (Fortunato, 2010).

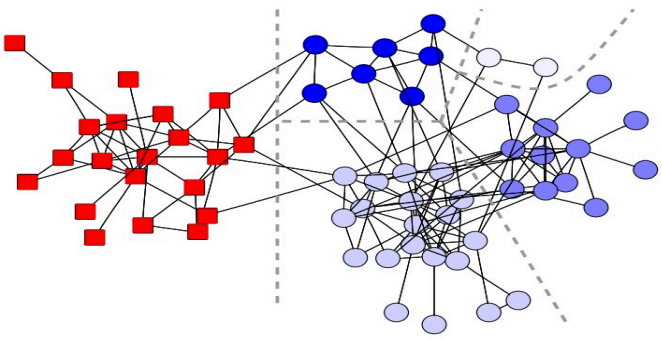


FIGURE 3.2: Community structure in the social network of bottle-nose dolphins population extracted using the algorithm of Girvan and Newman (Girvan and Newman, 2002). The squares and circles denote the primary split of the network into two groups and the circles are further subdivided into four smaller group as shown (Newman, 2004).

Modularity was introduced as an objective function by Newman (E J Newman, 2004) for the analysis of weighted networks. For a weighted network G with the weighted adjacency matrix A, the modularity Q is defined by:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{3.1}$$

where A_{ij} represents the weight between node i and node j; $m = \frac{1}{2} \sum_{ij} A_{ij}$ represents the total weights of the network; $k_i = \sum_j A_{ij}$ is the weighted degree of node i; c_i is the community label to which node i belongs; $\delta(c_i, c_j)$ is 1

if $c_i = c_j$ and 0 otherwise. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Therefore, modularity has been widely used to measure the quality of partitions, to compare algorithms with each other and, in some cases, it is also the function to be maximized by some algorithms.

The concepts of modularity and community detection have been popular in the complex networks discipline thanks to their application in large scale networks. Recently, the idea that using the modularity and community detection to approach image segmentation problem is more and more attracting for researchers from the field of image processing and computer vision. This strategy relies on the use of community detection algorithms in order to cluster pixels that belong to the same group of information. Louvain method (Blondel et al., 2008) is a classical community detection algorithm that will be detailed in Section 3.2.2.

3.2.2 Louvain Algorithm

The Louvain method (Blondel et al., 2008) is a hierarchical greedy algorithm designed to maximize the modularity on (weighted) graphs. Louvain method uses modularity increase to modify the partition and to detect the communities. It relies on the modularity increase caused by moving an isolated node i into a community C that can be computed by Equation

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right] \tag{3.2}$$

where $\sum_{in}$ is the sum of the weights of the links inside C, $\sum_{tot}$ is the sum of the weights of the links incident to nodes in C, k_i is the sum of the weights of the links incident to node i, $k_{i,in}$ is the sum of the weights of the links from i to nodes in C and m is the sum of the weights of all the links in the network. A similar expression is used in order to evaluate the change of modularity when a node i is removed from its community. These two expressions can therefore evaluate the change of modularity if node i is removed from its community and then moved into a neighboring community. It has to be noted that the computation of this variation of modularity does not depend on the whole partition.

More specifically, the Louvain algorithm consists of two phases that are repeated iteratively. Initially, every node is a singleton community. Next, during phase 1, all nodes are considered one by one and each node is moved in one of its neighboring community, including its own one, so as to increase as much as possible the modularity (using the previous formula). This process is repeated until no individual movement of node can improve the modularity. Therefore this first phase stops when the modularity reaches a local maximum. Then, phase 2 consists in building a new graph whose nodes are the communities found during the first phase. To build this new graph, edges

between nodes of the same community lead to self-loops while the weight of edges between new nodes are computed by the sum of the weights of edges between nodes in the corresponding two communities. An example of those two steps is presented in figure 3.3.

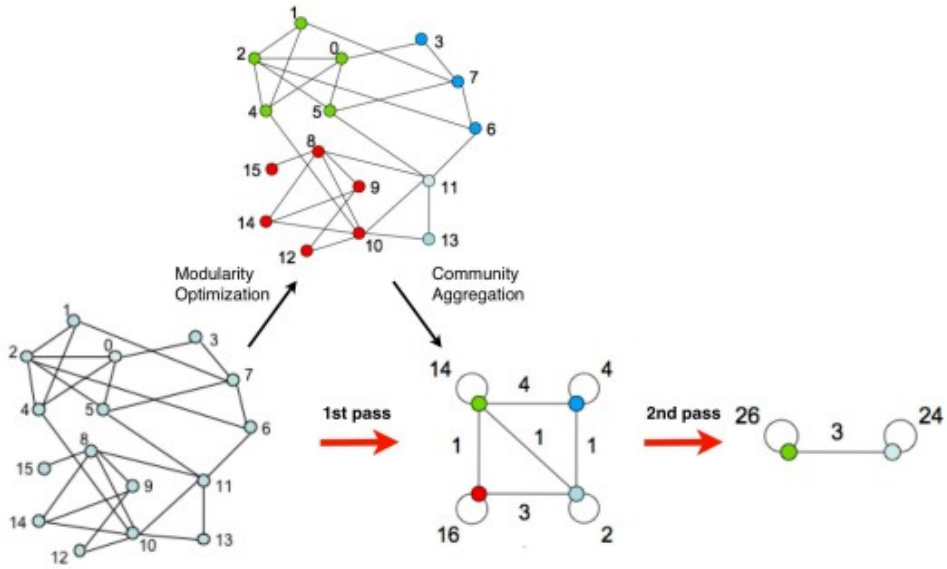


FIGURE 3.3: Process of community detection for Louvain method (Blondel et al., 2008): four communities are identified after phase 1 (modularity optimization) and Louvain creates a network with four nodes and matching weights after phase 2 (community agglomeration). When this process is repeated during the second pass, two communities are identified and a two-nodes network is created. Finally no more merging can occur and the process stops.

In Section 3.3, we introduce the way to build complex networks from images, and propose some typical patterns that improve the Louvain method performance in particular and community detection algorithms in general. Furthermore, we also evaluate the efficiency of these patterns through three image segmentation strategies detailed in the next Section.

3.3 Generating a complex network from an image

Complex networks can easily be used to model images. One image is represented as an undirected weighted graph $G = (V, E, w)$, where V is a set of vertices ($V = \{v_1, v_2, ..., v_n\}$) and E is a set of edges $E = \{e_1, e_2, ..., e_k\}$. Each vertex $v_i \in V$ corresponds to an individual pixel and similarity/closeness of pixels are modeled as edges: an edge $e_{ij} \in E$ connects vertices v_i and v_j. A weight on each edge, w_{ij}, is a nonnegative value that measures the affinity between v_i and v_j. The higher affinity is the stronger relation between the

pixels.

In our research, each node in the graph represents a pixel and edge weights are defined as:

$$w_{ij} = \begin{cases} 1 & if \quad d_{ij}^c \leq t \text{ for all color channels } c \\ nil & otherwise \end{cases} \tag{3.3}$$

where t is a threshold, d_{ij}^c is a measure of the similarity of pixels i and j intensity for color channel c among R, G and B (for red, green and blue respectively). It is defined by $d_{ij}^c = \left| I_i^c - I_j^c \right|$ where I_i^c and I_j^c represent the intensity of pixel i and j respectively for channel c.

However, a simple uniform image of size 1000 by 1000 would generate a graph with one million nodes and all possible connections between these nodes (that is close to 500 billion edges). Therefore we will only create edges between nodes that are close to each other, i.e., the corresponding pixels are close in the image. In our research, we attempt to build complex network using several proximity patterns and investigate their advantages and disadvantages.

3.3.1 Pattern definition

Figure 3.4 illustrates the patterns that we have considered.

In Figure 3.4 (a), for a given pixel, links towards other pixels are created if and only other considered pixels are inside N neighboring pixels for row and column directions, N being a parameter. Plus, equation 3.3 still applies and all distances d_{ij}^c of color channels must be lower than t for the edge to be considered. In this case, the weight is assigned $w_{ij} = 1$. The complex networks created with that typical pattern have several advantages in conjunction with Louvain algorithm thanks to reducing the number of edges of networks (each node connects to at most $2.N$ other nodes). It helps the Louvain algorithm to execute more quickly and accurately. This typical pattern also allows to set the N neighboring pixels parameter to a larger value in order to generate complex networks whose farther distance pixels are considered. Empirically, we tested Louvain with the complex networks created by this typical pattern and recorded proofs that both implementing cost and accurate score are very good. The obtained results of Louvain method are also adequate as input data for the agglomeration phase of homogeneous regions.

In Figure 3.4 (b), for a given pixel, links towards other pixels are created if and only if other considered pixels are inside $N \times N$ neighboring pixels. Plus all distances d_{ij}^c of color channels must be lower than t for the edge to be considered. In this case, the weight is assigned $w_{ij} = 1$. The complex networks which are created by that pattern contain much more information but are

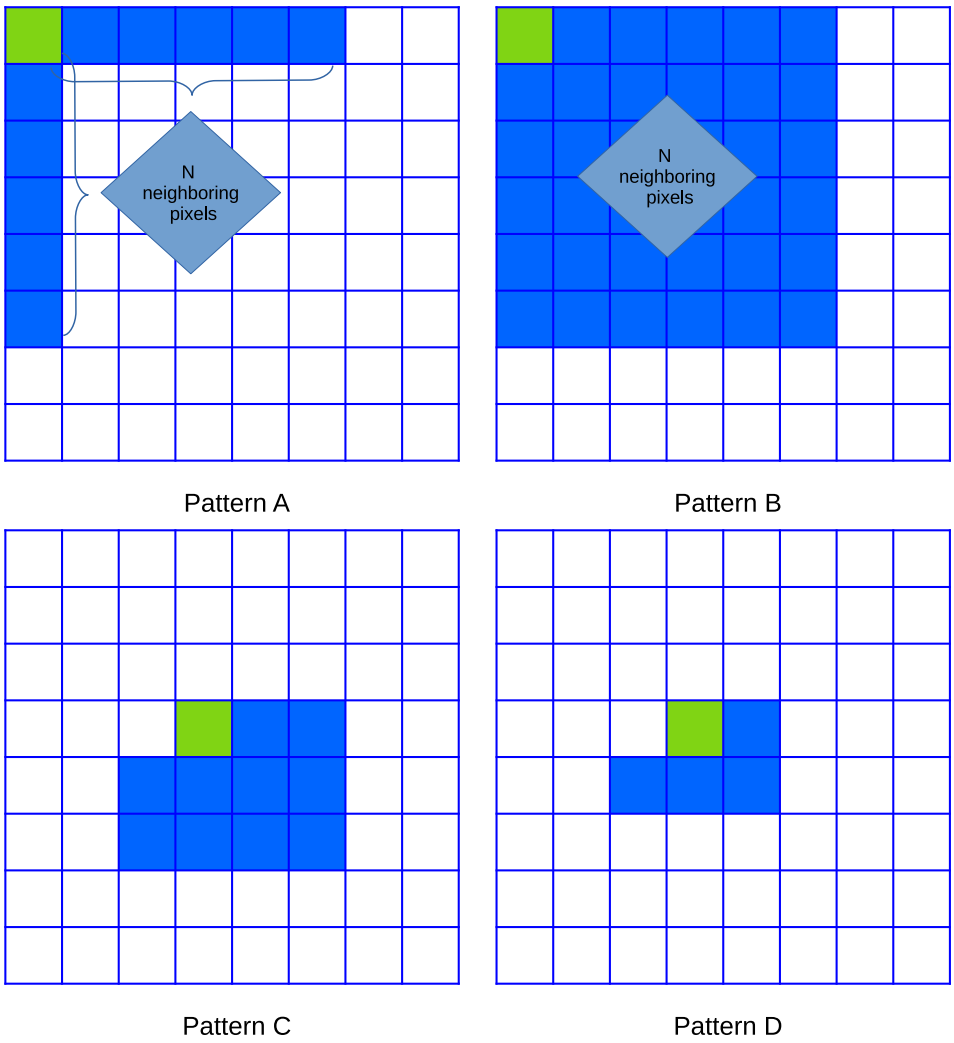


FIGURE 3.4: Four typical patterns for creating Complex Networks

also much denser that with the previous pattern. This typical complex network produces good super-pixels by applying the Louvain algorithm and excellent quality of input data for merging processes that will be detailed in Section 3.4.1. These generated complex networks, jointly used with the Louvain algorithm, produces more accurate results. However, this typical patterns show its disadvantages in large images and the N neighboring pixels parameter cannot too large because the generated complex networks is denser and it makes Louvain method to perform much slower. Empirically, we attempted to test Louvain method with the complex networks which is created by this typical pattern with several values of N and recorded proofs that its execution is very slow but that it gives accurate image segmentation results. The obtained results of Louvain method performing are good quality for the agglomeration phase of homogeneous regions.

We also tested the typical patterns illustrated in Figure 3.4 (c)(d) that

avoid duplicate links. With a proper choice of size N, these patterns and networks offer many benefits as redundancies of connections between pixels to pixels are reduced but they obtained pretty good image segmentation results. These typical patterns can deal with large images since they can generate suitable complex networks that can apply effectively in Louvain algorithm in order to detect community.

3.3.2 Pattern evaluation

To have a general visual comparison of the patterns, we tested these typical patterns to create complex networks, and experiment on some well-known datasets (hereafter the BSDS300 dataset). Then, we make an evaluation and comparison of these patterns. We compare all three kinds of measurement metrics proportions: PRI, VI and SC in a Chart (we will come back later to these metrics). Figure 3.5 presents the comparison that is produced based on *A new image segmentation approach based on the Louvain algorithm*, short name *Lv-ara method* (Nguyen, Coustaty, and Guillaume, 2018a) while Figure 3.6 comes from *An Efficient Agglomerative Algorithm Cooperating with Louvain Method for Implementing Image Segmentation*, is named *Lv-ahr* (Nguyen, Coustaty, and Guillaume, 2018b). Finally, Figure 3.7 comes from *A Combination of Histogram of Oriented Gradients and Color Features to Cooperate with Louvain Method Based Image Segmentation*, short name *Lv-mhr*, detailed in (Nguyen, Coustaty, and Guillaume, 2019).

As can be seen, the VI figures preserve pattern trend in all Figures represented for our three methods tested with four Patterns. Figure 3.5 and Figure 3.6 show that both PRI and VI values remain stable with all patterns while the SC values present a little different between two methods with the same kinds of Patterns. Figure 3.5 presents the highest SC proportion in Pattern A, the second highest fall in Pattern B, the next ranking is the Pattern C, and the worst is Pattern D. In Figure 3.6, the first ranking is both Pattern B and C, the third and the last one are Pattern D and Pattern A. However, the differences pointed out between those Patterns are not remarkable in these two Charts.

All in all, for four Patterns described above, the Pattern A denotes its outstanding characteristics. It is not only good for Louvain method's performance but is also stable and got very high score with all metrics in almost all our experimental results. Pattern B usually produces good results (from the metrics scores point of view) while Pattern C and D bring benefits in Louvain method performances. In particularly, Figure 3.7 shows that Pattern B wins the highest all measurement metrics scores to other Patterns. Pattern A occupied the second highest PRI and VI scores. Pattern C and Pattern D rank in the third and fourth position, respectively. The SC scores do not change in the three other Patterns with a score of 0.74.

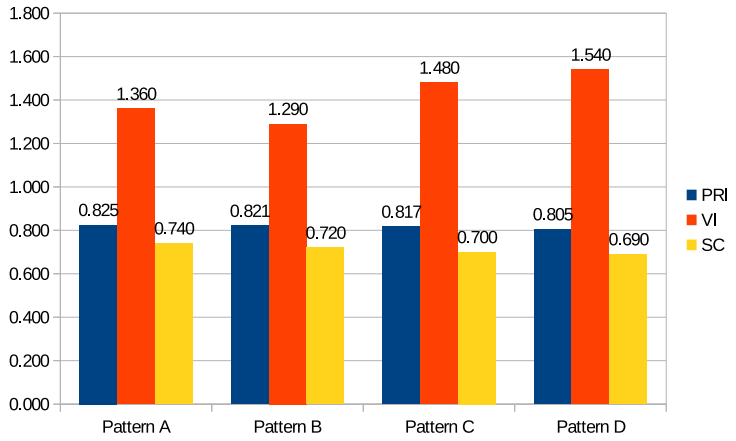


FIGURE 3.5: The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-ara method (Nguyen, Coustaty, and Guillaume, 2018a)

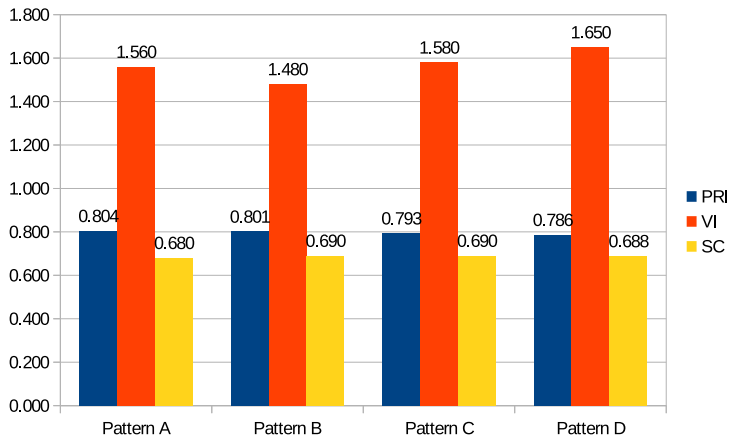


FIGURE 3.6: The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-ahr method (Nguyen, Coustaty, and Guillaume, 2018b)

3.4 Community Detection Based Image Segmentation

Although the Louvain algorithm has been widely applied for image segmentation purposes, the problem of over-segmentation has not yet been solved.

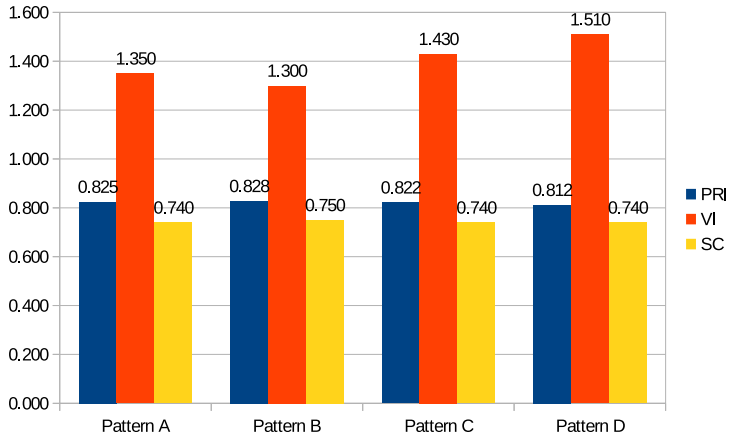


FIGURE 3.7: The comparison scores of the three segmentation metrics for four typical patterns for creating Complex Networks on BSDS300 Validation Set based on Lv-mhr method (Nguyen, Coustaty, and Guillaume, 2019)

The basic principle of image segmentation based on modularity is to produce many homogeneous regions that can belong to a single object. In order to solve over-segmentation problems, homogeneous regions must be combined whenever necessary. We will present all the steps of this approach in this section.

3.4.1 Aggregation of Regions Algorithm

We propose in this sub-section a solution to the problem of over-segmentation. Given an over-segmented image, consisting of a set of homogeneous regions R. The algorithm is based on the fact that if a region is small or similar to a neighboring region, both are merged. Let *regthres* be the threshold that defines the notion of small region, below which the regions are merged. The function $C(R_i)$ returns the number of pixels in the region R_i, $\delta(R_i, R_j)$ is 1 if R_i and R_j are adjacent regions and 0 otherwise, and *simthres* is the similarity distance threshold that defines the limit over which similar regions are merged, regardless of their size. The proposed algorithm can merge homogeneous regions to generate better segmentation results. It is described as *Pseudo-code* below:

Algorithm ARA

Input: A set of regions $R = \{R_1, R_2, ..., R_n\}$
01: **for** each region $R_i \in R$ **do**
02: **for** each region $R_j \in R; (i \neq j)$ **do**
03: **if** $\delta(R_i, R_j) = 1$ **then**
04: **if** $C(R_i) <= regthres\ OR\ C(R_j) <= regthres$ **then**

```
                    Merge region R_i and region R_j
                else
                    Compute similarity d(R_i,R_j) between region R_i
                     and region R_j
                    if d(R_i,R_j) >= simthres then
                        Merge region R_i and region R_j
                    end if
                end if
            end if
        end for
    end for
```

Output: The set of regions $R = \{R_1, R_2, \ldots, R_k\}; \forall k \leq n$.

