Prise en compte de l'esthétique dans la gestion des gammes de luminance des images

Jing Zhang

Thèse de Doctorat

Mention : Sciences et Technologies de l'Information et de la Communication
Spécialité : Informatique et applications

présentée à *l'École Doctorale en Sciences Technologie et Santé (ED 585)*

de l'Université du Littoral Côte d'Opale

par

Jing Zhang

pour obtenir le grade de Docteur de l'Université du Littoral Côte d'Opale

Prise en compte de l'esthétique dans la gestion des gammes de luminance des images

Soutenue le 12 juillet 2024, après avis des rapporteurs, devant le jury d'examen :

M. Christophe RENAUD	**Professeur, Université du Littoral Côte d'Opale**	**Président**
Mme. Solène KALÉNINE	**Chargée de Recherche, Université de Lille**	**Rapporteure**
M. Chaker LARABI	**Professeur, Université de Poitiers**	**Rapporteur**
Mme. Maud MARCHAL	**Professeure, INSA Rennes**	**Examinatrice**
M. Giuseppe VALENZISE	**Chargé de Recherche, Université Paris-Saclay**	**Examinateur**
M. Rémi COZOT	**Professeur, Université du Littoral Côte d'Opale**	**Directeur de thèse**
M. Rémi SYNAVE	**Maître de Conférences, Université du Littoral Côte d'Opale**	**Co-Directeur de thèse**
M. Samuel DELEPOULLE	**Maître de Conférences, Université du Littoral Côte d'Opale**	**Invité**
M. Daniel MENEVEAUX	**Professeur, Université de Poitiers**	**Invité**

PRISE EN COMPTE DE L'ESTHÉTIQUE DANS LA GESTION DES GAMMES DE LUMINANCE DES IMAGES

Résumé

L'analyse des caractéristiques esthétiques d'images numériques permettent d'améliorer la qualité esthétique du contenu visuel. En analysant les caractéristiques esthétiques qui influencent la perception visuelle à travers les données de l'image, les ordinateurs peuvent effectuer des tâches telles que l'édition assistée d'image, l'amélioration de la qualité esthétique et le filtrage de la meilleure image. Cette thèse intègre l'analyse des caractéristiques esthétique des images avec l'imagerie à haute gamme dynamique (HDR). Nous prenons en compte les propriétés du HDR et les caractéristiques esthétiques lors du traitement des images HDR. L'objectif est de maximiser la préservation des caractéristiques esthétiques originales des images lors de l'ajustement des effets d'affichage HDR, afin d'atteindre l'expérience visuelle plus agréable. Dans cette thèse, nous proposons deux approches d'auto-ajustement des images HDR et une méthode de reconstruction des lignes de force de la composition.

Concernant l'auto-ajustement des images HDR, nous développons un modèle basé sur un réseau de neurones pour prédire la courbe d'ajustement des images HDR, et un modèle utilisant un réseau de neurones convolutifs pour estimer la valeur d'ajustement de l'exposition, en analysant les caractéristiques potentielles des images HDR. Ces deux méthodes consistent à améliorer automatiquement la perception de la qualité esthétique des images HDR sur des dispositifs d'affichage HDR. Elles le font en entraînant des réseaux de neurones à apprendre des paramètres d'édition d'experts à partir de jeux de données HDR. Afin d'analyser l'esthétique de la composition d'une image, nous proposons de reconstruire les lignes de force. Tout comme la couleur, les lumières ou le grain de l'image, les lignes de force font partie des caractéristiques esthétique qui doivent être analysées. La méthode proposée identifie les lignes de force implicites dans l'image par un algorithme de regroupement des lignes. Nous avons initialement mené une analyse de cohérence entre experts pour démontrer la faisabilité de notre méthode. Par ailleurs, nous proposons une métrique pour comparer les deux ensembles de lignes de force.