In the ARA algorithm, the similarity of two adjacent regions R_i and R_j is computed by cosine similarities of a pair of vectors u_i, u_j that represent the two considered regions, as indicated in equation 3.4.

$$d(R_i, R_j) = Cosine(u_i, u_j) = \frac{u_i^{\mathsf{T}} u_j}{\|u_i\|.\|u_j\|} \tag{3.4}$$

However, the feature vectors used in the similarity measure remain a parameter and in this study we propose six possible feature vectors. The two first features rely on color information as this corresponds to the primary straightforward feature for image segmentation (Li and Wu, 2015; Mourchid, El Hassouni, and Cherifi, 2017). The first feature proposed (in section 4.2.1) relies on a distance using the mean and the standard deviation of color while the second solution proposes to compute a distance between reduced histogram of color channels (in section 4.2.2). The rest of four features used in this study relies on the combination between color features and texture features based on the Histogram of Oriented Gradients (HOG) feature (in section 4.3.1). These are listed HOGMeanSD, BHMF, MFO and MAFO features. The detailed description of these features will be displayed in Chapter 4.

3.4.2 Noise Removal

In our method, one main technique to emphasize is the noise removal process. As mentioned above, the results obtained with the Louvain algorithm consist of over-segmented results, which decreases *Probabilistic Rand Index (PRI)* score when it is evaluated. We proposed and evaluated strategies for better results and higher *PRI* scores. These strategies are divided into two phases of noise removal: the preliminary noise removal and getting the final segmentation.

The preliminary removing noise process is a crucial part of our algorithm. It merges the small regions neglected when considering the similarity $d(R_i, R_j)$ between two regions. This merging process uses two adjacent

regions and the number of pixels within must be less than a threshold *regthres* in our algorithm. Empirically, we set the threshold for small regions to *regthres* = 200 pixels for testing and evaluating on the validation dataset. On a small test sample this value gives the best results but more general experimentation and an automatic selection of the threshold would be a improvement. More specifically, the modularity and therefore the Louvain method suffers from the resolution limit problem that tends to favor typical sizes for communities (Fortunato and Barthélemy, 2007) and instability issues (Seifi et al., 2013) that are both directly related to this problem. Furthermore, while the order in which the merging of small regions is performed can influence the result, we have not studied this question and in our method the order is purely random.

Although the aggregation of the regions has reduced noise, the final image segmentations still contain noise that must be suppressed to produce clean images. Various noise reduction techniques have been introduced in the literature. The main image noise reduction techniques are supported by Open Source Computer Vision Library, such as: Averaging, Gaussian Blurring, Median blurring and Bilateral Filtering. In our method, Median Blurring technique has been applied to smooth images and achieve effective image segmentation.

Chapter 4

Extraction of Precise Features

4.1 Introduction

In image processing and computer vision, features are information extracted from images that are relevant for solving the computational tasks related to certain applications and which have a much lower dimension than the original image. Features corresponds to specific structures in the image such as points, edges, colors, gradients or objects. The features may also be the result of a general neighborhood operation or feature detection applied to the image. Other examples of features include motion in image sequences, shapes defined in terms of curves or boundaries between different image regions, or properties of such a region. Finally, features are generally represented as graphs or vectors/histograms for their efficiency in term of computational cost.

The concept of feature is very general and the choice of features may depend heavily on the specific problem studied. Basically, there are two primary types of features extracted from images depending on the application. These are the local and global features that constitute two approaches to object recognition. Global features describe the image as a whole to capture the entire object and therefore describe the whole image with single vector. Global features include contour representations, shape descriptors, and texture features and local features represents the texture in an image patch. Shape Matrices, Invariant Moments, Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005) and Co-HOG (Watanabe, Ito, and Yokoi, 2008) are some examples of global descriptors. On the contrary, local features describe areas of the image (key points in the image) and several points are available on the image which makes them more robust. SIFT (Lowe, 2004), SURF (Bay, Tuytelaars, and Van Gool, 2006), LBP (He and Wang, 1990), BRISK (Leutenegger, Chli, and Siegwart, 2011), MSER (Matas et al., 2002) and FREAK (Alahi, Ortiz, and Vandergheynst, 2012) are some examples of local descriptors. Despite of robustness in local feature, global feature is still useful for applications where rough segmentation of the object is available. Global texture features and local features provide different information about the image because the support on which texture is computed varies.

Features are sometimes called descriptors. Global descriptors are typically used for image retrieval, object detection and classification, while local descriptors are used for object recognition or identification. There is a large difference between detection and identification. Detection consists of looking for the existence of something or object, for example to determine if an object exists in an image or a video. In contrast, recognition consists of finding identity, for example, recognizing a person or thing from an object.

In short, for low level applications such as object detection and classification, global features are used and for higher level applications such as object recognition, local features are used. Combination of global and local features improves the accuracy of recognition with the side-effect of computational overheads.

In the following we will first present the color features and how we can extract them, then the texture features and finally how we can combine both to get better results.

4.2 Color features

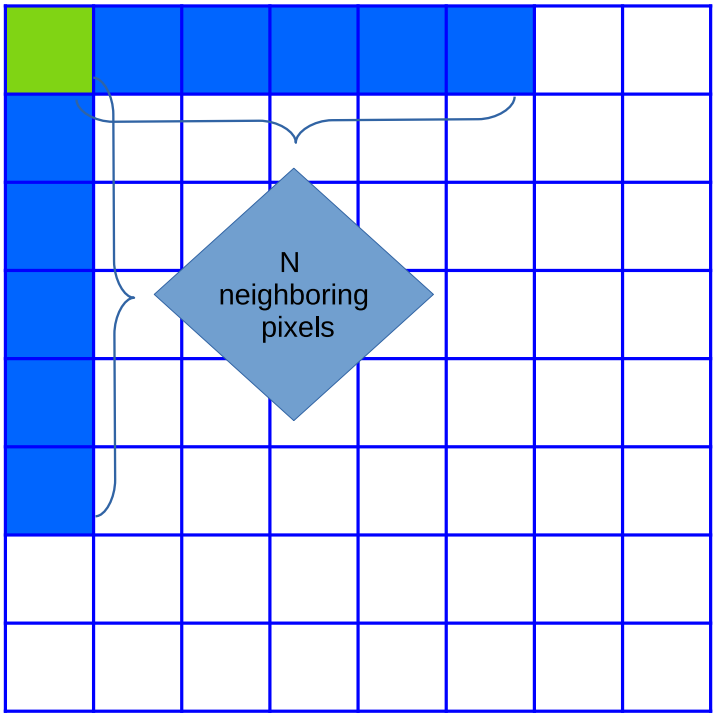
Color is one of the low-level features that is the subject of many researches as it is a simple and important feature from human perspective for image segmentation. To capture different aspects of the color, different color spaces are used such as RGB, L*a*b*, YUV, HSV and XYZ. (C Gonzalez and E Woods, 2002)

4.2.1 Mean and Standard Deviation of Color Approach

From this variety of color spaces, we selected the RGB color space in this thesis and the article from which parts of this chapter are taken from, namely *A new image segmentation approach based on the Louvain algorithm*, short name *Lvara* (Nguyen, Coustaty, and Guillaume, 2018a), since it is more meaningful and efficient in our work. With RGB color images, we have a flexibility in the way to compute the similarity distance between regions. We can consider every individual channel of color to build the complex network and compute distance similarity to manage the combination of regions in the algorithm ARA on Section 3.4.1 in Chapter 3.

In order to set up for this research, we build the complex networks based on pattern A that is described in Section 3.4 on Chapter 3. We recall that with this pattern, a node can be connected only to neighboring nodes located on the same line or the same column and at most at a given distance (20 in this chapter).

A complex network in this chapter is an undirected weighted graph $G = (V, E)$, where V is a set of vertices ($V = \{v_1, v_2, ..., v_n\}$) and E is a set of edges



Pattern A

FIGURE 4.1: Representation of Pattern A used in this chapter (Nguyen, Coustaty, and Guillaume, 2018a)

$E = \{e_1, e_2, ..., e_k\}$. Each vertex $v_i \in V$ corresponds to an individual pixel and similarity/closeness of pixels are modeled as edges: an edge $e_{ij} \in E$ connects vertices v_i and v_j. A weight on each edge, w_{ij}, is a nonnegative value that measures the affinity between v_i and v_j. Edge weights are defined as:

$$w_{ij} = \begin{cases} 1 & if \quad d_{ij}^c \leq t \text{ for all color channels } c \\ nil & otherwise \end{cases} \tag{4.1}$$

where t is a threshold, d_{ij}^c is a measure of the similarity of pixels i and j intensity for color channel c (among R, G and B). It is defined by $d_{ij}^c = \left| I_i^c - I_j^c \right|$ where I_i^c and I_j^c represent the intensity of pixel i and j respectively for channel c.

Therefore, for a given pixel, links towards other pixels are created if and only other considered pixels are inside 20 neighboring pixels for rows and columns directions. Plus all distances d_{ij}^c of color channels must be lower than t *(t=15 is the best value obtained experimentally)* for the edge to be considered. In this case, the weight is assigned $w_{ij} = 1$.

Then, we applied Louvain method for detecting pixels communities in

this network. As it is stated on Section 3.4, the image segmentation results obtained with the Louvain method segmentation process are over-segmented. To address this problem, we applied the algorithm ARA on Section 3.4.1 in Chapter 3. We proposed using Mean and Standard deviation feature for this approach.

For each region, we therefore construct a vector consisting of Mean and Standard deviation for every channel of colors as expressed in formulas 4.2 and 4.3. Each region is represented by a 6-dimensional vector, namely *MeanSD feature*.

$$Mean(R) = \frac{\sum_{i=1}^{n} C_i}{n} \tag{4.2}$$

$$SD(R) = \sqrt{\frac{\sum_{i=1}^{n} (C_i - Mean(R))^2}{n}} \tag{4.3}$$

where C_i is the color value channel of pixel i in the image and n is the number of pixels in the region R. Means and standard deviations are computed for all three color channels.

FIGURE 4.2: Vector representation for a region using RGB color space.

Mean	SD	Mean	SD	Mean	SD
0	1	2	3	4	5

First of all we display some image segmentation results of our algorithm in Figure 4.3. This figure highlights the fact that Louvain algorithm produces over-segmented images (middle line). It also shows that the proposed algorithm offers good results and produces sizable regions for all selected images and that our algorithm can aggregate homogeneous neighboring regions successfully even if pixels inside each region are dissimilar (bottom line), see Figure 4.3. However, we determine that the color feature hardly achieve instance image segmentation in case of repetitive patterns of different colors in many homogeneous objects or the quite different color parts inside an entity (for instance, Cheetah, Tiger and Zebra), see Figure 4.4. Because we consider the similarity of regions based on color property while they are quite dissimilar although they belong to one object. We attempt to solve this issue in some other approaches using a combination of color and texture properties

as below.

FIGURE 4.3: Image segmentation results: *Top:* Original images. *Middle:* after the execution of the Louvain method. *Bottom:* after the merging of homogeneous regions using both Mean and Standard deviation feature and the ARA algorithm.

The comparative experimental results of this research were carried out below according to two perspectives of evaluations: qualitative and quantitative. For qualitative evaluations, we present some segmentation results of related researches and our own results, it is shown in Figure 4.5.

From a quantitative point of view, we evaluated the image segmentation results using the *Probabilistic Rand Index* (PRI) that allows to compare a given segmentation with multiple ground-truth segmentations. We applied this measure on the validation set from the Berkeley segmentation dataset, and obtained the score of 0.819, while the ground-truth (segmentation made by human) got a score of 0.87 (Arbelaez et al., 2009; Mourchid, El Hassouni, and Cherifi, 2017; Li and Wu, 2015). Detailed results are given in table 4.1.

The evaluation results reflect the success of our agglomeration process for homogeneous regions. Our method exceeds all previous graph-based algorithms in term of PRI scores. We present some extensive experimental results on Section 5.4 in Chapter 5.

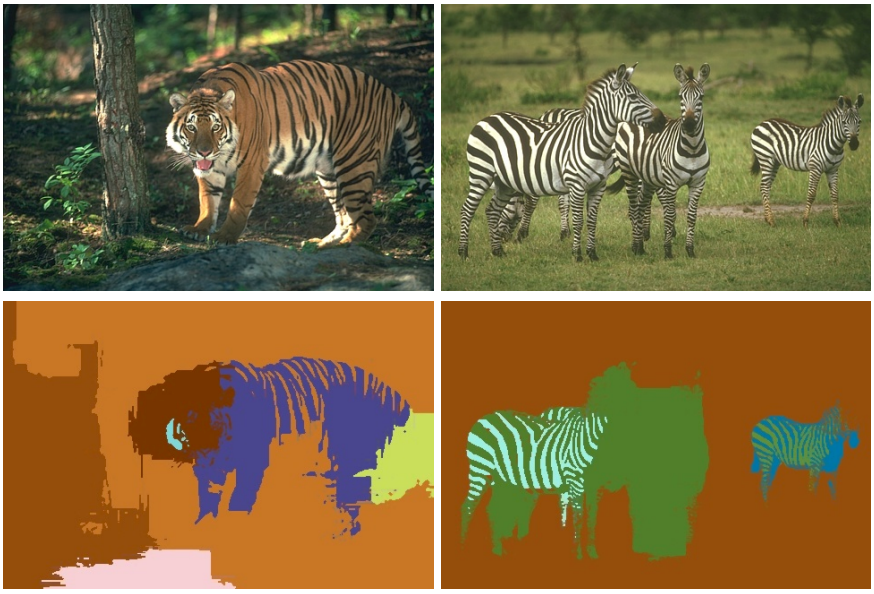


FIGURE 4.4: Image segmentation results: *Top:* Original images. *Middle:* after the execution of the Louvain method. *Bottom:* after the merging of homogeneous regions using both Mean and Standard deviation feature and the ARA algorithm.

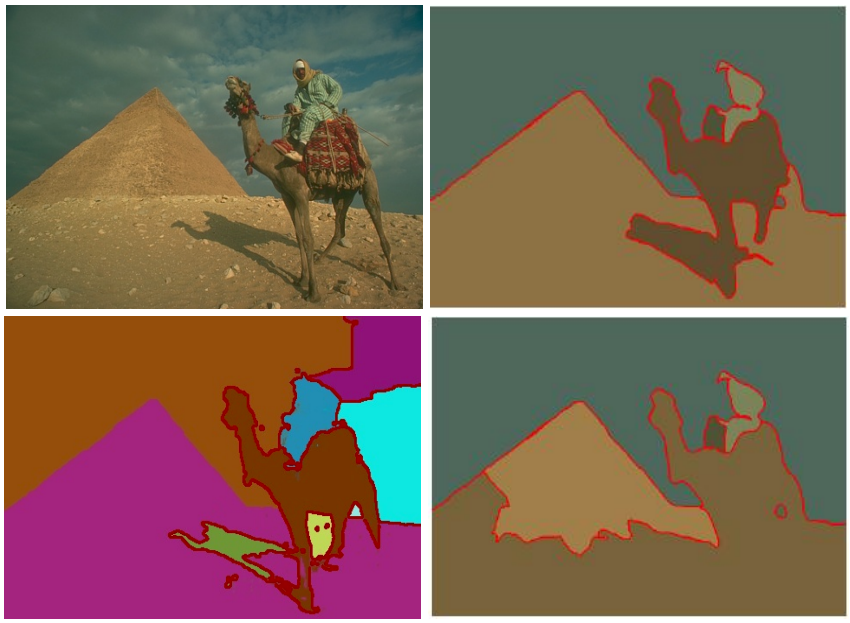


FIGURE 4.5: Visual Segmentation results of some methods: *Top left:* Original image. *Top right:* Segmentation result of HoS method (Li and Wu, 2015). *Bottom left:* Our segmentation result using mean and standard deviation of color features *Bottom right:* Segmentation result of HOG method. (Li and Wu, 2015; Nguyen, Coustaty, and Guillaume, 2018a)

Methods	PRI
Human	0.87
Our method: ARA + mean and standard deviation	**0.819**
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS (HOG))*	0.811
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.81
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG)*	0.803
Comaniciu and Meer, 2002 *(Mean Shift)*	0.78
Li and Wu, 2015 *(L*a*b (HoS))*	0.777
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.77
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.77
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.75
Li and Wu, 2015 *(RGB (HoS))*	0.749

TABLE 4.1: Comparative results of ARA algorithm plus MeanSD feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

4.2.2 Reduced Histogram of Color Approach

We also proposed a new feature based on a reduced histogram of color in cooperation with the ARA algorithm from Section 3.4.1 in Chapter 3 in order to merge homogeneous regions from the over-segmented step to obtain better image segmentation results. This proposal, presented in *An Efficient Agglomerative Algorithm Cooperating with Louvain Method for Implementing Image Segmentation*, is named *Lv-ahr* (Nguyen, Coustaty, and Guillaume, 2018b).

For this study, we used the same process to generate the network as mentioned on Section 4.2.2, and apply the Louvain method for undirected graph. For each color channel, we compute an image histogram (values between 0 and 255) and distribute it into 16 equivalently-sized (RH) bins and then aggregate values in each bin. Then, with these 16 bins for every color channel, we build the 48-dimensional vector for each region, namely *RH feature*. It is illustrated on Figure 4.6.

To present experimental results for this research, we arrange its results into two perspectives of evaluation: qualitative and quantitative evaluation.

For these qualitative results, we can see that the proposed algorithm offers good results and produces sizable regions for all selected images. Figure 4.7 presents some segmentation results of our algorithm on some images. Our algorithm can aggregate homogeneous neighboring regions successfully even if pixels inside each region are dissimilar.

For qualitative evaluations, we present some images of the comparison segmentation results as Figure 4.8.

FIGURE 4.6: Creating 48-dimensional vector for each region

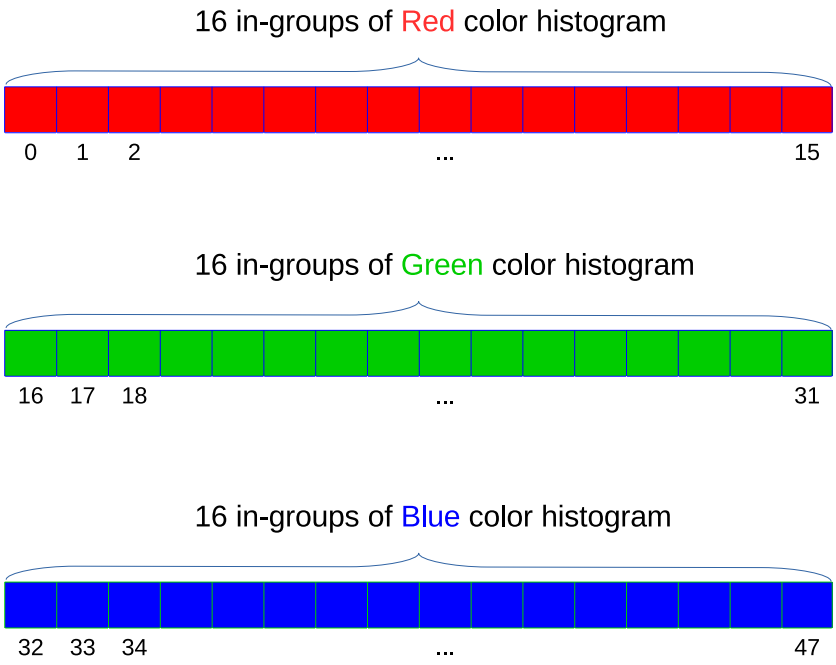


FIGURE 4.7: Image segmentation results: *Left:* Original images. *Middle:* after the execution of the Louvain method. *Right:* after the merging of homogeneous regions using both reduced histogram of colors and the ARA algorithm (Nguyen, Coustaty, and Guillaume, 2018b)

From a quantitative point of view, we evaluated the image segmentation results by the *Probabilistic Rand Index* (PRI) by comparing a test segmentation

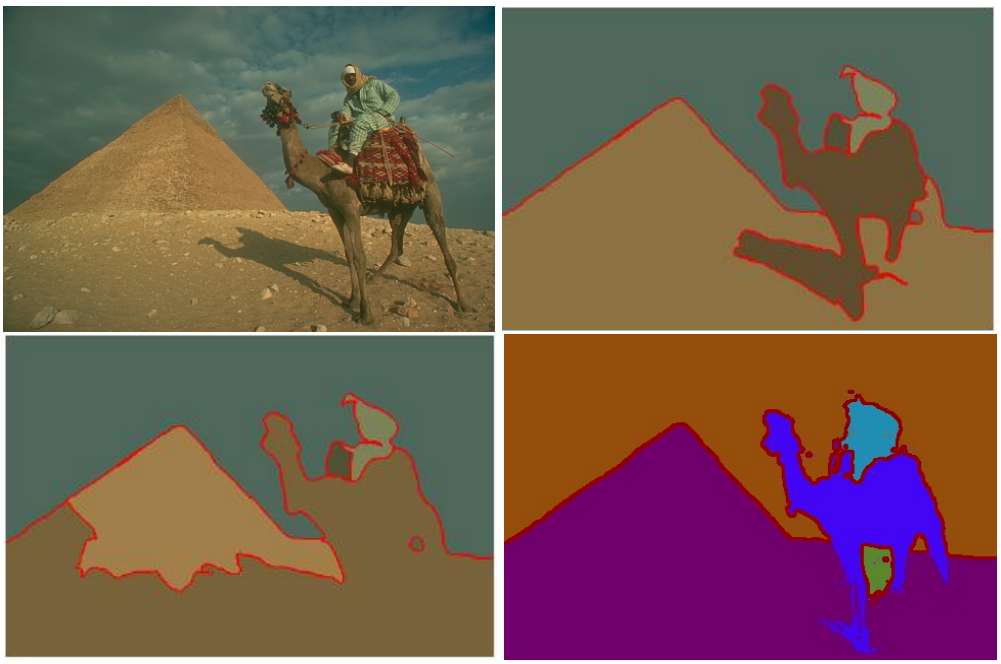


FIGURE 4.8: Visual Segmentation results of some methods: *Top left:* Original image. *Top right:* Segmentation result of HoS method (Li and Wu, 2015). *Bottom left :* Segmentation result of HOG method (Li and Wu, 2015). *Bottom right:* Our segmentation result by ARA and RH feature (Nguyen, Coustaty, and Guillaume, 2018b)

with multiple ground-truth images. We applied this measure on the validation set from the Berkeley segmentation dataset, and obtained the score of 0.80, while the ground-truth (segmentation made by human) got a score of 0.87 (Arbelaez et al., 2009). Detailed results are given in Table 4.2.

The evaluation results reflect the success of our agglomeration process for homogeneous regions. Our method has close performance with gPb-owt-ucm and Fast multi-scale (HOG), and exceeding all the rest algorithms in term of PRI scores. We also implement intensive and extensive experimental results for this precise feature to network and survey the efficiency of community detection with the undirected weighted graph function of Louvain method. Detail for this research results will be displayed on Section 5.4 in Chapter 5.

The two proposals using colors show good results from a qualitative point of view and succeed to merge homogeneous regions. However the quantitative results vary from one solution to the other and the simple mean and standard deviation offer better results. This leaves room for improvement using more complex features.

Methods	PRI
Human	0.87
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.81
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS (HOG))*	0.81
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG)*	0.80
Our method: ARA + reduced histogram of colors	**0.80**
Li and Wu, 2015 *(L*a*b (HoS))*	0.78
Li and Wu, 2015 *(RGB (HoS))*	0.75
Comaniciu and Meer, 2002 *(Mean Shift)*	0.78
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.77
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.77
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.75

TABLE 4.2: Comparative results of ARA algorithm plus RH feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

4.3 Texture features

The texture features are important low-level features which are used to define the characteristics of repetitive patterns inside an image or a region. While color corresponds to an important feature from the human perception point of view, textures are essentials to obtain a parallel analysis and to segment images not only from one category of information which could lead to some limitations (two different regions could share the same mean color value). The texture is commonly found in natural scenes, containing both natural and man-made objects. For example, sand, stones, grass, leaves, bricks and many more objects create a textured appearance in the images. It appears then interesting to study those features in our work to see if they could propose interesting results.

4.3.1 Texture-based feature using HOG Approach

In this feature, we compute the histogram of oriented gradients for grayscale images: given a grayscale image I, we extract the gradient magnitude and orientation using the 1D centered point discrete derivative mask 4.4, and 4.5 in the horizontal and vertical directions to compute the gradient values.

$$D_X = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{4.4}$$

and

$$D_Y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{4.5}$$

We obtain the x and y derivatives using a convolution operation 4.6, 4.7.

$$I_X = I * D_X \tag{4.6}$$

and

$$I_Y = I * D_Y \tag{4.7}$$

The magnitude of the gradient is presented by the equation 4.8, and the oriented gradient given by Equation 4.9. In this feature, however, the oriented gradient is used to build the similarity feature vectors.

$$|G| = \sqrt{I_X^2 + I_Y^2} \tag{4.8}$$

$$\theta = \arctan \frac{I_Y}{I_X} \tag{4.9}$$

In the implementation, we use the function *atan2* that returns a value in the interval $(-\pi, \pi]$. The orientation of gradient at a pixel is $\theta = atan2(I_Y, I_X)$ radians. The angle degrees are transformed by $\alpha = \theta * 180/\pi$, that give values in the range (-180, 180] degrees. To shift into unsigned gradient we apply formula 4.10 and obtain the range of the gradient [0, 360). The orientation of gradient is put into 9 equally-sized bins that represent 9 elements in the similarity feature vectors. For each region, we compute the *HOG feature* as the percentage of oriented gradient bins.

$$\alpha = \begin{cases} \alpha, & if \quad \alpha \geq 0 \\ \alpha + 360, & if \quad \alpha < 0 \end{cases} \tag{4.10}$$

We did not perform any experiments using only the HOG feature, but we instead associated this texture feature with the previous color features. Indeed, only one individual color feature or texture feature could not offer a good implementing for ARA algorithm as we stated above.

4.4 Combination of colors and textures

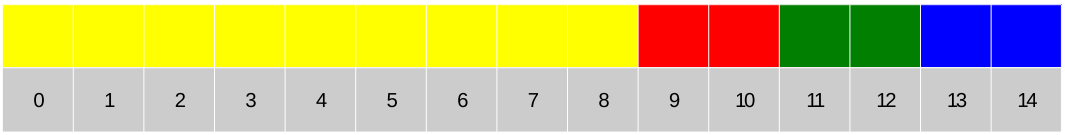
As we have stated previous, both color features and texture features are very precious properties for a variety of applications in the image segmentation field. However, individual color features or texture features are not quite robust in images that are composed of repetitive patterns of different colors in many homogeneous objects. In this research, we present some combinations of color feature and texture feature properties.

To assess the efficiency of this feature, we still build an undirected graph based on the pattern A and use Louvain method to find communities.

4.4.1 Mean and Standard deviation plus HOG feature

The new feature, namely *HOGMeanSD feature*, has been built as a 15-dimensional vector feature for each region. It consists of 6 elements from Mean and Standard deviation (color feature) and 9 elements provided by HOG feature (see in Figure 4.9), the solution comes from *A Combination of Histogram of Oriented Gradients and Color Features to Cooperate with Louvain Method Based Image Segmentation*, short name *Lv-mhr*, detailed in (Nguyen, Coustaty, and Guillaume, 2019).

FIGURE 4.9: Creating 15-dimensional vector for each region



We present the experiment of this method on the BSDS300 dataset and assess the experimental results using the PRI image segmentation metric. We also present the results both from a qualitative and quantitative point of view.

For qualitative evaluations, we present some images of the segmentation results in Figure 4.10 and Figure 4.11, collected from the dataset BSDS300. For these qualitative results, we can see that the proposed algorithm offers good results and produces sizable regions for all selected images. The HOG-MeanSD feature cooperating with ARA algorithm can aggregate homogeneous neighboring regions successfully even if pixels inside each region are dissimilar.

From a quantitative point of view, we evaluated the segmentation results using PRI evaluation metric presented in section 5.3 by comparing a test segmentation with multiple ground-truth images. We applied the PRI evaluation metric on the BSDS300 dataset, detailed results are given in Table 4.3.

The evaluation results shows that our algorithm is successful since it exceeds all previous graph-based algorithms in terms of PRI scores. Empirically, the threshold range for the agglomeration process is the *simthres* only taking range from 0.940 to 0.999 *(with 0.005 intervals)*. The best results are recorded when the value of cosine similarity distance greater or equal to 0.995 ($simthres \geq 0.995$). Cosine similarity distance domain that offers good image segmentation results in our algorithm fall into [0.990, 0.999].

Moreover, we extend the experimental evaluation in all three datasets that are presented in Chapter 5.

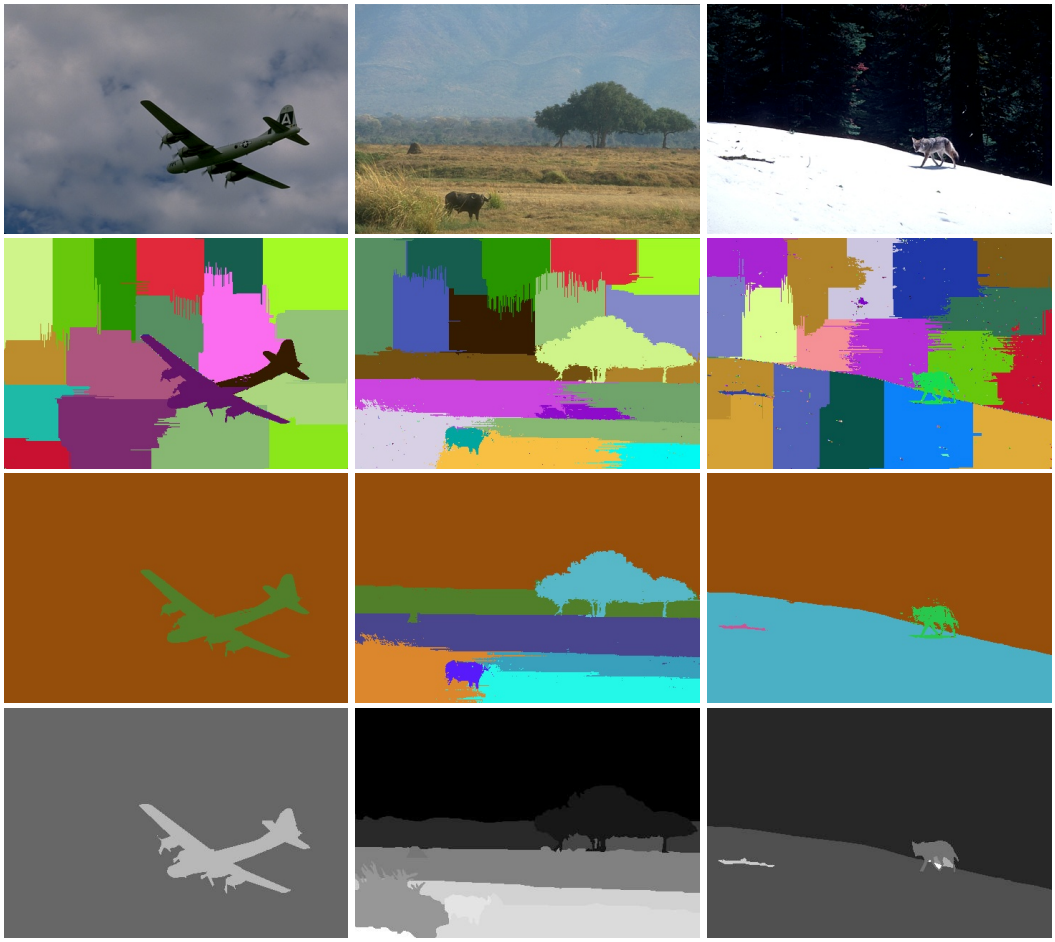


FIGURE 4.10: Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both HOGMeanSD feature and the ARA algorithm. *Fourth line:* Ground-truth.

4.4.2 Balancing HOG and MeanSD Features

It is noticed that the HOGMeanSD feature method above (Nguyen, Coustaty, and Guillaume, 2019), the total of elements in each representative vector consist of 15 elements including 9 elements coming from HOG feature and 6 elements coming from MeanSD feature. According to many previous research, the color property is a worthwhile feature for image segmentation (Li and Wu, 2015; Mourchid, El Hassouni, and Cherifi, 2017; Nguyen, Coustaty, and Guillaume, 2018a; Nguyen, Coustaty, and Guillaume, 2018b). Especially, it is more essential when segmenting images using community detection. We therefore propose a new strategy which gives more weight to color features compared to HOG. This new composed feature, namely *BHMF feature*, consists of a 27-dimensional feature vector for every region that includes 18 elements from Mean and Standard deviation $(2(Mean\ and\ SD)\ for\ each\ channel\ \times\ 3\ channels \times 3\ times)$ and 9 elements that coming from HOG feature. It is illustrated in Figure 4.13.

FIGURE 4.11: Image segmentation results: *First:* Original images. *Second:* after the execution of the Louvain method. *Third:* after the merging of homogeneous regions using both HOG-MeanSD feature and the ARA algorithm. *Fourth:* Ground-truth.

Methods	PRI
Human	0.870
Our algorithm: ARA + HOGMeanSD feature	**0.822**
Nguyen, Coustaty, and Guillaume, 2018a *(Lv-ara)*	0.819
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS(HOG))*	0.811
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.810
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG))*	0.803
Nguyen, Coustaty, and Guillaume, 2018b *(Lv-ahr)*	0.80
Comaniciu and Meer, 2002 *(Mean Shift)*	0.780
Li and Wu, 2015 *(L*a*b (HoS))*	0.777
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.770
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.770
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.750
Li and Wu, 2015 *(RGB (HoS))*	0.749

TABLE 4.3: Comparative results of ARA algorithm plus HOG-MeanSD feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

We build complex network based on Pattern A and tested on the mentioned BSDS300 dataset. The experimental results denote that our method produces efficient image segmentation results compared to the state of the art. The detailed assessment of this method is displayed in Table 4.4 and Figures 4.14 and 4.15. The extended experimental evaluation will be implemented in all three datasets and three evaluation metrics that are presented in Chapter 5.

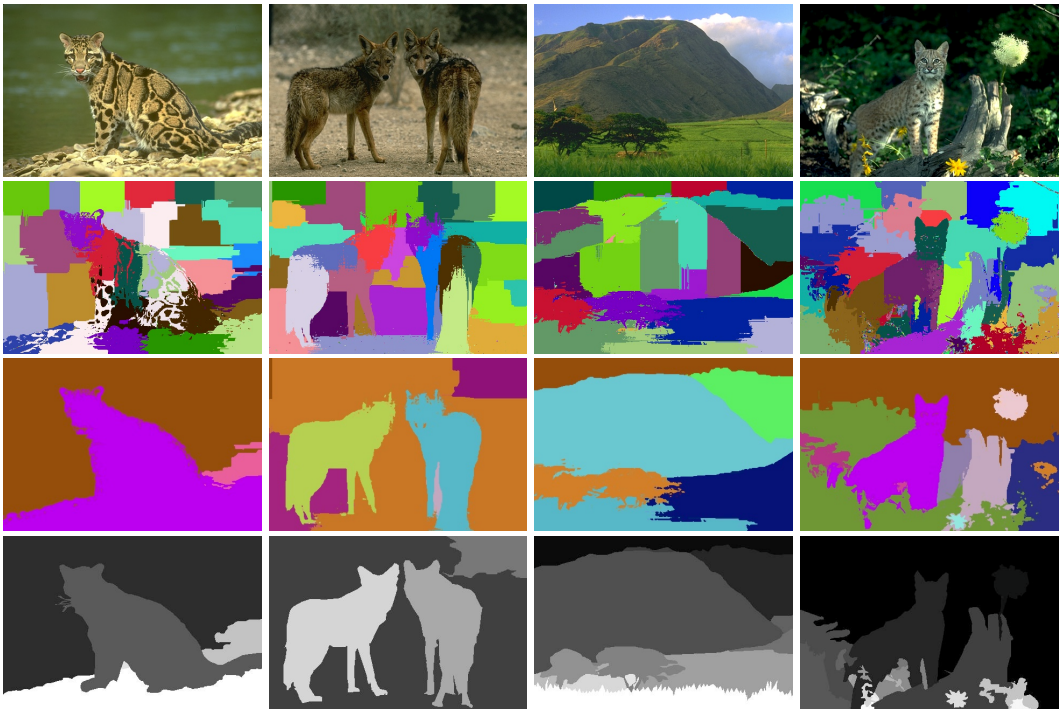
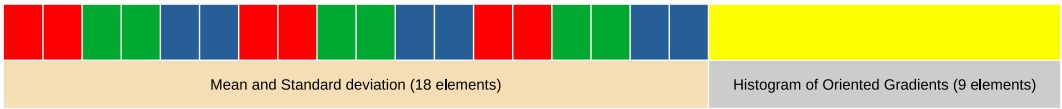


FIGURE 4.12: Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth.

FIGURE 4.13: Creating 27-dimensional vector for each region



4.4.3 Merging all features in one

In general, each feature displays some strengths and weaknesses in some particular cases of images segmentation. Our hypothesis suppose that those strengths and weaknesses complement one another and by combining them we could get better results. For this strategy, we make a combination of RH, MeanSD and HOG features into one feature, namely *MFO feature*. To do that, we build a 63-dimensional feature vector for every region including 48 elements selected from Histogram of color (16 for each color channel) and 6 elements from Mean and Standard deviation (2 for each color channel) plus 9 elements coming from HOG feature. It is illustrated in Figure 4.16. We tested this mix of features on the BSDS300 dataset with Pattern A. The experimental results are presented in Table 4.5. Although the experimental results show that the combination of those features did not give better results, we

Methods	PRI
Human	0.870
Our algorithm: ARA + BHMF feature	**0.828**
Nguyen, Coustaty, and Guillaume, 2019 *(Lv-mhr)*	0.822
Nguyen, Coustaty, and Guillaume, 2018a *(Lv-ara)*	0.819
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS(HOG))*	0.811
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.810
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG))*	0.803
Nguyen, Coustaty, and Guillaume, 2018b *(Lv-ahr)*	0.800
Comaniciu and Meer, 2002 *(Mean Shift)*	0.780
Li and Wu, 2015 *(L*a*b (HoS))*	0.777
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.770
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.770
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.750
Li and Wu, 2015 *(RGB (HoS))*	0.749

TABLE 4.4: Comparative results of ARA algorithm plus BHMF feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

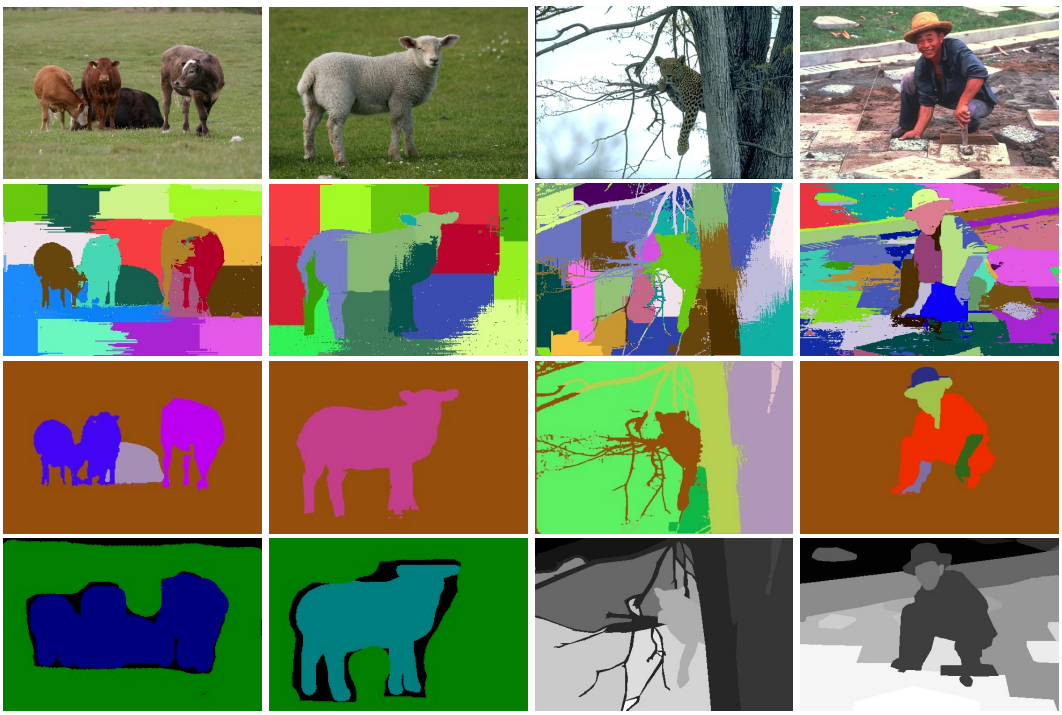


FIGURE 4.14: Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth.