Mots clés : esthétique d'image, imagerie HDR, perception visuelle, apprentissage automatique

Abstract

Aesthetic analysis of digital images enhances the visual content's aesthetic quality. By analyzing aesthetic features that influence visual perception through image data, computers can perform tasks like assisted image editing, aesthetic quality enhancement, and filtering for the best image. This thesis merges aesthetic image analysis with high dynamic range (HDR) imaging. We consider both the properties of HDR and the aesthetic characteristics of images during HDR image processing. The aim is to maximize the preservation of the original aesthetic features of images when adjusting HDR image display results, thereby achieving a pleasant visual experience. In this thesis, we propose a composition leading lines reconstruction method and two HDR image auto-adjustment methods.

Regarding automatic adjustment of HDR images, we are developing a model based on a neural network to predict the adjustment curve of HDR images, and a model using a convolutional neural network to estimate the exposure adjustment value, by analyzing potential features of HDR images. Both methods automatically enhance the aesthetic quality perception of HDR images on HDR display devices by training neural networks to learn expert editing parameters from an HDR database. In order to analyze the aesthetics of image composition, we propose to reconstruct the leading lines of the image. Just like color, lighting, or the grain of the image, the leading lines are among the aesthetic features that need to be analyzed. The proposed method identifies implicit leading lines in the image through a line regrouping algorithm. We initially carried out an inter-expert consistency analysis to demonstrate the feasibility of our method. In addition, we propose a metric for comparing the two sets of leading lines.

Keywords: image aesthetics, HDR image, visual perception, machine learning

Remerciements

Je souhaite premièrement remercier mon directeur de thèse, Rémi Cozot, pour son encadrement exemplaire, son écoute attentive et son soutien constant. Sa patience, ses encouragements et sa confiance ont été essentiels tout au long de ma thèse.

Je souhaite remercier également mon encadrant, Rémi Synave, pour ses conseils et son soutien. Grâce à son humour et son optimisme, il rend les difficultés moins pénibles. Je tiens également à remercier Samuel Delepoulle, pour ses explications minutieuses tout au long de mon travail de recherche. Sa rigueur et son attention aux détails m'ont été très bénéfiques.

Je voudrais remercier sincèrement Céline Loscos, pour ses conseils et son aide précieuse. Je remercie Francesco Banterle, avec qui nous avons passé de très bons moments de recherche. Je remercie Jérôme Buisine pour avoir patiemment répondu à toutes mes questions.

Je voudrais remercier Chaker Larabi et Solène Kalénine pour avoir accepté de rapporter ce document, et plus globalement tous les membres du jury pour avoir accepté d'examiner mes travaux de thèse.

Je voudrais remercier Daniel Méneveaux et Gilles Roussel pour leur écoute et leurs encouragements. Je tiens à remercier tous les membres de mon équipe dans notre laboratoire LISIC. Je chéris le temps que j'ai passé avec eux pendant mes travaux de recherche.

J'aimerais remercier mes parents pour leur soutien inconditionnel, ainsi que tous mes amis et ma famille pour leurs encouragements. Je remercie mon amie Li Ruijuan pour son soutien enthousiaste, chaque fois que je lui parle, j'en retire beaucoup. Je tiens particulièrement à remercier Xiaozhou pour son accompagnement et son soutien quotidiens.

Table des matières

Glossaire

E | G | L | T

E

esthétique Dans le domaine de recherche de cette thèse, l'esthétique représente l'esthétique de l'image, qui se réfère à la perception visuelle que les images apportent aux humains, et au sentiment de beauté que les images transmettent au système visuel humain à travers la présentation de leurs propres caractéristiques. Les caractéristiques esthétiques se rapportent plus au style de l'image ; la qualité esthétique se rapporte plus au sens de la beauté. 4, 6

G

gamme dynamique La gamme dynamique est le rapport entre les valeurs maximales et minimales de la luminance. 4

L

lumière Rayonnement émis par des crops portés à haute température ou par des substances excitées qui est perçu par les yeux et fournit un éclairage. 4

luminance Luminance est la densité de l'intensité lumineuse par rapport à la surface projetée dans une direction déterminée en un point déterminé d'une surface réelle ou imaginaire. 4

luminosité Qualité de ce qui est lumineux. 21

T

Tone Mapping Tone Mapping est l'opération qui consiste à ajuster la gamme dynamique d'un contenu à haute gamme dynamique pour l'adapter à la gamme dynamique réduite d'un dispositif d'affichage. 6