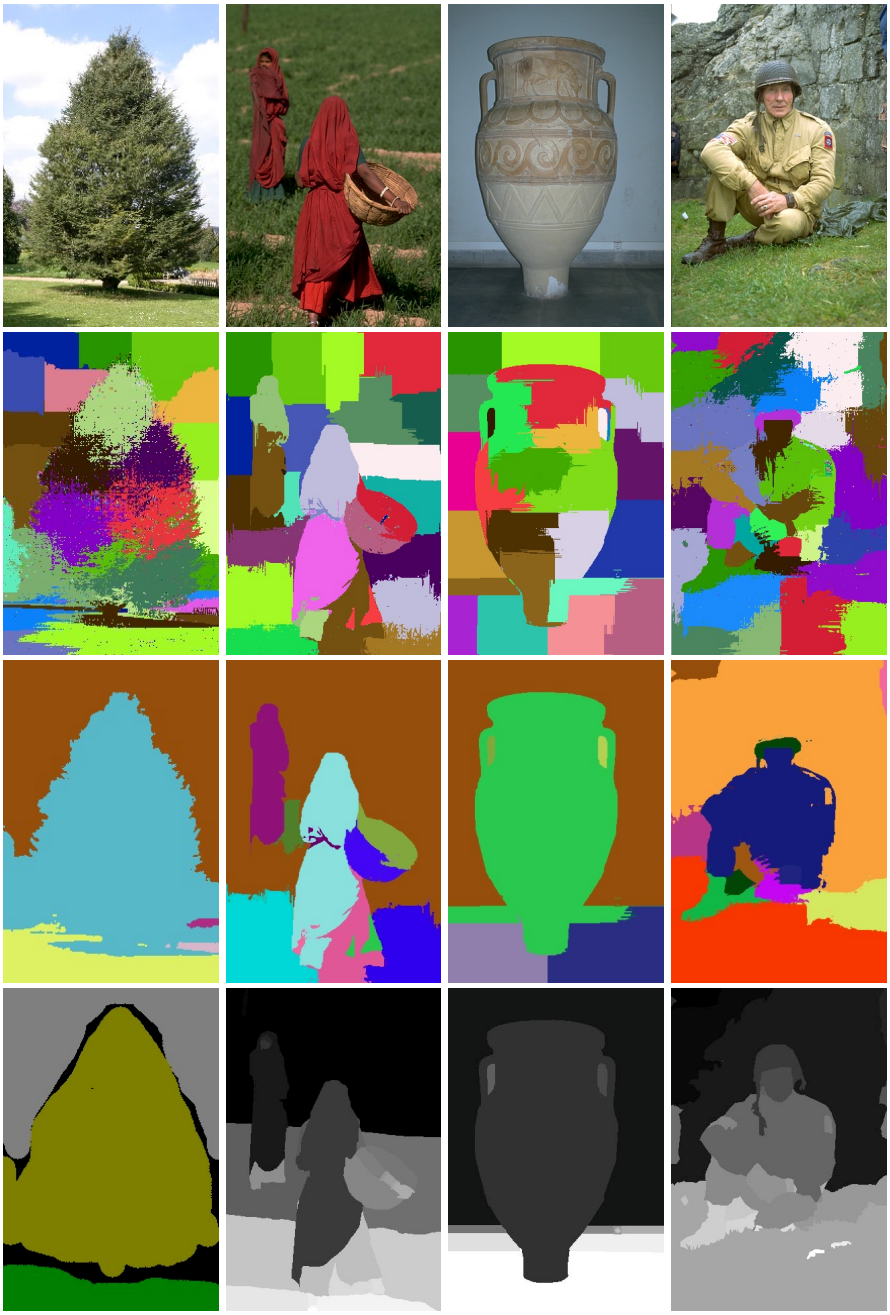


FIGURE 4.15: Image segmentation results: *Top:* Original images. *Second line:* after the execution of the Louvain method. *Third line:* after the merging of homogeneous regions using both BHMF feature and the ARA algorithm. *Fourth line:* Ground-truth.

believe that different features can offer a good complement to each other. As we described above, the reduced histogram of color (RH feature) gave worst result than both the mean and standard deviation (MeanSD feature) and the combination of the mean and standard deviation and histogram of oriented gradients (HOGMeanSD feature) methods. While the total of the RH feature

occupied 48 elements in total 63 elements of this vector feature. In addition, the MeanSD feature also belongs to color property, as a consequence, it seems that mixing all features add confusion and is not effective. However, we will implement this strategy with more extended experiments all the patterns with all the data sets, and evaluation image segmentation metrics listed as above. They are presented on Section 5.4 in Chapter 5.

FIGURE 4.16: Creating 63-dimensional vector for each region

Methods	PRI
Human	0.870
Nguyen, Coustaty, and Guillaume, 2019 *(Lv-mhr)*	0.822
Nguyen, Coustaty, and Guillaume, 2018a *(Lv-ara)*	0.819
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS(HOG))*	0.811
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.810
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG))*	0.803
Nguyen, Coustaty, and Guillaume, 2018b *(Lv-ahr)*	0.800
Our algorithm: ARA + MFO feature	**0.795**
Comaniciu and Meer, 2002 *(Mean Shift)*	0.780
Li and Wu, 2015 *(L*a*b (HoS))*	0.777
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.770
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.770
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.750
Li and Wu, 2015 *(RGB (HoS))*	0.749

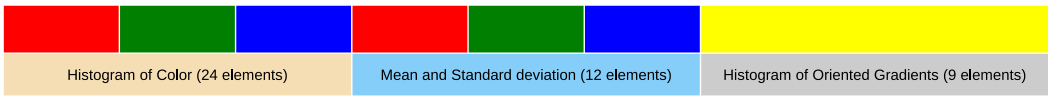
TABLE 4.5: Comparative results of ARA algorithm plus MFO feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

4.4.4 Merging and Adapting all Features in One

The idea that all features should combine in one so that they can mutually support to each other sounds very good. However, the experimental results show that such a combination does not offer the results that we were expected. We realized that the component parts are unbalanced. Indeed, it consists of 48 elements selected from Histogram of color (16 for each color channel) and 6 elements from Mean and Standard deviation (2 for each color channel) and 9 elements that coming from HOG feature. Moreover, Histogram of

color and Mean and Standard deviation approaches represent 54 elements while texture features correspond to only 9 elements. To balance it, we built a 45-dimensional vector for each region that includes 24 elements selected from Histogram of color (8 for each color channel $\times$ 3) and 12 elements from Mean and Standard deviation (2 for each color channel $\times$ 3 channels of color $\times$ 2) and 9 elements that coming from HOG feature, namely *MAFO feature*. It is described in Figure 4.17. We tested this mix of features on the BSDS300 dataset with Pattern A. The experimental results denote that the PRI scores increase significantly from 0.795 (MFO feature) to 0.809 (MAFO feature). It helps the ordered rank of this strategy increase two stepped items from the eighth to sixth in the presented results Table 4.5 and Table 4.6.

FIGURE 4.17: Creating 45-dimensional vector for each region



Methods	PRI
Human	0.870
Nguyen, Coustaty, and Guillaume, 2019 *(Lv-mhr)*	0.822
Nguyen, Coustaty, and Guillaume, 2018a *(Lv-ara)*	0.819
Mourchid, El Hassouni, and Cherifi, 2017 *(FMS(HOG))*	0.811
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.810
Our algorithm: ARA + MAFO feature	**0.809**
Mourchid, El Hassouni, and Cherifi, 2017 *(HOG))*	0.803
Nguyen, Coustaty, and Guillaume, 2018b *(Lv-ahr)*	0.800
Comaniciu and Meer, 2002 *(Mean Shift)*	0.780
Li and Wu, 2015 *(L*a*b (HoS))*	0.777
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.770
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.770
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.750
Li and Wu, 2015 *(RGB (HoS))*	0.749

TABLE 4.6: Comparative results of ARA algorithm plus MAFO feature with other methods using the PRI metric on the Berkeley segmentation BSDS300 dataset (validation set).

4.5 Conclusion

All in all, the extraction of precise features plays a very important role in the effective performance of ARA algorithm. Both color property and texture property have their own merits and demerits, therefore, an appropriate combination of these properties could give better results when used in ARA algorithm. During the study period with several experiments and implementations, we determine that the color property helps ARA algorithm to merge homogeneous regions effectively while texture property deals with heterogeneous regions more effectively. Depending on the dataset and the use of the segmentation, we can select the extracted features for ARA algorithm to perform effectively and efficiently. A proper balance of color and texture properties is indispensable: we cannot simply use an equalized distribution for both color property and texture property. For example, if the dataset consists of images in which objects contained homogeneous color inside, the color property might take an advantage. Otherwise, if the dataset includes a large amount of images of cheetahs, tigers and zebras or some patterns like that the texture property might be advantages one.

Chapter 5

Extended Experimental Results for Image Segmentation

5.1 Introduction

This Chapter provides complete experiments that were performed to assess our algorithms and go beyond the simple results presented in the previous Chapter. To evaluate the proposed models, we used three publicly available datasets for image segmentation: Berkeley Segmentation Data Set 300 *(BSDS300)* (Martin et al., 2001), Berkeley Segmentation Data Set 500 *(BSDS500)* (Arbelaez et al., 2011) and Microsoft Research Cambridge Object Recognition Image Database *(MSRC)* (Shotton et al., 2006). Three widely used evaluation segmentation metrics, Variation of Information *(VI)* (Meilă, 2005), Segmentation Covering *(SC)* (Arbelaez et al., 2009) and Probabilistic Rand Index *(PRI)* (Pantofaru and Hebert, 2005), have been applied to measure the accuracy of the proposed algorithms. The quantitative and qualitative evaluations are presented below in Section 5.4.

In the rest of this Chapter we first present the datasets, then the quality assessment functions and finally we present the results of the different algorithms with all the datasets and functions.

5.2 Datasets

The Berkeley Segmentation Data Set 300 (BSDS300) has been built with the aim of providing an empirical basis for research on image segmentation and boundary detection. This dataset comprises 300 images, including 200 images for training and 100 images for validation. Each image has 481 x 321 pixels, which yields a graph of 154401 vertices. The BSDS300 also provides multiple ground-truth segmentation images that are manually generated by many human subjects. For every image, there are from 5 to 10 ground-truth segmentation maps.

The Berkeley Segmentation Data Set 500 (BSDS500) is an extension of BSDS300. This dataset comprises 500 images, including 200 images for training, 200 new testing images and 100 images for validation. Every image has

481 x 321 pixels and 5 ground-truth segmentation maps in average. It supplies a benchmark for comparing different segmentation and boundary detection algorithms.

The Microsoft Research Cambridge Object Recognition Image Database (MSRC) contains a set of 591 natural images of size 320 x 213 with one ground-truth per image grouped into categories. Its intended use is research, in particular object recognition research.

5.3 Evaluation metrics

In the literature, evaluation of image segmentation algorithms generally relies on three classical metrics: Variation of Information *(VI)* (Meilă, 2005), Segmentation Covering *(SC)* (Arbelaez et al., 2009) and Probabilistic Rand Index *(PRI)* (Pantofaru and Hebert, 2005). Especially, PRI is very beneficial for evaluation on BSDS300 and BSDS500 datasets which provide multiple ground-truth.

The Probabilistic Rand Index (PRI) (Pantofaru and Hebert, 2005) is a classical evaluation criterion for clustering. It measures the probability that pairs of pixels have consistent labels in the set of manual segmentations (ground-truth). Given a set of ground-truth segmentations $\{S_k\}$, PRI is defined as:

$$PRI(S_{test}, \{S_k\}) = \frac{1}{T} \sum_{i<j} [c_{ij} p_{ij} + (1 - c_{ij})(1 - p_{ij})] \tag{5.1}$$

where c_{ij} is the event that the algorithm gives the same label to pixels i and j, and p_{ij} corresponds to the probability that pixels i and j have the same label as estimated by the sample mean of the corresponding Bernoulli distribution on the ground-truth dataset. T is the total number of pixel pairs. The PRI values range in [0,1] and a larger value likely indicates a greater similarity.

The Variation of Information (VI) metric was introduced for the evaluation of clustering (Meilă, 2005). It measures the distance between two clusterings in terms of the information difference between them. VI is defined by:

$$VI(C, C') = H(C) + H(C') - 2I(C, C') \tag{5.2}$$

where $H(C)$ and $H(C')$ are the entropy associated with segmentation C and ground-truth segmentation C', respectively and $I(C, C')$ is the mutual information between the two segmentations C and C'. Let segmentation C and ground-truth segmentation C' have N levels of gray and distributions are uniform, *i.e.* $PN = 1/N$. The maximal values of entropies are $H(C) = logN$ and $H(C') = logN$, and let mutual information $I(C, C')$ be equal to zero. Hence, the range of this metric is $[0, 2logN]$, and the smaller value corresponding to the better segmentation results.

The Segmentation Covering (SC) metric measures the average matching between a proposed segmentation and a ground-truth labeling. It was introduced by Arbelaez *et al.* Arbelaez et al., 2009. It is defined by:

$$SC(S, S_g) = \frac{1}{N} \sum_{R \in S} |R|. \max_{R' \in S_g} O(R, R') \tag{5.3}$$

where N denotes the total number of pixels in the image and the *overlap* between two regions R and R' is defined as:

$$O(R, R') = \frac{|R \cap R'|}{|R \cup R'|} \tag{5.4}$$

5.4 Quantitative evaluation

5.4.1 Comparison of patterns and solutions

At first we evaluated our segmentation approaches from a quantitative point of view. We then used the three evaluation metrics presented before and the three datasets stated above individually. Furthermore, we evaluated all our strategies based on the four patterns and display the experimental results in Table 5.1, 5.2, 5.3 and 5.4

In Table 5.1, we show the experimental results of our method using the Reduced Histogram of color, namely *RH feature*, with the four patterns. The pattern that gives highest PRI score in BSDS300 and BSDS500 datasets is pattern A. Pattern B wins all PRI, VI and SC scores in MSRC_V2 dataset and VI, SC scores in BSDS300 and BSDS500 datasets.

Moving to Table 5.2, we again witness the domination of patterns A and B in which the former wins PRI score on BSDS300 and MSRC_V2 datasets and VI, SC scores on BSDS500 dataset. The latter occupies the other highest scores on three datasets. The results in this Table are evaluated based on our method using Mean and Standard deviation of color, namely *MeanSD feature*, for all patterns and datasets.

Table 5.3 presents the evaluation of our algorithm by using Mean and Standard deviation plus Histogram of Oriented Gradients features, named as *HOGMeanSD feature*. One can observe that this combination of features offers better results than the two former strategies. As this feature is built by combining a color feature and a texture feature, the heterogeneity of information gives better results than previously. Especially, in image segmentation problem, we need to cluster some pixels that do not always have homogeneous properties. Moreover, we can see that pattern B almost get the best scores with this heterogeneous feature while pattern A maintains the best PRI score on BSDS300 dataset and shares the first one with pattern B in BSDS500

BSDS300 validation set											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.804	1.56	0.68	0.801	**1.48**	**0.69**	0.793	1.58	0.69	0.786	1.65	0.68
BSDS 500 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.819	1.51	0.71	0.815	**1.46**	**0.72**	0.806	1.57	0.71	0.797	1.67	0.70
MSRC_V2 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.710	1.79	0.70	**0.714**	**1.74**	**0.72**	0.701	1.78	0.71	0.693	1.85	0.69

TABLE 5.1: Comparative results of Lv-ahr algorithm using the three different metrics with the Quantized Histogram of Color features (RH) for regions merging (Nguyen, Coustaty, and Guillaume, 2018b).

BSDS 300 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.825	1.36	**0.74**	0.821	**1.29**	0.72	0.817	1.48	0.70	0.805	1.54	0.69
BSDS 500 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.832	**1.37**	**0.77**	**0.840**	1.46	0.72	0.820	1.53	0.72	0.809	1.62	0.71
MSRC_V2 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.733	1.66	0.74	0.727	**1.58**	**0.75**	0.71	1.70	0.71	0.70	1.75	0.70

TABLE 5.2: Comparative results of Lv-ara algorithm using the three different metrics with the MeanSD features for regions merging (Nguyen, Coustaty, and Guillaume, 2018a).

dataset.

In a global analysis, we can see that pattern B is the best pattern for all metric used in this study. The next best pattern is pattern A in three Table 5.1, 5.2 and 5.3 while pattern C is slightly lower the pattern A in terms of metric scores. Pattern D gives the worst results, even though the difference between the four patterns is very small. While pattern B gives the best results, it also

BSDS 300 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.825	1.35	0.74	**0.828**	**1.30**	**0.75**	0.822	1.43	0.74	0.812	1.51	0.74
BSDS 500 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.834	1.29	0.75	**0.834**	**1.26**	**0.75**	0.826	1.32	0.74	0.818	1.40	0.73
MSRC_V2 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.739	1.63	0.75	**0.74**	**1.53**	**0.77**	0.73	1.69	0.76	0.72	1.75	0.74

TABLE 5.3: Comparative results of Lv-mhr algorithm using the three different metrics with the HOGMeanSD features for regions merging (Nguyen, Coustaty, and Guillaume, 2019).

generates the most complex networks with a high computational cost as discussed in Section 3.4.

We can also observe that the three features proposed and the combination of those strategies offer very different results. The HOGMeanSD feature and MeanSD feature gave very similar results and it is difficult to distinguish which one is the best, while RH feature got lower results.

Every feature has its own merits and demerits in some particular cases of images segmentation. Our hypothesis is that those strengths and weaknesses complement one another and by combining them we could get better results, hence the *MFO feature* that combines all features in one. Result for this feature are given in Table 5.4). Pattern C which, as we mentioned above, generates networks easy to handle for Louvain, and using this pattern we obtain the highest PRI score *(PRI=0.840)* for the dataset BSDS500. This MFO feature provides the pretty stable VI and SC scores for the three patterns A, B and C. Finally, the two other combinations of features are the BHMF feature and the MAFO feature. Results are presented in Table 5.5 and Table 5.6, respectively. BHMF feature with ARA algorithm offers the best image segmentation results for all the evaluation metrics for all tested dataset using pattern A.

5.4.2 Comparison with the state of the art

We now present some comparison with the state of the art to assess the relevance of our method. Table 5.7 and Table 5.8 presents a comparison of our methods using pattern A.

BSDS 300 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.795	1.48	0.73	**0.796**	1.45	0.74	0.792	1.49	0.73	0.785	1.56	0.73
BSDS 500 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.813	1.47	0.74	0.813	1.40	0.74	**0.840**	1.45	0.74	0.801	1.54	0.75
MSRC_V2 dataset results											
Pattern (a)			Pattern (b)			Pattern (c)			Pattern (d)		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.734	1.73	0.71	0.728	1.69	0.72	0.714	1.70	0.72	0.703	1.76	0.70

TABLE 5.4: Comparative results of Lv-ara algorithm using the three different metrics with the MFO features for regions merging

Our BHMF-feature method								
BSDS300			BSDS500			MSRC_V2		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.828	1.33	0.77	0.838	1.28	0.77	0.743	1.61	0.754

TABLE 5.5: The results of Lv-ara algorithm using the three different metrics with the BHMF features for regions merging

Our MAFO-feature method								
BSDS300			BSDS500			MSRC_V2		
PRI	**VI**	**SC**	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
0.809	1.48	0.735	0.824	1.43	0.74	0.733	1.79	0.696

TABLE 5.6: The results of Lv-ara algorithm using the three different metrics with the MAFO features for regions merging

5.5 Qualitative evaluation

Lastly, we propose a qualitative evaluation of our results. We present some results of our segmentation results on the three datasets represented below by three Figures: Figure 5.1 with images taken at random from BSDS300, Figure 5.2 for BSDS500 and Figure 5.3 for MSRC_V2 dataset. From these qualitative results, we can see that the proposed algorithm offers good results and produces sizable regions for all selected images. Although each feature produces poor and rich quality of image segmentation differently, our algorithm

Methods	PRI	VI	SC
Human	0.870	1.160	-
Our method: BHMF feature	**0.828**	**1.330**	**0.770**
Our method: HOGMeanSD feature	0.825	1.350	0.740
Our method: MeanSD feature	0.825	1.360	0.740
Our method: MAFO feature	0.809	1.480	0.735
Our method: RH feature	0.804	1.560	0.680
Our method: MFO feature	0.795	1.480	0.730
Mourchid, El Hassouni, and Cherifi, 2017) *(FMS(HOG))*	0.811	-	-
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.810	1.470	0.750
Mourchid, El Hassouni, and Cherifi, 2017) *(HOG))*	0.803	-	-
Comaniciu and Meer, 2002 *(Mean Shift)*	0.780	1.630	0.660
Li and Wu, 2015 *(L*a*b (HoS))*	0.777	1.879	-
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.770	1.790	0.680
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.770	1.810	0.660
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.750	1.840	0.660
Li and Wu, 2015 *(RGB (HoS))*	0.749	2.149	-

TABLE 5.7: Quantitative comparisons of our methods using **pattern A** to others methods tested on BSDS300.

Methods	PRI	VI	SC
Human	0.870	1.170	-
Khan and Sundaramoorthi, 2018 *(Learned STLD)*	**0.860**	1.540	0.670
Our method: BHMF feature	0.838	**1.280**	**0.770**
Our method: HOGMeanSD feature	0.834	1.290	0.750
Our method: MeanSD feature	0.832	1.370	0.770
Our method: MAFO feature	0.824	1.430	0.740
Our method: RH feature	0.819	1.510	0.710
Our method: MFO feature	0.813	1.470	0.740
Arbelaez et al., 2011 *(gPb-owt-ucm)*	0.830	1.480	0.740
Zohrizadeh, Kheirandishfard, and Kamangar, 2018 *(IS4(MCG))*	0.830	1.350	0.630
Felzenszwalb and Huttenlocher, 2004 *(Felz-Hutt)*	0.800	1.870	0.690
Comaniciu and Meer, 2002 *(Mean Shift)*	0.790	1.640	0.660
Arbelaez et al., 2011 *(Canny-owt-ucm)*	0.790	1.890	0.660
Cour, Benezit, and Shi, 2005 *(NCuts)*	0.780	1.890	0.670

TABLE 5.8: Quantitative comparisons of our methods using **Pattern A** to others methods tested on BSDS500.

can aggregate homogeneous neighboring regions successfully even if pixels

inside each region are dissimilar. In these cases, our method achieve object-level segmentation to some extent.

These Figures also highlight the fact that the combination between HOG (texture property) and Mean and Standard deviation (color property) to create feature (for instance, HOGMeanSD and BHMF features) proposes a better agglomeration of homogeneous regions than the others in case of considering the included perspective and tolerates the dissimilarities in whole the image in order to produce better image segmentation results.

The MeanSD feature preserves the homogeneity and considers more accurate local feature so it can produce object-level in case of the presence of local homogeneity feature in considering objects. Since this feature replies to the consideration of similarity between regions (sub-segments) in object. Moreover, Mean and Standard deviation features allows extracting homogeneous regions in objects which do not have dissimilarities inside based on the given criteria. Finally, RH feature can offer good image segmentation results for images whose the color presented high contrast between objects and background. For example, when we reduce the histogram of color from 256 to 16 items means that it merged the group with the 16 units step of color density, and from 256 to 8 items means making groups with 32 units step of color density. This reason can be an answer for the question why RH feature produces lower scores. Because it might merge all the sub-segments matching the criterion for group histogram while the dissimilarities between the objects and background in many images do not always satisfy with this matching criterion. Otherwise, in case of repetitive patterns of different colors in many homogeneous objects or the quite different color parts inside an entity (for instance, Cheetah, Tiger and Zebra) this feature might not merge successfully sub-segments to produce object. It is determined that the color feature hardly achieve instance image segmentation in case of the repetitive patterns of different colors in many homogeneous objects or the quite different color parts inside an entity as we mentioned above.

We have also made a qualitative comparison of our method to HoS and HOG methods (Li and Wu, 2015) in two Figures 5.4 and 5.5. While Shijie Li's method produces images segmentation toward object-level with maximum objects in general, our method achieves, to some extent, object-level in more specific objects. Shijie Li's method reconstructs the neighborhood system for each region based on color property and histogram of states (HoS). Then, they estimate the distribution of the similarity feature for each region by the way that it is computed and adapted to update the similarity matrix W based on the color property and histogram of states whenever implementing the merging regions. These processes of the recomputed and adapted similarity matrix W lead to the reduction of dissimilarity distances of merged regions and considering regions. For example, region R_1 more similar to R_2, and R_2 is similar to R_3 however R_1 are not similar to R_3. As a consequence,

when the merging occurs between R_1 and R_2 the similarity matrix W is recomputed and adapted lead to the merged region $R_{1,2}$ is similar to the region R_3. It means that the next repeated step the $R_{1,2}$ can merge with the region R_3 to create big region $R_{1,2,3}$. In our algorithm, we compute all the similarity feature vectors representative for all regions at once and we implement the merging processes without recomputing the similarity feature vectors.

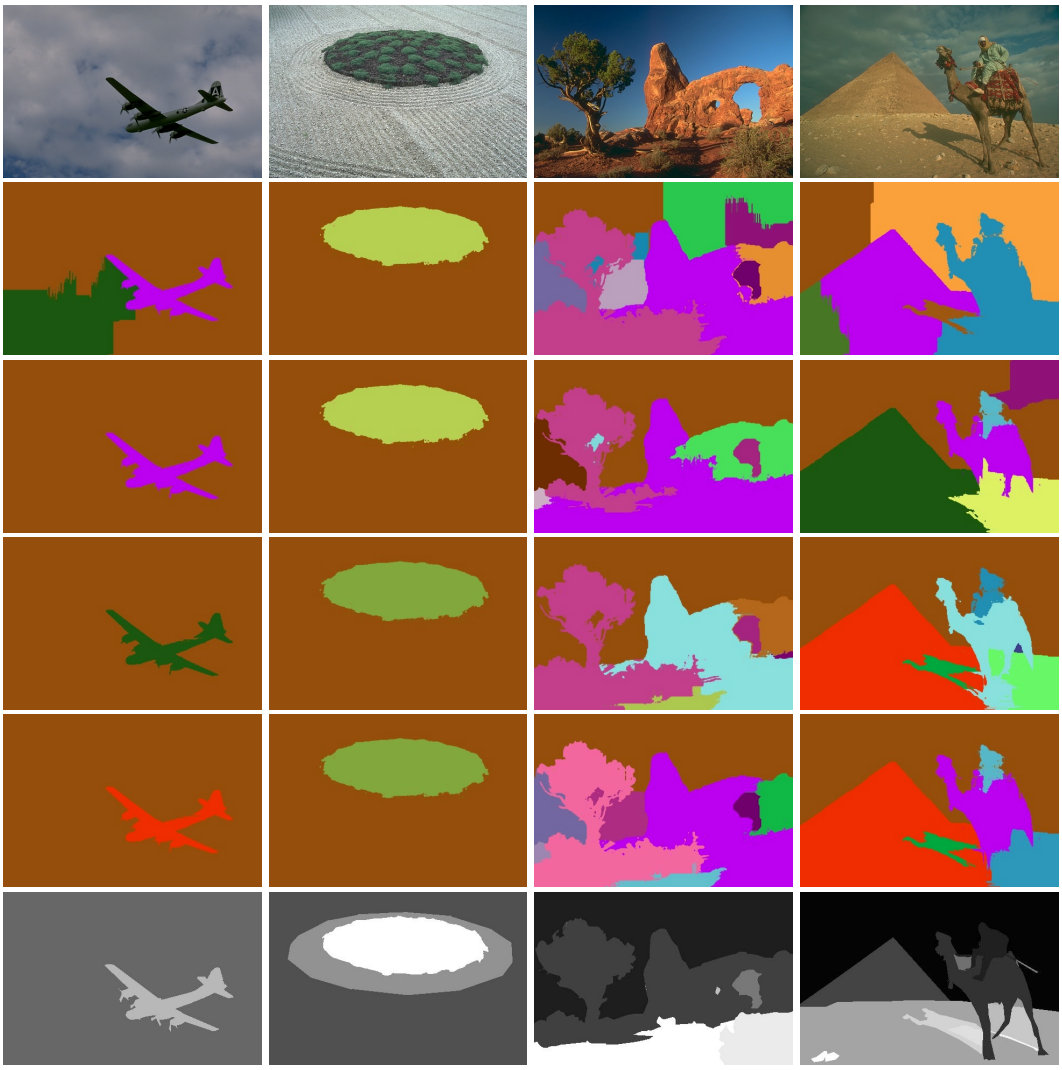


FIGURE 5.1: Qualitative evaluation of results using Best pattern (Pattern A) on BSDS 300 dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth.

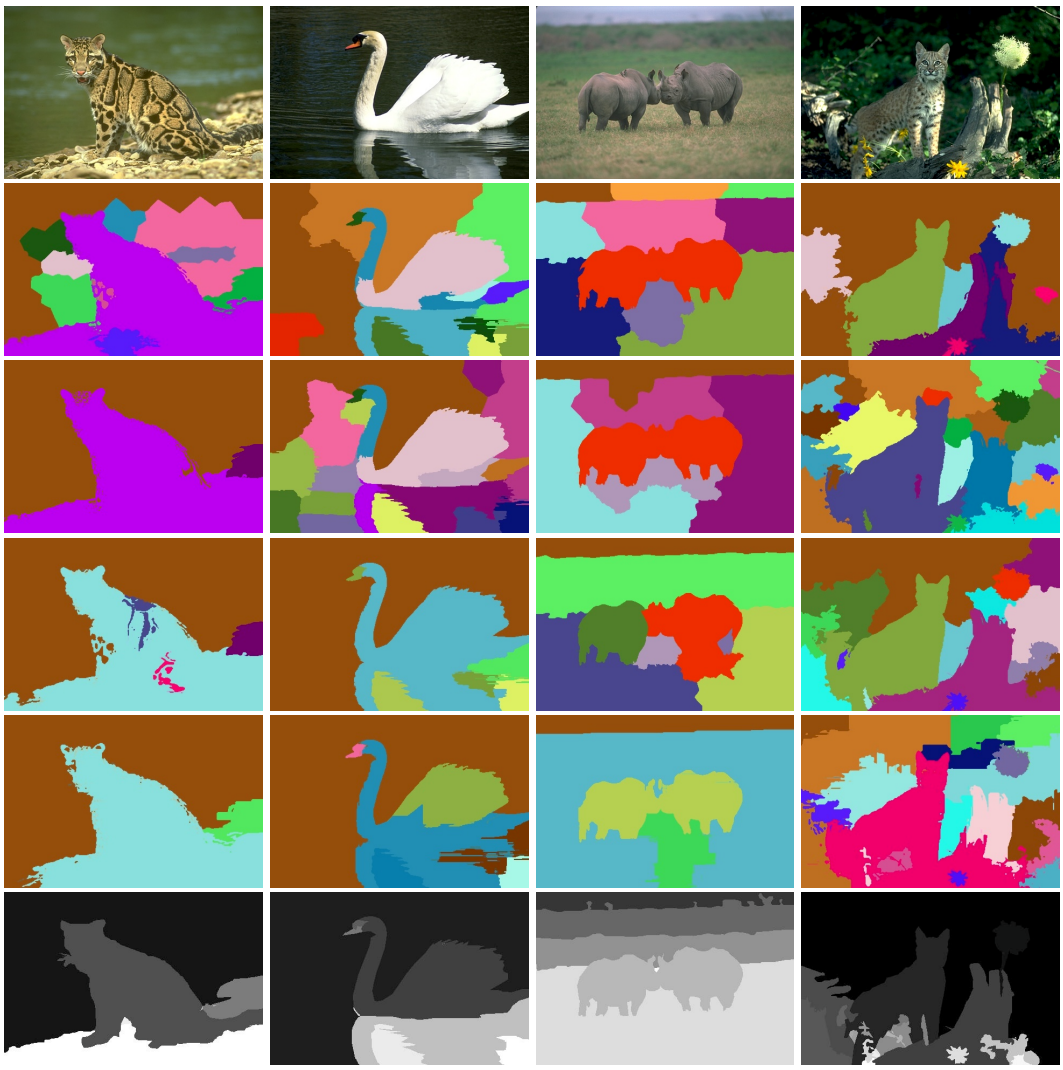


FIGURE 5.2: Qualitative evaluation of results using Best pattern (Pattern A) on BSDS 500 dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth.

5.6 Conclusion

In this Chapter, we made the intensive and extensive experiments with all our proposed methods on the four mentioned patterns with three datasets and three evaluation metrics. The details of these experimental results have been reported and took into account the merits and pointed out the drawbacks of the particularly methods. In addition, we also made comparisons to other methods using both quantitative and qualitative evaluation methods, and discussed on the experimental results. With these deep experiments, we could understand our methods accurately and clearly in more detail.

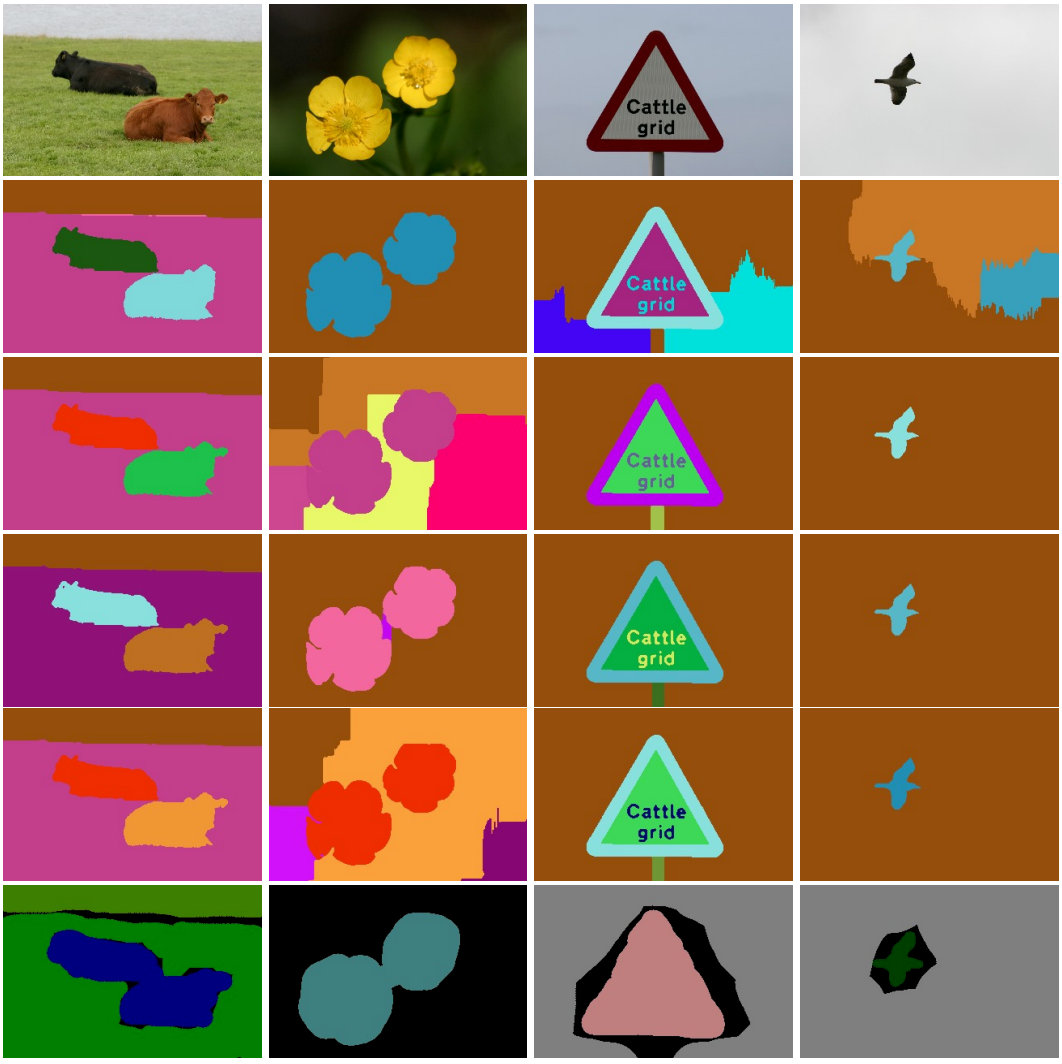


FIGURE 5.3: Qualitative evaluation of results using Best pattern (Pattern A) on MSRC dataset. *First line*: Original images, *Second line*: Results of using RH feature, *Third line*: Results of using MeanSD feature, *Fourth line*: Results of using HOGMeanSD feature, *Fifth line*: Results of using BHMF feature, *Sixth line*: Ground-truth.

FIGURE 5.4: Visual Segmentation results of some methods: *First column:* original image. *Second column:* Segmentation result of HOG method (Li and Wu, 2015). *Third column:* Segmentation result of HoS method. (Li and Wu, 2015) *Fourth column:* Our segmentation results

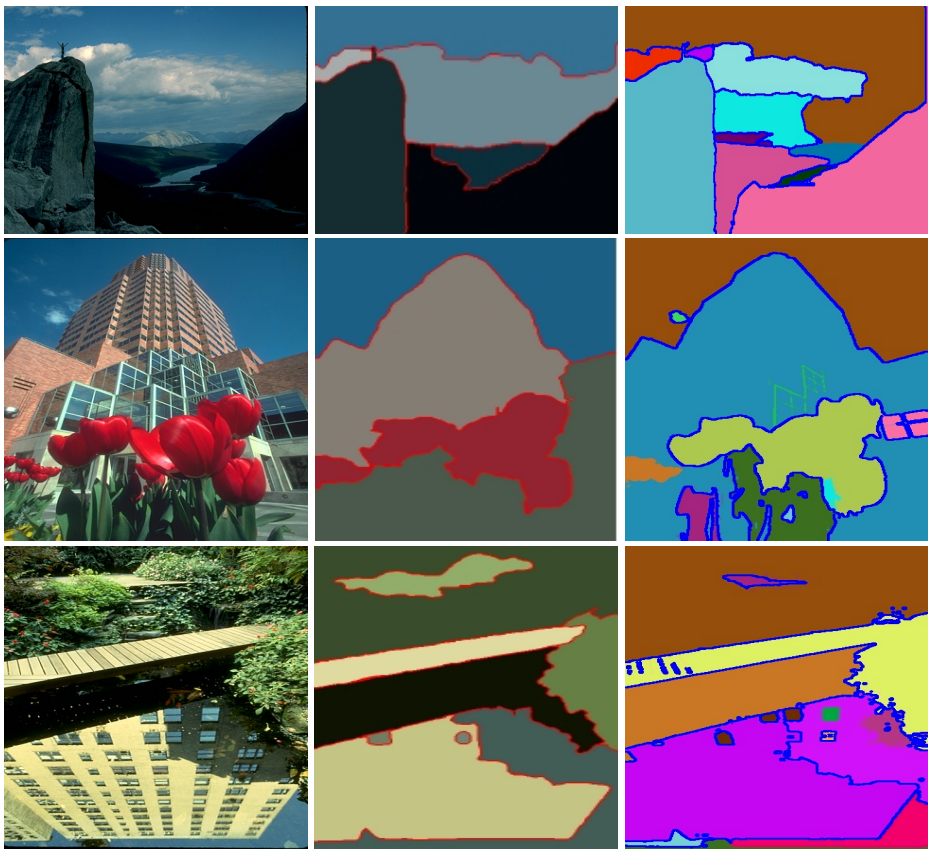


FIGURE 5.5: Qualitative comparison of segmentation results by HoS methods: *First column:* original image. *Second column:* Segmentation result of HoS method (Li and Wu, 2015). *Third column:* Our segmentation results

Chapter 6

A Consensus-Based Image Segmentation Algorithm

In this Chapter we propose a new solution to the image segmentation problem based on the general principle of consensual clustering (Strehl and Ghosh, 2002). This approach is similar to the ones previously developed but that relies on less cleaning and merging steps. The basic idea is to compute several segmentation of a given image, or more precisely several partitions of the associated graph, and to find similarities between all these partitions. The result of the consensual clustering is a hierarchy of communities that can be assimilated to a hierarchy of segmentation.

Our general framework is therefore as follows: first we convert one image into a graph, then we compute several partitions of this graph using any community detection algorithm. In practice we use the Louvain method as before (Blondel et al., 2008). Similarities between these partitions are computed and form a hierarchy. Finally, choosing a level in the hierarchy wisely gives the segmentation of the original image.

The Chapter is organized as follows. In Section 6.1 we present the general framework including the construction of a graph, the community detection and the consensual community computation. We then present and discuss the results in Section 6.2.

6.1 Proposed Approach

Classical graph-based approaches described before (Youssef Mourchid, Mohammed El Hassouni and Hocine Cherifi, 2015; Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi, 2016; Nguyen, Coustaty, and Guillaume, 2018a), generally use a supervised and complex fusion strategy to avoid over-segmentation. In contrast, the main contribution of this Chapter relies on a consensus clustering algorithm (Strehl and Ghosh, 2002) to automatically segment images.

To reach this goal, we first consider images as complex networks and apply any non-deterministic community detection algorithm several times. The

different clusterings obtained are then reconciled using a consensus clustering algorithm that computes shared information between the clusterings. Finally, the consensus clustering algorithm returns a hierarchy of clusterings that can be viewed as a hierarchy of image segmentations.

As before, we use the Louvain method (Blondel et al., 2008) to detect communities and we use the consensus clustering method described in (Seifi et al., 2013) that follows the general framework from (Strehl and Ghosh, 2002). The Louvain method is particularly useful in this context since it is non-deterministic and therefore it returns a different partition for every execution. It is also very fast and it provides high quality partitions (Fortunato, 2010).

6.1.1 Complex Networks from an image

As before, we only consider unweighted graphs, i.e. all edges have the same weight, where an edge e_{ij} exists if i and j are close and d^c_{ij}, the similarity of pixels i and j is lower than a threshold. d^c_{ij} is defined by $d^c_{ij} = \left| I^c_i - I^c_j \right|$ where I^c_i and I^c_j represent the intensity of pixel i and j respectively for color channel c (among R, G and B).