Acronymes

Symboles | A | B | C | E | F | G | H | I | L | M | N | P | S | T | U | V

Symboles

cd/m^2 *Candela per meter squared, Candela par mètre carré* . 4

A

ACQUINE *Aesthetic Quality Inference Engine, Moteur d'inférence de qualité esthétique*. 58

ASP *Adaptive spatial pooling, Pooling spatial adaptatif* . 60

B

BOV *Bag-of-Visual-words, Sac-de-mots*. 58

C

CNN *Convolutional Neural Networks, Réseaux neuronaux convolutifs*. 59

E

EMD *Earth Mover's Distance, Distance normalisée de Earth Mover's* . 61

F

FV *Fisher Vector, le vecteur de Fisher*. 58

G

GCN *Graph Convolutional Network, Réseau convolutionnel graphique*. 62

H

HDR *High Dynamic Range, Grande gamme dynamique*. 5, 100

HDR-VDP *Visual Difference Predictor for HDR images, Prédicteur de différence visuelle pour les images HDR*. 40, 51

I

ILSVRC *ImageNet Large-Scale Visual Recognition Challenge*. 59

L

LDR *Low Dynamic Range, Faible gamme dynamique* . 5
LED *Light-Emitting Diode, Diode électroluminescente* . 12

M

MLSP *Multi-level Spatially Pooled activation blocks, Blocs d'activation multi-niveaux à pool spatiale*. 60
MSE *Mean Squared Error, Erreur quadratique moyenne*. 38, 46

N

NIMA *Neural Image Assessment, Évaluation de l'image neuronale* . 61

P

PCA *Principal Component Analysis, Analyses en composantes principales*. 57
PQ *Perceptual Quantizer, Quantificateur perceptif* . 51
PSNR *Peak Signal to Noise Ratio*. 40, 51
PU *Perceptually Uniform, Uniformité perceptive* . 30, 40, 51

S

SSIM *Structural Similarity, Similarité structurelle*. 40, 51
SVM *Support Vector Machines, Machine à vecteurs de support, ou séparateur à vaste marge*. 57

T

TMO *Tone Mapping Operator, Opérateur de mappage tonalité* . 6
TMQI *Tone Mapping Image Quality Index, Indice de qualité des images de mappage de tonalité*. 30
TVHD *High-definition television, Télévision Haute Définition* . 16

U

UTVHD *Ultra High Definition Television, Télévision à ultra-haute définition* . 16

V

VESA *Video Electronics Standards Association, Association pour les normes de l'électronique vidéo*. 5

Chapitre 1

Introduction

L'évolution de la technologie numérique a élevé l'image au rang de moyen de communication incontournable dans plusieurs domaines, tels que la diffusion d'informations, les médias sociaux, l'art et le design. Fondées sur les images numériques, des disciplines comme la vision par ordinateur et la photographie informatisée ont ouvert de nouvelles perspectives et présenté des défis inédits aux approches traditionnelles. En tant que vecteur d'informations visuelles, les images capturent les détails objectifs des scènes tout en cherchant à exprimer les émotions avec l'esthétique voulues par leurs créateurs. La restitution fidèle de la richesse d'une scène à travers le traitement numérique des ombres et lumière représente un enjeu majeur dans le domaine du traitement de l'image.

En général, on admet que les premières images produites au cours de l'histoire étaient des peintures murales dans les grottes ou des gravures rupestres, elles sont regroupées sous le nom d'art pariétal. Avec le développement de la civilisation, les fresques, les sculptures, les gravures, les dessins et les peintures sont également devenus successivement des moyens d'expression d'images visuelles. Tout en enregistrant les histoires de leur temps, les artistes ont également exprimé leur compréhension subjective de l'esthétique à travers les lignes et les couleurs. Au début du XIXe siècle, l'inventeur français Joseph Nicéphore Niépce a pris la première photographie de l'histoire en utilisant la technique de l'héliographie à la chambre obscure. Il s'agit de la première tentative réussie dans l'histoire de l'humanité d'utiliser des procédés chimiques pour capturer et conserver des images visuelles, ce qui constitue la base des techniques photographiques ultérieures.