To account for proximity, we define the admissible neighborhood of pixels using patterns: only pixels within a given pattern can be connected. We are then able to build different complex networks from a given image by varying the proximity pattern. We only use the two first pattern defined before, patterns (a) and (b) that have been shown to give good results (pattern (b)) and a good compromise and speed and quality (pattern (a)).

Recall that for pattern (a), a given pixel can only have edges towards other pixels if these pixels are inside N neighbouring pixels on the same row or column. If this proximity property is met, the edge is really created if the similarity d^c_{ij} is lower than a threshold t. Pattern (b) allows to create edges between a given pixel and all the other pixels inside a $N \times N$ square neighbourhood and again the similarity criterion. The complex networks which are created by that pattern contain more edges and therefore more information for community detection algorithms to perform well. However, the N value must be kept low since the global density increases quadratically which slows both the construction of the networks and the execution of the algorithms.

6.1.2 Louvain method and consensual clustering

The Louvain method (Blondel et al., 2008) has already been presented before. Although it has been widely applied for image segmentation purposes,

it cannot overcome the over-segmentation problem directly. The main principle behind the modularity-based image segmentation algorithms is to produce many homogeneous regions that could belong to one object and homogeneous regions are then merged. Furthermore we can also observe that several executions can gives slightly different results since the Louvain method is a non-deterministic heuristic.

Here we propose an alternative approach that uses consensual clustering (Strehl and Ghosh, 2002) and more precisely the solution defined for graphs in (Seifi et al., 2013). The idea is to execute the Louvain method several times and to find similarities between all the obtained clusterings.

Consensual clustering takes as input a set of k partitions of the same set (that are supposed to be somehow similar) and a parameter $p \in [0, 1]$ and will extract similarities that are present in at least a fraction p of the k partitions. For instance if nodes i and j are in the same community in 50 percent of the experiments and if $p \leq 0.5$ then i and j will be in the same .5-consensual community. Note that the reverse is not always true and i and j can still be in the same .5-consensual community even though they are in the same community in half of the experiments. For instance if i and k are together in half of the experiments, and k and j also then by transitivity i and j are together.

Another way to understand this notion is to compute the matrix M where M_{ij} is the fraction of the experiments where i and j are grouped. Then, given a threshold p, all values lower than p are set to 0. The p-consensual communities are the connected components of the graph whose adjacency matrix is the thresholded version of M.

Finally it is interesting to note that consensual communities form a hierarchy. Indeed an increase of p can only result in the split of some communities: p_2-consensual communities form a sub-partition of p_1-consensual communities is $p_2 > p_1$.

The global algorithm therefore follows the following steps:

- given an image, a proximity pattern and a distance N, construct the unweighted graph that corresponds to the image;
- execute Louvain method on the network a sufficient number of times k;
- compute the complete consensual-community hierarchy using all values of $p = i/k$ for $i \in \{0, k\}$;
- evaluate the quality of the image segmentation that corresponds to the p-consensual communities.

Previous results from this thesis have already shown that a typical value of $N = 20$ is a good compromise between quality and speed. Similarly it has been shown (Seifi et al., 2013) that the consensual community computation

converges really fast and that k can be kept low. We selected a value of 100 that gives a hierarchy with that many levels and that number of segmentations. Finally the only parameter left is p if we need a single segmentation and not a hierarchy.

6.2 Experimental Evaluation

This Section provides experiments that were performed to assess our approach. To evaluate the proposed model, we used two widely used and publicly available datasets for image segmentation and three metrics to evaluate the quality of the segmentation. As before, we use the same datasets: BSDS500 (Arbelaez et al., 2011) and MSRC (Shotton et al., 2006) and we evaluate the approach with the three classical metrics: Variation of Information (Meilă, 2005), Segmentation Covering (Arbelaez et al., 2009) and Probabilistic Rand Index (Pantofaru and Hebert, 2005).

6.2.1 Results

At first we evaluated our segmentation approach from a quantitative point of view. Given that the method yields a hierarchy of segmentations, we need to choose a level in this hierarchy, *i.e.* a threshold for the consensual step. Intuitively, if the threshold is very low (close to 0) then it is very likely that any two pixels will be in the same consensual community (remember that consensual communities are obtained by transitive closure) and we will have a single region. On the contrary if the threshold is very high (close or equal to 1) then two pixels will be in the same region if and only if all clusterings agree that these two pixels are to be grouped. This will create many regions. In between, we will observe fewer and hopefully more meaningful regions.

This is validated by Figure 6.1 that indeed shows an increase of the PRI when the segmentation gets closer to the ground-truth and a final decrease when the image gets over-segmented. In this specific case the maximal value of PRI is obtained for a threshold around 0.844. Figure 6.2 present the images that correspond to this curve.

In the following all experiments have been performed running Louvain 100 times. This number of executions ensure a quick computation but already provides a deep hierarchy and several levels of segmentation.

Table 6.1 present the comparative results of our approach. The presented scores always correspond to the optimal choice of threshold during the consensus step, i.e., the one that gives the best score. We will discuss this later on. Using our method we present two different scores: OIS corresponds to the best choice of threshold for each image individually, i.e., one image might be segmented with a threshold of 0.86 while another one could use a threshold of 0.93. In some sense OIS corresponds to the best theoretical value one

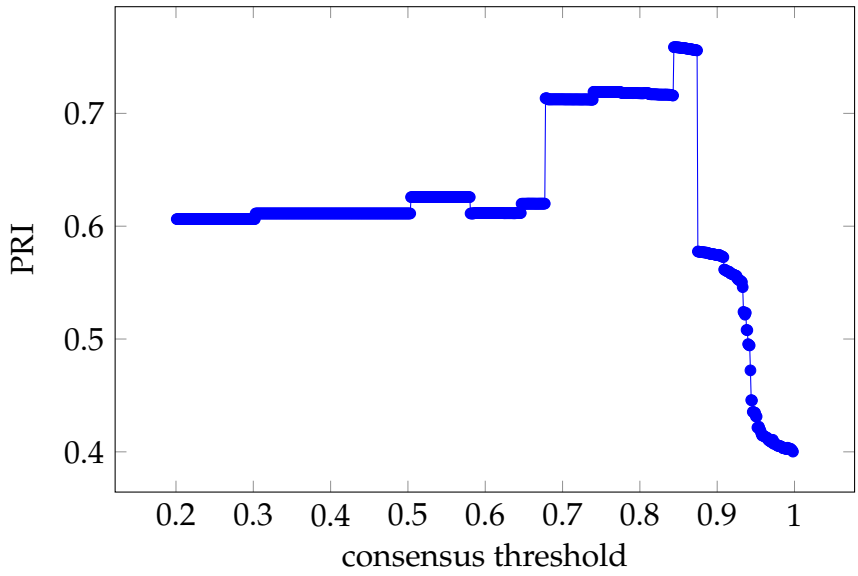


FIGURE 6.1: Evolution of PRI versus consensus threshold

can get with our approach. On the contrary, ODS uses a similar threshold for all images in the dataset which explains the lowest quality. See (Arbelaez et al., 2011) for more details on the evaluation method.

While we presented two different patterns, the experimental results show that the pattern (b) provides better results when compared to the pattern (a). The reason for this is that pattern (b) creates graphs with more information since the neighborhood of each node is larger. Communities are therefore more meaningful and the segmentation as well. One should however remember that the computational cost is higher with pattern (b).

More generally, the results show that the OIS approach gives excellent results but determining the best threshold for each image is beyond our capabilities. These scores should therefore be seen as bounds on the quality that can be reached with our approach. ODS also gives very good results.

6.3 Conclusion

This Chapter proposed a new image segmentation approach that relies on an analogy between image segmentation and community detection on graphs and use consensual clustering for community detection. We performed extensive experiments using different patterns, different features and different datasets. Our method, while very simple, produces efficient image segmentation results compared to the state of the art.

This method provides room for many improvements. In particular, the graph is currently built in a very simple way and creating weighted graphs

Dataset	**BSDS500**			**MSRCv2**		
Metrics	**PRI**	**VI**	**SC**	**PRI**	**VI**	**SC**
Methods	ODS / OIS	ODS / OIS	ODS / OIS	ODS / OIS	ODS / OIS	ODS / OIS
Human	0.87/0.87	1.16/1.16	-/-	-/-	-/-	-/-
Square method (Pattern (b))	0.806/**0.826**	**1.262/1.034**	**0.822/0.861**	0.682/**0.749**	1.585/**1.279**	**0.774/0.823**
L-shape method (Pattern (a))	0.768/0.810	1.357/1.064	0.806/0.859	0.674/0.745	1.653/1.294	0.754/0.810
Khan and Sundaramoorthi, 2018	**0.86**/0.86	1.54/1.54	0.67/0.67	-/-	-/-	-/-
Nguyen, Coustaty, and Guillaume, 2019	0.834/-	1.26/-	0.75/-	**0.74**/-	**1.53**/-	0.77/-
Mourchid, El Hassouni, and Cherifi, 2017	0.811/-	-/-	-/-	-/-	-/-	-/-
Nguyen, Coustaty, and Guillaume, 2018b	0.821/-	1.46/-	0.72/-	0.727/-	1.58/-	0.75/-

TABLE 6.1: Quantitative comparisons on BSDS500 and MSRCv2 datasets

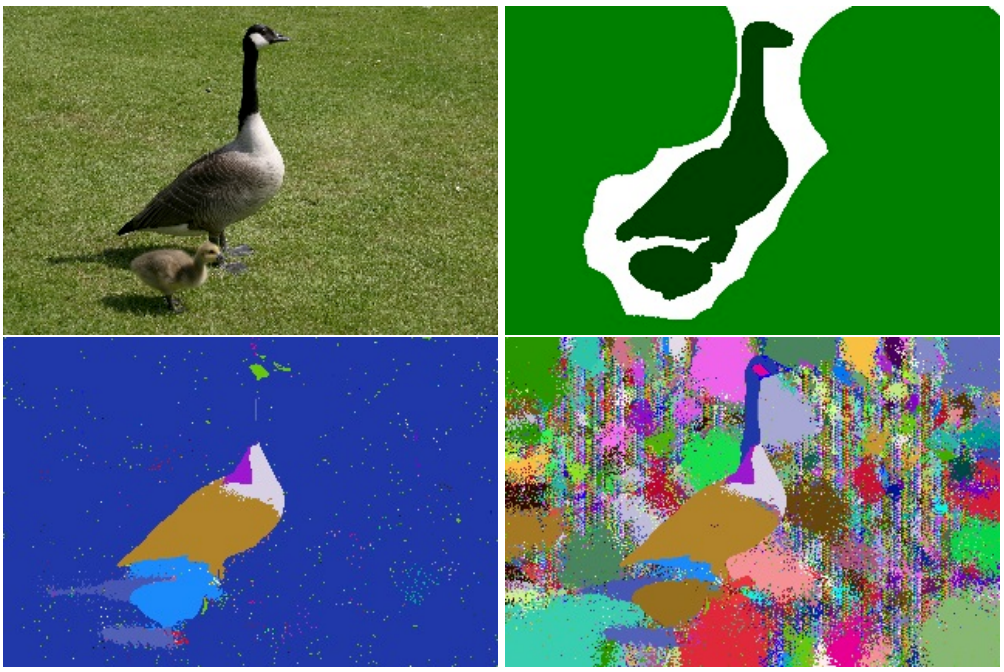


FIGURE 6.2: Qualitative evaluation on image 12_19_s of MSRC dataset. From top to bottom and left to right: original image, ground-truth, best segmentation (threshold=0.844), most segmented image (threshold=0.999)

(with a weight that represent the similarity more precisely) would much likely increase the quality of the community detection step. Another probable improvement would be to clean the output. Indeed we can observe some noise one the results with very small communities of few pixels. Merging these pixels with the enclosing community would improve a little the results.

Finally the selection of the threshold needs some further insights even though the hierarchy itself is very informative.

Chapter 7

Using Extracted Features for Content Based Image Retrieval

Inspired by the efficiency of extracted features for image segmentation, we attempt to set up an image search engine using Bag-of-Visual-Words (BoVW) in which our extracted features are used. In this Chapter, we propose a detailed description of our proposed image retrieval system, which consists of a technique, a mechanism, and experimental results.

7.1 Introduction

Image retrieval (IR) is an active research topic in the computer vision field. Especially, with the dramatic increase in multimedia data, the traditional information retrieval techniques do not meet the demand for users. Content-based image retrieval (CBIR) has become one of the hot topics, and has achieved significant development.

Image retrieval is the discipline of the research that concerned in looking, browsing and recovering digital images from an extensive database. Traditional image retrieval systems utilize embedded metadata such as captioning, keywords, or descriptions of the images in order to perform image retrieval through the annotation words. Recently, the amount of digital multimedia data has increased rapidly. Therefore, information retrieval techniques have shifted from text-based to become semantic-based or content-based.

The text-based image retrieval (TBIR) is a traditional image retrieval method that utilizes the metadata, and are based on artificial notes to retrieve image results. This image retrieval method not only requires huge labors to manually annotate the images, but also takes time consuming and complicated work.

The content-based image retrieval (CBIR) means that the search analyzes the content of the image based on features related to color, texture, shape or any other informative that can be derived from image properties themselves. This image retrieval method is desirable since most image search engines

repy purely on metadata, and this process produces a lot of unrelated images.

The semantic-based image retrieval (SBIR) is an automatic method of extracting semantically meaningful representations from low-level features of images to support query by high-level semantic concept in user's mind. Reducing semantic gap between the low-level features and high-level semantic concept is a main concern of semantic image retrieval efforts.

In fact, there is a difference between what image features can distinguish and what people perceives from the image. Semantic based image retrieval can be made by extraction of low-level features of images to identify meaningful and interesting objects based on the similar characteristics of the visual features. In a sense, the object features will go into semantic image extraction process to get the semantic description of images to be stored in database. Image retrieval can be queried based on high-level concepts. The query may be done based on a set of textual words that will go into semantic features translator to get the semantic features from the query. The semantic mapping process is used to find the best concept to describe the segmented or clustered objects based on the low-level features. This mapping will be done through supervised or unsupervised learning tools to associate the low-level features with object concept and will be annotated with the textual word through image annotation process (Wang, Mohamad, and Ismail, 2010; Wang, Mohamad, and Ismail, 2009). Semantic content can be obtained by textual annotation or by high sophisticated inference procedures based on visual content. Recent successes in deep learning enabled detection of objects belonging to many classes greatly outperforming traditional computer vision techniques (Wan et al., 2014; Sadeghi-Tehran et al., 2019; Potapov et al., 2018).

7.2 Bag of Visual Works Model

The Bag-of-Visual-Works (BoVW) model is one of the most popular CBIR models that represent images as a collection of local features. For this reason, some researchers tend to name it as a bag of features. These local features are typically groups of local descriptors. The total number of local descriptors that is extracted for each image may be colossal. In addition, searching nearest neighbors for each local descriptor in the image query consumes a long time. Therefore, BoVW was proposed as an approach to tackle this issue by quantizing descriptors into "visual words" which decreases the descriptors' sum drastically. Thus, BoVW makes the descriptor more robust to change. This model is very close to the traditional description of texts in information retrieval, but it is considered for images retrieval (Csurka et al., 2004; Lowe, 2004). Normally, Bag-of-Visual-Words usually reply on the Keypoint detectors and Keypoint descriptors. However, there are many researches using methodology with graphs (Silva et al., 2013; Silva et al., 2018; Cortés Llosa, Conte, and Cardot, 2018), Strokes (Mandal et al., 2018; Okawa, 2016), Bag-of-ARSRG (attributed relational SIFT (scale-invariant feature transform) regions

graph) words (BoAW) (Manzo and Pellino, 2019), and *etc.*.

Bag-of-Visual-Words is one of the de facto standard of image features for retrieval and recognition (Bay, Tuytelaars, and Van Gool, 2006) with deep-learning approaches. In our work, we focused on the BoVW models as we want to test the relevance of our image segmentation approach and the associated features in the context of CBIR. This BoVW method comprises three main stages: (1) finding the region of interests (interesting part) of an image using Keypoint detectors in general; (2) computing a summary of the Region through the use of a feature vector (like the Keypoint descriptors); and (3) building the vocabulary to define a common subspace (the vocabulary) for all images. This subspace will then make images comparable. We present detailed description in the following sequential subsections.

7.2.1 Keypoint Detection

Keypoint detection is a process that uses a feature detector to determine regions of an image that have unique content, such as corners. Feature detection is used to look for points of interest (keypoints) in the image that remain locally invariant. Therefore, it can detect them even in the presence of scaling or rotation.

Detecting Keypoints is the first step of the BoVW model. In order to extract feature from interest points, these features are computed at predefined locations and scales (Liu, 2013). Feature extraction is an individually separated process from feature representation in BoVW approaches (Mikolajczyk et al., 2005). In the literature, it is witnessed the existence of several keypoint detectors that were used in research, such as Harris-Laplace, Difference of Gaussian (DoG), Hessian Laplace, and Maximally Stable Extremal Regions (MSER) (Mukherjee, Mukhopadhyay, and Mitra, 2014; Tsai, 2012; Dang et al., 2019).

7.2.2 Keypoint Descriptors

Feature descriptor (Keypoint descriptor) involves computing a local descriptor, which is usually done on regions centered on detected interest points. Local descriptors depend on image processing to transform a local pixel neighborhood into a compact vector representation.

According to the content of Keypoints, they are described as multidimensional numerical vectors. In other words, features descriptors are utilized to determine how to represent the neighborhood of pixels near a localized keypoint (Mikolajczyk et al., 2005). The most popular feature descriptors from the litterature are the scale-invariant feature transform (SIFT) (Lowe, 2004), speeded up robust features (SURF) (Bay, Tuytelaars, and Van Gool, 2006), gradient location and orientation histogram (GLOH) (Mikolajczyk and

Schmid, 2005) and histogram of oriented gradients (HOG) (Dalal and Triggs, 2005).

7.2.3 Building Vocabulary

In the previous stage, the amount of extracted feature descriptors are quite large. To solve this problem, we need to apply clustering algorithm to cluster the feature descriptors in order to generate a visual vocabulary, for instance, K-Means technique (Hartigan and Wong, 1979). Each cluster is treated as a distinct visual word in the vocabulary, which is represented by its respective cluster center. The size of the vocabulary is determined when using the clustering algorithm, and depends on the size and the types of the dataset (Bay, Tuytelaars, and Van Gool, 2006).

The Bag-of-Visual-Words model can be formulated as follows. First of all, BoVW usually defines the training dataset as S including images represented by $S = s_1, s_2, \ldots, s_n$, where s is the extracted visual features. After that, utilizing the clustering algorithm like K-Means, which is based on a fixed number to visual words W represented by $W = w_1, w_2, \ldots, w_v$, where v is the cluster number. Then, the data is summarized in a $V \times N$ occurrence table of counts $N_{ij} = n(w_i, s_j)$, where $n(w_i, s_j)$ denotes how often the word w_i is occurred in an image s_j (Lowe, 2004).

7.3 Our CBIR Architecture

Based on the results we got in the previous Sections on the image segmentation task, we propose a system for image retrieval based on extracting local features using BoVW model from the segmented regions. The system uses the features introduced for the image segmentation methods and computes the descriptor for those segmented regions. For the validation of our system we decided to use the well-known K-Means algorithm to obtain the visual vocabulary. As shown in Figure 7.2, the proposed system consists of two stages: a training stage and a testing stage.

7.3.1 Training stage

In this stage, the proposed system will implement repeatedly for all images in the dataset.
1) For each image in the dataset:

- Read each image individually
- Convert image to a complex network (Graph)
- Segment the image based on community detection: Louvain method and ARA algorithm

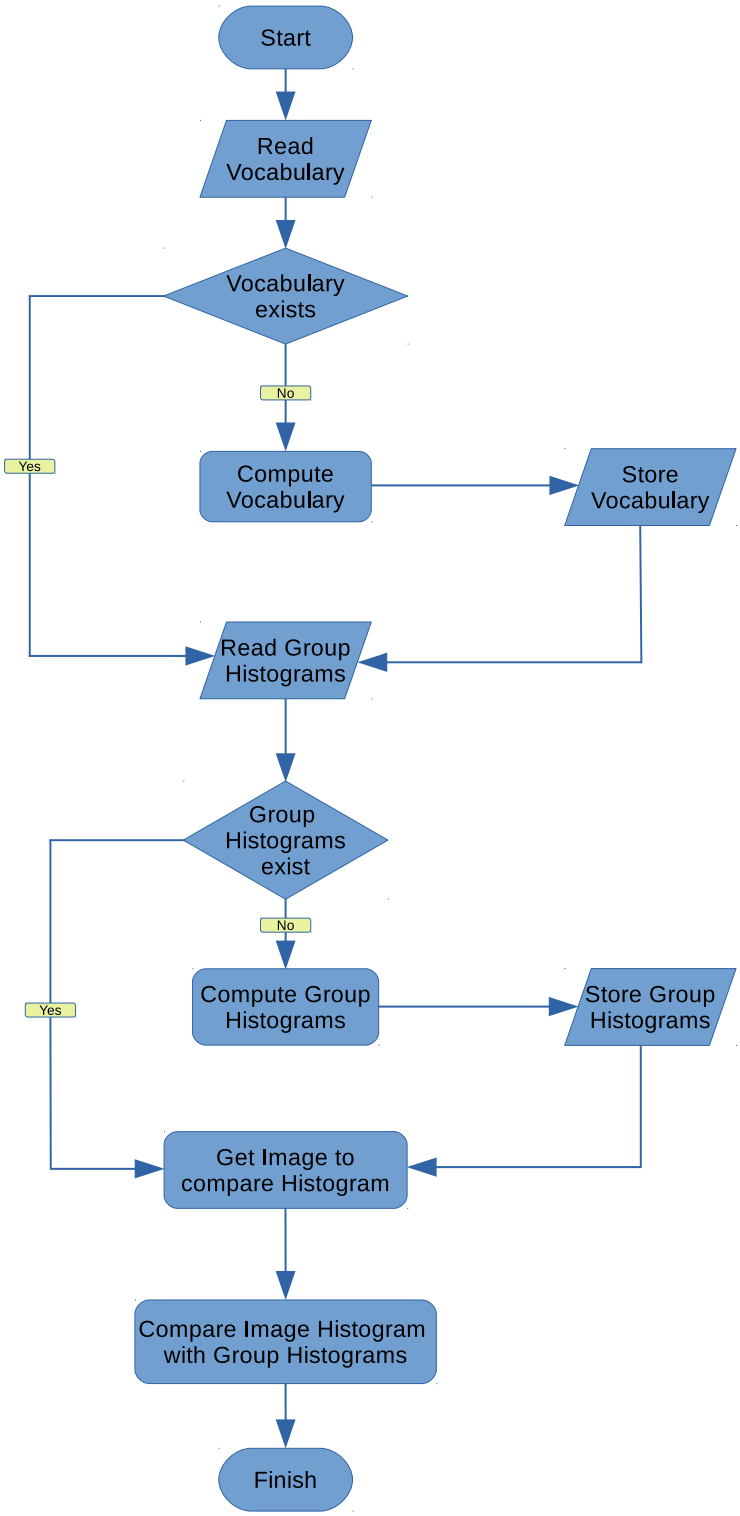


FIGURE 7.1: A general Bag of Visual Word Model

- Extract features based on the sub-segments and associate these characteristics to feature descriptors

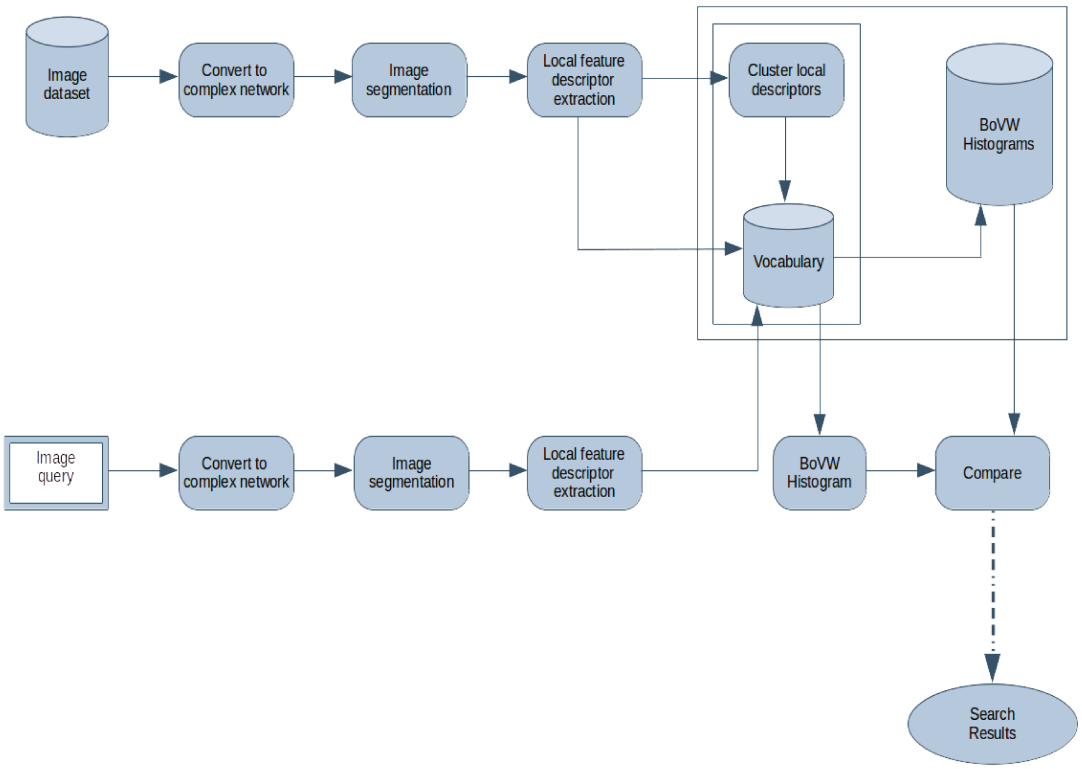


FIGURE 7.2: An illustration of our proposed CBIR Model

- Cluster the set of local descriptors using a K-Means algorithm to construct a vocabulary of K clusters.

2) For each feature descriptor in the image:

- Find the nearest visual word from the vocabulary for each feature vector using distance matching, for instance, BFMatcher, knnMatch, radiusMatch or FlannBasedMatcher
- Compute the Bag-of-Words image descriptor as a normalized histogram of vocabulary words encountered in the image
- Save the Bag-of-Words descriptors for all images in the dataset

7.3.2 Testing stage

At the test stage, the proposed system will perform for each input image query.

- The query image is converted to graph
- Segment the image based on community detection: Louvain method and ARA algorithm
- Extract features based on the sub-segments and associate these characteristics to feature descriptors

- Compute the Bag-of-Words vector with the method defined above
- Compare the histogram of query image with groups histograms and retrieve the best results

7.4 Experimental Results

We have implemented our CBIR systems to search the similar images to a query image from an image database which is collected from Corel dataset.

7.4.1 Dataset

In order to evaluate the proposed method, we selected the Corel dataset which is composed of 10 classes (Liu, Yang, and Li, 2015). Every class consists of 100 images on the same topic namely, Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. There is a total of 1000 images in the dataset for the experiment. We tested 1 image to all images in the dataset, so total test cases are 1000×1000 images. The experimental results are shown using qualitative and quantitative evaluations.

7.4.2 Results

For qualitative evaluations, we present some retrieved images from image queries implemented by proposed CBIR system. Detail of this qualitative evaluations are displayed in Figures 7.3, 7.4, 7.5 and 7.6.

For quantitative evaluations, we present a comparison of the proposed method with some methods using the same Corel dataset where each image is taken as a query image. If a retrieved image is within the same class of the query image, it is considered successful; otherwise it fails. The precision is given as Equation 7.1. In Table 7.1, we present Top 1 retrieved image excluding the query image itself on every test cases. We also display the statistics of Top 5 and Top 10 first candidates images on retrieved images results.

$$Precision = \frac{Number\ of\ Relevant\ Images}{Number\ of\ Relevant\ Images + Number\ of\ Irrelevant\ Images} \tag{7.1}$$

7.5 Conclusion

With recent advances in multimedia techniques and social networks, content-based image search has become an active topic of research. Thus, we have proposed in this Chapter a content-based image recovery system for real images. We have also implemented this system to look for the most similar images to a given image. This CBIR system is one of the representative applications for image segmentation. The combination of this CBIR system with the

FIGURE 7.3: Qualitative evaluation of results using image query.
First line: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

FIGURE 7.4: Qualitative evaluation of results using image query.
First line: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

Bag-of-Visual-Words model and the local features we have proposed and extracted makes it more efficient. Although our CBIR system has not been implemented on a large dataset or in deep and extensive experimental results, it has demonstrated its usefulness. We strongly believe that our research will be more advanced and deeper in several future studies.

FIGURE 7.5: Qualitative evaluation of results using image query.
First line: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

FIGURE 7.6: Qualitative evaluation of results using image query.
First line: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

Class name	Top 1	Top 5	Top 10
Africans (0-99)	70.00%	73.02%	59.48%
Beaches (100-199)	04.04%	34.35%	26.16%
Buildings (200-299)	25.25%	41.90%	33.18%
Buses (300-399)	76.77%	79.80%	75.20%
Dinosaurs (400-499)	95.96%	95.00%	92.00%
Elephants (500-599)	06.06%	26.93%	17.33%
Flowers (600-699)	66.67%	74.05%	69.25%
Horses (700-799)	56.57%	60.73%	51.88%
Mountains (800-899)	11.11%	35.15%	26.82%
Food (900-999)	79.80%	80.60%	76.80%
Total (10 Clases)	40.22%	60.15%	52.81%

TABLE 7.1: Precision proportion of Top 1, Top 5 and Top 10 implemented on dataset including all classes.

Methods	10 classes	9 classes	6 classes	5 classes
Huang's method (Huang and Chen, 2016)	74.41%	78.71%	86.51%	94.23%
Khosla's method (Khosla, Rajpal, and Singh, 2015)	55.00%	-	-	-
Srivastava's method (Srivastava, Prakash, and Khare, 2014)	67.16%	-	-	-
Choudhary's method (Choudhary et al., 2014)	-	75.00%	-	-
Choudhari's method (Chaudhari and Patil, 2012)	-	-	-	72.00%
Janani's method (Janani, 2014)	-	-	79.00%	-
Our method	49.22%	53.13%	74.79%	76.20%

TABLE 7.2: Comparison of the proposed method with others.

Chapter 8

Conclusion and future directions

8.1 Conclusions

This Thesis focused on a new image segmentation approach that relies on an analogy between image segmentation and community detection on graph with an extra agglomeration of homogeneous regions. Our methods produce efficient image segmentation results compared to the state of the art. The novelty of this research is to consider the strategy that needs to be implemented to generate a suitable network both in terms of the construction patterns and features to be computed after the execution of Louvain. Such a well constructed network is necessary to achieve good results, both in terms of execution time and quality, with Louvain algorithm. We performed extensive experiments using different patterns, different features and different datasets. The results show that the different algorithms that we proposed can reliably segment images and avoid over-segmentation. This produces more accurate objects and enhance computing performance efficiently.

Inspired by the efficiency of extracted features for image segmentation, we also attempted to set up an image search engine using Bag-of-Visual Word (BoVW) in which we used the features extracted for image segmentation. The content based image retrieval system is utilized for searching the closest images, in term of similarity, with a given query image. Our CBIR system relies on the cooperation of Bag-of-Visual-Words model and our extracted features offering better performances due to the indexation of local image features. Although the system was not implemented on huge dataset, its performances show the interest to continue in this direction

The following is a summary of the research that has been described in this thesis. Chapter 1 discussed the purpose and definition of image segmentation, and stated the objectives, the motivation, the problem description, our approach and the primary contributions of research. Chapter 2 described the state-of-the-art in image segmentation and made a taxonomy of image segmentation methods. Chapter 3 introduced complex networks, the concept of modularity for community detection and the Louvain algorithm. We

described the way to approach image segmentation using community detection, introduced how to generate an appropriate complex network, proposed an ARA algorithm for solving the over-segmentation issue and removed noise strategy. Chapter 4 introduced image features and extracted precise features for image processing and computer vision fields. We also proposed solutions to extract features that benefit in our research. Chapter 5 focused on extensive experiments to assess our methods. In this Chapter, we also introduced datasets, evaluation metrics, experimental results for each methods that has been described in Chapter 4. Chapter 6 introduced a new solution to the image segmentation problem based on the principle of consensual clustering that offers good results without the need of dealing with over-segmentation. The proposed approach is detailed and experiments are performed to show its interest. Chapter 7 introduced a content based image retrieval system that is built by the incorporation of our extracted features into Bag-of-Visual-Words model. This image search engine can search and grab images similar to a query image. Finally, this Chapter summarizes the Thesis and suggests potential directions for future research.

8.2 Future work

In the future research, we would like to extend our existing methods described in Chapter 4, and Chapter 7, i.e. the segmentation methods and the CBIR system. We would also like to do experimental evaluation of all methods listed in Chapter 4, and Chapter 7. Especially, we will implement extensively in a variety kinds of datasets and set up CBIR model in more cases of extraction features that we obtained when implemented image segmentation methods.

For image segmentation methods, there is still some space mainly in improving our segmentation methods, in particular:

- Making a combination between color property and spatial features to assign edge weights when we generate a graph. At the moment, edges weights are only 0 or 1 (i.e. edge is present or not) but since most community detection algorithms and the modularity can handle weighted graphs, discarding the weights is clearly a limitation of our approaches. This could be a short term evolution, however it would require testing several weighting functions and extensive experiments.

- Suggest and investigate more patterns to generate complex networks. In particular it might be interesting to test patterns that go only horizontally, vertically or diagonally to see whether such patterns might offer better performances on images with specific textures.

- Besides, the ARA algorithm could be used with a mix of image features properties both color, texture, spatial informative and shape features to produce better image segmentation results. In particular we could set up the implementation methods with more color spaces.

- We could also build a weighted graph which each region (superpixel) is a node, and the Cosine similarity distance (or any other distance) between two regions is an edge weight. The Louvain method would be applied once again for community detection instead of using ARA algorithm. To do that we can implement every proposed feature mentioned in Chapter 4.

- Inspired by the concepts of Cellular Automaton (CA), we can generate a variety of texture features that might help the ARA algorithm. It might be fascinating to study image segmentation in this branch.

- It might be interesting to study the Region Connection Calculus (RCC) theory. The set of relations well-known as RCC8 in the literature, forms a Jointly Exhaustive and Pairwise Disjoint (JEPD) set. Building a weighted graph whose regions are nodes and edge weights are determined by RCC8 relations might offer a different view on the problem. We could also use ARA algorithm when the criteria for merging two regions satisfy JEPD set.

- The consensual solution that we proposed could also benefit from all the previous modifications (adding weights, different patterns, more complex or mixed features). Furthermore the results provided with this solution use no cleaning or denoising and it is very likely that some cleaning would improve the results. Finding the correct threshold is also necessary for this approach to perform well.

- Finally it might be interesting to study relations between network embedding and image segmentation. Network embedding is used to project a network in a low dimensional space and the embedding might help in reducing the noise present in the image and therefore provide better defined communities.

For content based image retrieval system, we plan to implement experimental evaluation on more intensive and extensive with various datasets both quantity of images and a variety of objects in images. Because our CBIR system depends on the successes of the image segmentation result, all successful proposed methods for image segmentation problem could also improve the effectiveness of the CBIR system. We would also like to build an image search engine based on a mix image features properties obtained from our image segmentation methods and BoVW models.

Appendix A

List of Publications

International Conference and Workshop Papers

Thanh-Khoa Nguyen, Mickael Coustaty, and Jean-Loup Guillaume (2019). A Combination of Histogram of Oriented Gradients and Color Features to Cooperate with Louvain Method based Image Segmentation. In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, INSTICC. SciTePress*, pp. 280–291. ISBN : 978-989-758-354-4. DOI : 10.5220/0007389302800291.

Thanh-Khoa Nguyen, Mickael Coustaty, and Jean-Loup Guillaume (2018a). A New Image Segmentation Approach Based on the Louvain Algorithm. In: *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6. DOI : 10.1109/CBMI.2018.8516531.

Thanh-Khoa Nguyen, Mickael Coustaty, and Jean-Loup Guillaume (2018b). An Efficient Agglomerative Algorithm Cooperating with Louvain Method for Implementing Image Segmentation. In: *Advanced Concepts for Intelligent Vision Systems. Ed. by Jacques Blanc-Talon et al. Cham: Springer International Publishing*, pp. 150–162. ISBN : 978-3-030-01449-0.
DOI: https://doi.org/10.1007/978-3-030-01449-0_13

International Journal Papers

Thanh-Khoa Nguyen, Mickael Coustaty, and Jean-Loup Guillaume. A New Image Segmentation Approach Based on the Louvain Algorithm. *In Journal of Multimedia Tools and Applications (MTAP Journal)*. (under review)

Mickael Coustaty, Maximilien Danisch, Jean-Loup Guillaume, Thanh-Khoa Nguyen, Guillaume Serrano. A Consensus-Based Image Segmentation Algorithm Using Complex Networks. *In Journal of Pattern Recognition Letters*. (under review)

Book Chapters

Thanh-Khoa Nguyen, Jean-Loup Guillaume, and Mickael Coustaty. An Enhanced Louvain Based Image Segmentation approach using Color Properties and Histogram of Oriented Gradients. *In Book Series of Communications in Computer and Information Science (CCIS Book series)*. (accepted)

Bibliography

Abin, Ahmad Ali, Farzaneh Mahdisoltani, and Hamid Beigy (2011). "A New Image Segmentation Algorithm: A Community Detection Approach". In: *IICAI*.

Alahi, A., R. Ortiz, and P. Vandergheynst (2012). "FREAK: Fast Retina Keypoint". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510–517. DOI: 10.1109/CVPR.2012.6247715.

Albert, Réka and Albert-László Barabási (2002). "Statistical mechanics of complex networks". In: *Rev. Mod. Phys.* 74 (1), pp. 47–97. DOI: 10.1103/RevModPhys.74.47. URL: https://link.aps.org/doi/10.1103/RevModPhys.74.47.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási (1999). "Internet: Diameter of the World-Wide Web". In: *Nature* 401.6749, pp. 130–131. ISSN: 00280836. DOI: 10.1038/43601.

Arbelaez, P. et al. (2009). "From contours to regions: An empirical evaluation". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2294–2301. DOI: 10.1109/CVPR.2009.5206707.

Arbelaez, Pablo et al. (2011). "Contour Detection and Hierarchical Image Segmentation". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.5, pp. 898–916. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.161. URL: http://dx.doi.org/10.1109/TPAMI.2010.161.

Barabasi, Albert-Laszlo and Reka Albert (1999). "Albert, R.: Emergence of Scaling in Random Networks. Science 286, 509-512". In: *Science (New York, N.Y.)* 286, pp. 509–12. DOI: 10.1126/science.286.5439.509.

Barrat, Alain, Marc Barthlemy, and Alessandro Vespignani (2008). *Dynamical Processes on Complex Networks*. 1st. New York, NY, USA: Cambridge University Press. ISBN: 0521879507, 9780521879507.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool (2006). "SURF: Speeded Up Robust Features". In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 404–417. ISBN: 978-3-540-33833-8.

Beucher, Serge and Christian Lantuéjoul (1979). *Use of Watersheds in Contour Detection*. workshop published. URL: http://cmm.ensmp.fr/~beucher/publi/watershed.pdf.

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions". In: *Bulletin of the Calcutta Mathematical Society* 35, pp. 99–109.

Blondel, V. D. et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 10, p. 10008. DOI: 10.1088/1742-5468/2008/10/P10008. arXiv: 0803.0476 [physics.soc-ph].

Boccaletti, Stefano et al. (2006). "Complex networks: Structure and dynamics". In: *Physics Reports* 424, pp. 175–308. DOI: 10.1016/j.physrep.2005.10.009.

Browet, Arnaud, P. A. Absil, and Paul Van Dooren (2011). "Community Detection for Hierarchical Image Segmentation". In: *Combinatorial Image Analysis*. Ed. by Jake K. Aggarwal et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 358–371. ISBN: 978-3-642-21073-0.

C Gonzalez, Rafael and Richard E Woods (2002). *Digital Image Processing (2nd Edition)*. ISBN: 0201180758.

Chaudhari, R. and A. M. Patil (2012). "Content Based Image Retrieval Using Color and Shape Features". In: *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*. Vol. Vol. 1. No. 5, pp. 386–392.

Chen, Y. L. et al. (2008). "Vision-based nighttime vehicle detection and range estimation for driver assistance". In: *2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 2988–2993. DOI: 10.1109/ICSMC.2008.4811753.

Chien, Shao-Yi, Yu-Wen Huang, and Liang-Gee Chen (2003). "Predictive watershed: a fast watershed algorithm for video segmentation". In: *IEEE Trans. Circuits Syst. Video Techn.* 13, pp. 453–461.

Choudhary, R. et al. (2014). "An integrated approach to Content Based Image Retrieval". In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2404–2410. DOI: 10.1109/ICACCI.2014.6968394.

Christoudias, C. M., B. Georgescu, and P. Meer (2002). "Synergism in low level vision". In: *Object recognition supported by user interaction for service robots*. Vol. 4, 150–155 vol.4. DOI: 10.1109/ICPR.2002.1047421.