Actuellement, l'essor des médias numériques offre davantage de possibilités d'enregistrer le monde réel et de partager l'inspiration créative. Mais les photographies que nous voyons aujourd'hui sont-elles suffisantes pour rendre compte du monde réel ? La gamme dynamique de la luminance très étendue dans les environnements naturels, la luminance d'un ciel nocturne clair est d'environ $10^{-3}\ cd/m^2$, tandis que la luminance

d'une scène éclairée par le soleil est d'environ 10^5 cd/m^2 et la luminance du soleil lui même est supérieure à 10^9 cd/m^2 (voir la figure 1.1). Dans une scène d'intérieur classique, le ratio de la luminance entre les parties les plus claires et les plus sombres est d'environ $1 : 10^3$, et dans des scènes particulières, comme un lever de soleil, ce ratio de gamme dynamique peut atteindre environ $1 : 10^5$. Cependant, les images numériques classiques sont généralement codées sur 24 bits par pixel (3 canaux de couleur × 8 bits), un seul canal stockant des valeurs entières comprises entre 0 à 255, ce qui permet de stocker une gamme dynamique limitée et ne permet pas de capturer simultanément les détails clairs et sombres.

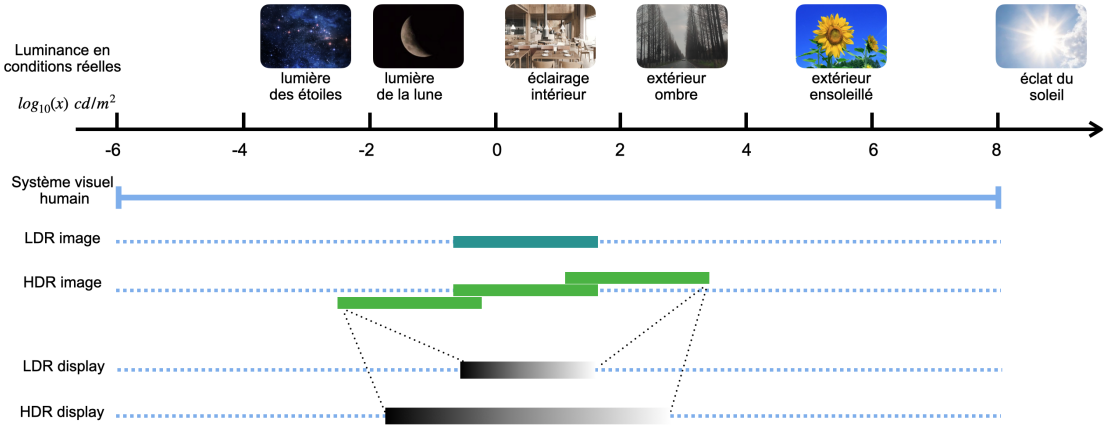


FIGURE 1.1 – Luminance en condition réelle.

Avec le développement de l'imagerie à grande gamme dynamique (HDR), nous pouvons capturer et afficher une gamme dynamique beaucoup plus large que les images classiques. Les images HDR sont codées sur 96 bits par pixels (3 canaux de couleur × 32 bits), ce qui permet aux images de mieux refléter les conditions d'éclairage réelles, créant ainsi des effets visuels plus vifs et plus réalistes. Elle est généralement créée en combinant plusieurs photos avec différents niveaux d'exposition, chacune capturant un niveau de luminance différent dans la scène (Debevec et al., 1997). Les informations dans ces photos sont ensuite composées en une seule image qui contient tous les détails, comme le montre la figure 1.1. La gamme dynamique des images HDR est plus large que celle des images à faible gamme dynamique (LDR).

Après avoir enregistré ces données à grande dynamique, elles doivent être présentées avec précision sur un dispositif d'affichage. Pour l'écran conventionnel, le pic de luminance se situe généralement entre 200 et 500 cd/m^2, et supporte une profondeur des couleurs de 8 bits. Selon la certification de DisplayHDR de VESA (« Vesa Certified

DisplayHDR », s. d.), pour les écrans HDR existants au niveau de l'application sur le marché, le pic de luminance peut atteindre jusqu'à 1 400 cd/m^2, et supporte une profondeur des couleurs plus large (au minimum 10 bits). Mais ils sont cependant loin de couvrir une grande partie de la gamme dynamique stockée dans les images HDR. Par conséquent, les informations HDR doivent être compressées à l'aide de techniques de Tone Mapping (TMO) afin qu'elles puissent être rendues sur les dispositifs d'affichage existants.