Clauset, Aaron, M E J Newman, and Cristopher Moore (2005). "Finding community structure in very large networks". In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 70, p. 066111. DOI: 10.1103/PhysRevE.70.066111.

Comaniciu, Dorin and Peter Meer (2002). "Mean Shift: A Robust Approach Toward Feature Space Analysis". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 24.5, pp. 603–619. ISSN: 0162-8828. DOI: 10.1109/34.1000236. URL: http://dx.doi.org/10.1109/34.1000236.

Cortés Llosa, Xavier, Donatello Conte, and Hubert Cardot (2018). "Bags of Graphs for Human Action Recognition: Joint IAPR International Workshop, S+SSPR 2018, Beijing, China, August 17–19, 2018, Proceedings". In: pp. 429–438. ISBN: 978-3-319-97784-3. DOI: 10.1007/978-3-319-97785-0_41.

Cour, T., F. Benezit, and J. Shi (2005). "Spectral segmentation with multiscale graph decomposition". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2, 1124–1131 vol. 2. DOI: 10.1109/CVPR.2005.332.

Csurka, Gabriella et al. (2004). "Visual categorization with bags of keypoints". In: *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22.

Dalal, N. and B. Triggs (2005). "Histograms of oriented gradients for human detection". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

Dang, Quoc Bao et al. (2019). "A comparison of local features for camera-based document image retrieval and spotting". In: *International Journal on Document Analysis and Recognition (IJDAR)* 22.3, pp. 247–263. ISSN: 1433-2825. DOI: 10.1007/s10032-019-00329-w. URL: https://doi.org/10.1007/s10032-019-00329-w.

Deng, Yining and b. s. Manjunath (2001). "Unsupervised Segmentation of Color-Texture Regions in Images and Video". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 23.8, pp. 800–810. ISSN: 0162-8828. DOI: 10.1109/34.946985. URL: https://doi.org/10.1109/34.946985.

Dorogovtsev, S.N. and J.F.F. Mendes (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press. URL: https://EconPapers.repec.org/RePEc:oxp:obooks:9780198515906.

E J Newman, M (2004). "Analysis of Weighted Networks". In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 70, p. 056131. DOI: 10.1103/PhysRevE.70.056131.

E. J. Newman, Mark (2003). "Newman MEJ.. The structure and function of complex networks. SIAM Rev 45: 167-256". In: *SIAM Review* 45. DOI: 10.1137/S003614450342480.

Easley, David and Jon Kleinberg (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press. DOI: 10.1017/CBO9780511761942.

Euler, Leonhard (1736). "Solutio problematis ad geometriam situs pertinentis". In: *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8, pp. 128–140.

Felzenszwalb, Pedro F. and Daniel P. Huttenlocher (2004). "Efficient Graph-Based Image Segmentation". In: *International Journal of Computer Vision* 59.2, pp. 167–181. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000022288.19776.77. URL: http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77.

Fortunato, Santo (2010). "Community detection in graphs". In: *Physics Reports* 486.3–5, pp. 75 –174. ISSN: 0370-1573. DOI: http://dx.doi.org/10.1016/j.physrep.2009.11.002. URL: http://www.sciencedirect.com/science/article/pii/S0370157309002841.

Fortunato, Santo and Marc Barthélemy (2007). "Resolution limit in community detection". In: *Proceedings of the National Academy of Sciences* 104.1, pp. 36–41. ISSN: 0027-8424. DOI: 10.1073/pnas.0605965104. eprint: http://www.pnas.org/content/104/1/36.full.pdf. URL: http://www.pnas.org/content/104/1/36.

Girvan, M. and M. E. J. Newman (2002). "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 7821–7826. ISSN: 0027-8424. DOI: 10.1073/pnas.122653799. eprint: https://www.pnas.org/content/99/12/7821.full.pdf. URL: https://www.pnas.org/content/99/12/7821.

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". In: *Applied Statistics* 28.1, pp. 100–108. ISSN: 00359254. DOI: 10.2307/2346830. URL: http://dx.doi.org/10.2307/2346830.

He, Dong-Chen and Li Wang (1990). "Texture Unit, Texture Spectrum, And Texture Analysis". In: *IEEE Transactions on Geoscience and Remote Sensing* 28.4, pp. 509–512. ISSN: 0196-2892. DOI: 10.1109/TGRS.1990.572934.

He, K. et al. (2016). "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

Horowitz, S. L. and T. Pavlidis (1974). "Picture Segmentation by a directed split-and-merge procedure". In: *Proceedings of the 2nd International Joint Conference on Pattern Recognition, Copenhagen, Denmark*, pp. 424–433.

Huang, Yin-Fu and Bo-Rong Chen (2016). "Content-Based Image Retrieval System for Real Images". In: *GCAI 2016. 2nd Global Conference on Artificial Intelligence*. Ed. by Christoph Benzm\"uller, Geoff Sutcliffe, and Raul Rojas. Vol. 41. EPiC Series in Computing. EasyChair, pp. 95–108. DOI: 10.29007/w4sr. URL: https://easychair.org/publications/paper/Z3T.

Ikonomatakis, N. et al. (1997). "Region growing and region merging image segmentation". In: *Proceedings of 13th International Conference on Digital Signal Processing*. Vol. 1, 299–302 vol.1. DOI: 10.1109/ICDSP.1997.628077.

Janani, Rajeswari Gopalakrishnan (2014). "An Improved CBIR Method Using Color andTexture Properties with Relevance Feedback". In: *International Journal of Innovative Research in Computer and Communication Engineering*. Vol. Vol. 2. Special Issue 1, pp. 47–54.

Junwei, L. and L. Shaokai (2013). "A novel image segmentation technology in intelligent traffic light control systems". In: *2013 3rd International Conference on Consumer Electronics, Communications and Networks*, pp. 26–29. DOI: 10.1109/CECNet.2013.6703263.

Khan, N. and G. Sundaramoorthi (2018). "Learned Shape-Tailored Descriptors for Segmentation". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 666–674. DOI: 10.1109/CVPR.2018.00076.

Khosla, G., N. Rajpal, and J. Singh (2015). "Evaluation of Euclidean and Manhanttan metrics in Content Based Image Retrieval system". In: *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 12–18.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'12. USA: Curran Associates Inc., pp. 1097–1105. URL: http://dl.acm.org/citation.cfm?id=2999134.2999257.

Leutenegger, Stefan, Margarita Chli, and Roland Siegwart (2011). "BRISK: Binary Robust invariant scalable keypoints". In: pp. 2548–2555. DOI: 10.1109/ICCV.2011.6126542.

Li, S. and D. O. Wu (2015). "Modularity-Based Image Segmentation". In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.4, pp. 570–581. ISSN: 1051-8215. DOI: 10.1109/TCSVT.2014.2360028.

Li, Wenye (2013). "Modularity Segmentation". In: *Neural Information Processing*. Ed. by Minho Lee et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 100–107. ISBN: 978-3-642-42042-9.

Linares, Oscar A. C. et al. (2016). "Segmentation of large images based on super-pixels and community detection in graphs". In: *CoRR* abs/1612.03705. URL: http://arxiv.org/abs/1612.03705.

Liu, D. and T. Chen (2008). "DISCOV: A Framework for Discovering Objects in Video". In: *IEEE Transactions on Multimedia* 10.2, pp. 200–208. ISSN: 1520-9210. DOI: 10.1109/TMM.2007.911781.

Liu, Guang-Hai, Jing-Yu Yang, and ZuoYong Li (2015). "Content-based image retrieval using computational visual attention model". In: *Pattern Recognition* 48.8, pp. 2554 –2566. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2015.02.005. URL: http://www.sciencedirect.com/science/article/pii/S0031320315000539.

Liu, Jialu (2013). "Image Retrieval based on Bag-of-Words model". In: *CoRR* abs/1304.5168. arXiv: 1304.5168. URL: http://arxiv.org/abs/1304.5168.

Liu, T., M. Seyedhosseini, and T. Tasdizen (2016). "Image Segmentation Using Hierarchical Merge Tree". In: *IEEE Transactions on Image Processing* 25.10, pp. 4596–4607. ISSN: 1057-7149. DOI: 10.1109/TIP.2016.2592704.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965. URL: https://doi.org/10.1109/CVPR.2015.7298965.

Lowe, David G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints". In: *International Journal of Computer Vision* 60.2, pp. 91–110. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000029664.99615.94. URL: https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Mahalanobis, P. C. (1936). "On the generalised distance in statistics". In: *Proceedings National Institute of Science, India*. Vol. 2. 1, pp. 49–55. URL: http://ir.isical.ac.in/dspace/handle/1/1268.

Mandal, Ranju et al. (2018). "Bag-of-Visual-Words for Signature-Based Multi-Script Document Retrieval". In: *CoRR* abs/1807.06772. arXiv: 1807.06772. URL: http://arxiv.org/abs/1807.06772.

Manzo, Mario and Simone Pellino (2019). "Bag of ARSRG Words (BoAW)". In: *Machine Learning and Knowledge Extraction* 1, pp. 871–882. DOI: 10.3390/make1030050.

Martin, D. et al. (2001). "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics". In: *Proc. 8th Int'l Conf. Computer Vision*. Vol. 2, pp. 416–423.

Matas, Jiri et al. (2002). "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions". In: vol. 22(10). DOI: 10.5244/C.16.36.

Medioni, Gerard and Yoshio Yasumoto (1987). "Corner detection and curve representation using cubic B-splines". In: *Computer Vision, Graphics, and Image Processing* 39.3, pp. 267 –278. ISSN: 0734-189X. DOI: https://doi.

org/10.1016/S0734-189X(87)80181-0. URL: http://www.sciencedirect.com/science/article/pii/S0734189X87801810.

Meilă, Marina (2005). "Comparing clusterings: an axiomatic view". In: *In ICML '05: Proceedings of the 22nd international conference on Machine learning*. ACM Press, pp. 577–584.

Meyer, F. (1989). "Skeletons and perceptual graphs". In: *Signal Processing* 16.4. Special Issue on Advances in Mathematical Morphology, pp. 335 –363. ISSN: 0165-1684. DOI: https://doi.org/10.1016/0165-1684(89)90030-3. URL: http://www.sciencedirect.com/science/article/pii/0165168489900303.

Mikolajczyk, K. and C. Schmid (2005). "A performance evaluation of local descriptors". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.10, pp. 1615–1630. DOI: 10.1109/TPAMI.2005.188.

Mikolajczyk, K. et al. (2005). "A Comparison of Affine Region Detectors". In: *International Journal of Computer Vision* 65.1, pp. 43–72. ISSN: 1573-1405. DOI: 10.1007/s11263-005-3848-x. URL: https://doi.org/10.1007/s11263-005-3848-x.

Mori, G. (2005). "Guiding model search using segmentation". In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Vol. 2, 1417–1423 Vol. 2. DOI: 10.1109/ICCV.2005.112.

Mourchid, Youssef, Mohammed El Hassouni, and Hocine Cherifi (2017). "An Image Segmentation Algorithm based on Community Detection". In: *Complex Networks & Their Applications V*. Ed. by Hocine Cherifi et al. Cham: Springer International Publishing, pp. 821–830. ISBN: 978-3-319-50901-3.

Mukherjee, J., J. Mukhopadhyay, and P. Mitra (2014). "A survey on image retrieval performance of different bag of visual words indexing techniques". In: *Proceedings of the 2014 IEEE Students' Technology Symposium*, pp. 99–104. DOI: 10.1109/TechSym.2014.6807922.

Newman, M. E. J. (2004). "Detecting community structure in networks". In: *The European Physical Journal B* 38.2, pp. 321–330. ISSN: 1434-6036. DOI: 10.1140/epjb/e2004-00124-y. URL: http://dx.doi.org/10.1140/epjb/e2004-00124-y.

Newman, Mark (2010). *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc. ISBN: 0199206651, 9780199206650.

Newman, Mark EJ and Michelle Girvan (2004). "Finding and evaluating community structure in networks". In: *Physical Review E* 69.2, p. 026113.

Newman, M.E.J. (2003). "Fast algorithm for detecting community structure in networks". In: *Physical Review E* 69. URL: http://arxiv.org/abs/cond-mat/0309508.

Nguyen, T., M. Coustaty, and J. Guillaume (2018a). "A New Image Segmentation Approach Based on the Louvain Algorithm". In: *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6. DOI: 10.1109/CBMI.2018.8516531.

Nguyen, Thanh-Khoa, Mickael Coustaty, and Jean-Loup Guillaume (2018b). "An Efficient Agglomerative Algorithm Cooperating with Louvain Method

for Implementing Image Segmentation". In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by Jacques Blanc-Talon et al. Cham: Springer International Publishing, pp. 150–162. ISBN: 978-3-030-01449-0.

Nguyen, Thanh-Khoa, Mickael Coustaty, and Jean-Loup Guillaume (2019). "A Combination of Histogram of Oriented Gradients and Color Features to Cooperate with Louvain Method based Image Segmentation". In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP*, INSTICC. SciTePress, pp. 280–291. ISBN: 978-989-758-354-4. DOI: 10.5220/0007389302800291.

Okawa, M. (2016). "Offline Signature Verification Based on Bag-of-VisualWords Model Using KAZE Features and Weighting Schemes". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 252–258. DOI: 10.1109/CVPRW.2016.38.

Pantofaru, Caroline and Martial Hebert (2005). *A Comparison of Image Segmentation Algorithms*. Tech. rep. CMU-RI-TR-05-40. Pittsburgh, PA: Carnegie Mellon University.

Pastor-Satorras, Romualdo and Alessandro Vespignani (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York, NY, USA: Cambridge University Press. ISBN: 0521826985.

Potapov, Alexey et al. (2018). "Semantic Image Retrieval by Uniting Deep Neural Networks and Cognitive Architectures". In: *CoRR* abs/1806.06946. arXiv: 1806.06946. URL: http://arxiv.org/abs/1806.06946.

Ren and Malik (2003). "Learning a classification model for segmentation". In: *Proceedings Ninth IEEE International Conference on Computer Vision*, 10–17 vol.1. DOI: 10.1109/ICCV.2003.1238308.

Rosten, E., R. Porter, and T. Drummond (2010). "Faster and Better: A Machine Learning Approach to Corner Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.1, pp. 105–119. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2008.275.

Rosten, Edward and Tom Drummond (2005). "Fusing Points and Lines for High Performance Tracking". In: *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*. ICCV '05. Washington, DC, USA: IEEE Computer Society, pp. 1508–1515. ISBN: 0-7695-2334-X-02. DOI: 10.1109/ICCV.2005.104. URL: https://doi.org/10.1109/ICCV.2005.104.

Sadeghi-Tehran, Pouria et al. (2019). "Scalable Database Indexing and Fast Image Retrieval Based on Deep Learning and Hierarchically Nested Structure Applied to Remote Sensing and Plant Biology". In: *Journal of Imaging* 5.3. ISSN: 2313-433X. DOI: 10.3390/jimaging5030033. URL: https://www.mdpi.com/2313-433X/5/3/33.

Scott, John (2000). *Social Network Analysis: A Handbook*. 2nd. London: Sage Publications. DOI: 10.1002/cpe.607.

Seifi, Massoud et al. (2013). "Stable Community Cores in Complex Networks". In: *Complex Networks*. Ed. by Ronaldo Menezes, Alexandre Evsukoff, and Marta C. González. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 87–98. ISBN: 978-3-642-30287-9. DOI: 10.1007/978-3-642-30287-9_10. URL: https://doi.org/10.1007/978-3-642-30287-9_10.

Sharma, Gaurav, Wencheng Wu, and Edul Dalal (2005). "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations". In: *Color Research Application* 30, pp. 21–30. DOI: 10.1002/col.20070.

Shi, Jianbo and J. Malik (2000). "Normalized cuts and image segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8, pp. 888–905. ISSN: 0162-8828. DOI: 10.1109/34.868688.

Shi, Jianbo and Carlo Tomasi (1994). "Good features to track". In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600. DOI: 10.1109/CVPR.1994.323794.

Shotton, Jamie et al. (2006). "TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation". In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–15. ISBN: 978-3-540-33833-8.

Silva, Fernanda et al. (2013). "Image classification based on bag of visual graphs". In: pp. 4312–4316. ISBN: 978-1-4799-2341-0. DOI: 10.1109/ICIP.2013.6738888.

Silva, Fernanda B. et al. (2018). "Graph-based bag-of-words for classification". In: *Pattern Recognition* 74, pp. 266 –285. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2017.09.018. URL: http://www.sciencedirect.com/science/article/pii/S0031320317303680.

Simonyan, Karen and Andrew Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv 1409.1556*.

Smith, Stephen M. and J. Michael Brady (1997). "SUSAN—A New Approach to Low Level Image Processing". In: *International Journal of Computer Vision* 23.1, pp. 45–78. ISSN: 1573-1405. DOI: 10.1023/A:1007963824710. URL: https://doi.org/10.1023/A:1007963824710.

Srivastava, P., O. Prakash, and A. Khare (2014). "Content-Based Image Retrieval using moments of wavelet transform". In: *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*, pp. 159–164. DOI: 10.1109/ICCAIS.2014.7020550.

Strehl, A. and J. Ghosh (2002). "Cluster ensembles – a knowledge reuse framework for combining multiple partitions". In: *Journal on Machine Learning Research (JMLR)* 3, pp. 583–61710008.

Sumengen, B. and B. S. Manjunath (2005). "Multi-scale edge detection and image segmentation". In: *2005 13th European Signal Processing Conference*, pp. 1–4.

Szegedy, Christian et al. (2014). "Going Deeper with Convolutions". In: *CoRR* abs/1409.4842. arXiv: 1409.4842. URL: http://arxiv.org/abs/1409.4842.

Tsai, Chih-Fong (2012). "Bag-of-Words Representation in Image Annotation: A Review". In: *ISRN Artificial Intelligence* 2012. DOI: 10.5402/2012/376804.

Verdoja, Francesco and Marco Grangetto (2015). "Fast Superpixel-Based Hierarchical Approach to Image Segmentation". In: *Image Analysis and Processing — ICIAP 2015*. Ed. by Vittorio Murino and Enrico Puppo. Cham: Springer International Publishing, pp. 364–374. ISBN: 978-3-319-23231-7.

Vincent, L. and P. Soille (1991). "Watersheds in digital spaces: an efficient algorithm based on immersion simulations". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.6, pp. 583–598. ISSN: 0162-8828. DOI: 10.1109/34.87344.

Wan, Ji et al. (2014). "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study". In: *Proceedings of the 22Nd ACM International Conference on Multimedia*. MM '14. ACM, pp. 157–166. ISBN: 978-1-4503-3063-3. DOI: 10.1145/2647868.2654948. URL: http://doi.acm.org/10.1145/2647868.2654948.

Wang, Hui, Dzulkifli Mohamad, and Nor Azman Ismail (2009). "Image Retrieval: Techniques, Challenge, and Trend". In: *World Academy of Science, Engineering and Technology* 60.

Wang, Hui Hui, Dzulkifli Mohamad, and Nor Azman Ismail (2010). "Approaches, Challenges and Future Direction of Image Retrieval". In: *CoRR* abs/1006.4568. arXiv: 1006.4568. URL: http://arxiv.org/abs/1006.4568.

Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press. DOI: 10.1017/CBO9780511815478.

Watanabe, Tomoki, Satoshi Ito, and Kentaro Yokoi (2008). "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection". In: *Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*. PSIVT '09. Springer-Verlag, pp. 37–47. ISBN: 978-3-540-92956-7. DOI: 10.1007/978-3-540-92957-4_4. URL: http://dx.doi.org/10.1007/978-3-540-92957-4_4.

Youssef Mourchid, Mohammed El Hassouni and Hocine Cherifi (2015). "A New Image Segmentation Approach using Community Detection Algorithms". In: *15th International Conference on Intelligent Systems Design and Applications, Marrakesh, Marocco*.

Youssef Mourchild, Mohammed El Hassouni and Hocine Cherifi (2016). "Image segmentation based on community detection approach". In: *International Journal of Computer Information Systems and Industrial Management Applications* 8, pp. 195–204. ISSN: 2150-7988.

Zhou, J. Y., E. P. Ong, and C. C. Ko (2000). "Video object segmentation and tracking for content-based video coding". In: *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532)*. Vol. 3, 1555–1558 vol.3. DOI: 10.1109/ICME.2000.871065.

Zhu, Song-Chun et al. (2005). "What are Textons?" In: *International Journal of Computer Vision* 62.1, pp. 121–143. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000046592.70770.61. URL: https://doi.org/10.1023/B:VISI.0000046592.70770.61.

Zohrizadeh, Fariba, Mohsen Kheirandishfard, and Farhad Kamangar (2018). "Image Segmentation using Sparse Subset Selection". In: *CoRR* abs/1804.02721. arXiv: 1804.02721. URL: http://arxiv.org/abs/1804.02721.