L'objectif d'un TMO est de réduire la gamme dynamique d'une image tout en conservant son impression générale. Cependant, le processus de réduction de la gamme dynamique s'accompagne généralement d'une perte de style de l'image. Comme le montre la figure 1.2, l'opération de réduction d'échelle modifie la distribution de l'histogramme des pixels de l'image, ce qui rend le style de l'image finale affichée différent du style original de l'image HDR. Les photographes professionnels appliquent souvent une correction d'exposition ou un ajustement de la courbe pour préserver les détails des ombres et des lumières d'une scène. Au cours du processus d'ajustement, les informations esthétiques telles que la composition de l'image, la distribution de la lumière et le style de tonalité doivent être prises en considération, et un affinement des valeurs de l'image doit être effectué afin d'obtenir une perception visuelle cohérente entre le dispositif d'affichage et la scène réelle.

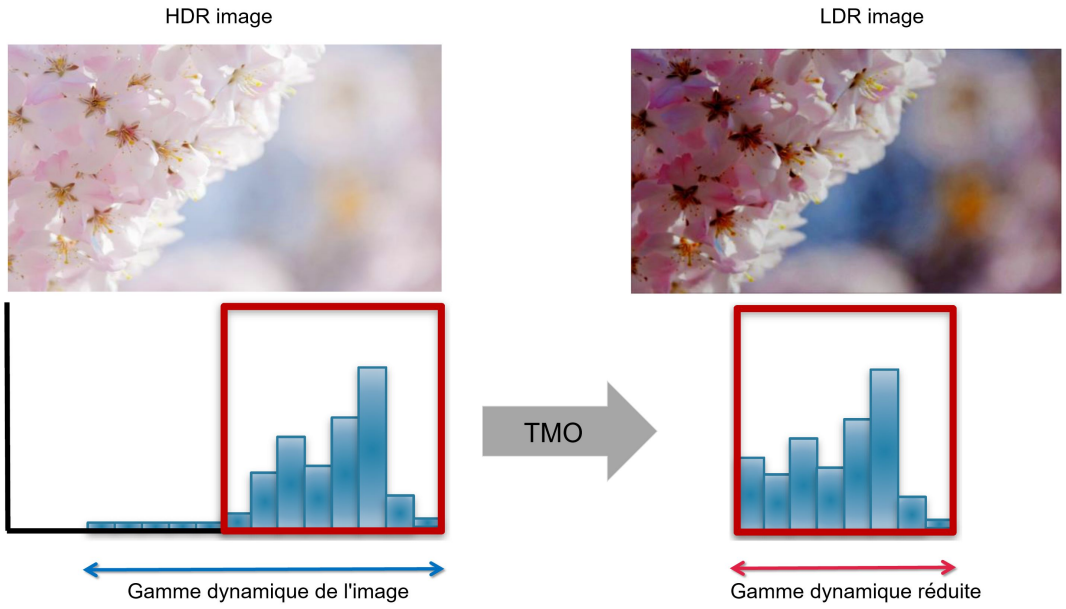


FIGURE 1.2 – Changement de style d'image de HDR à LDR.

1.1 Contexte

Les divergences inhérentes entre les capacités d'enregistrement des images et celles d'affichage des écrans induisent une représentation partielle des informations des caractéristiques esthétiques du monde réel sur nos dispositifs numériques. En effet, la perception complète de ces informations esthétiques est limitée, car elle ne reflète pas intégralement la réalité. De surcroît, considérant la subjectivité intrinsèque du concept esthétique, l'analyse, la compréhension et la restitution optimale des nuances esthétiques dans les images pour une reproduction fidèle sur les écrans constituent un domaine de recherche et d'exploration complexe.

Grâce à une meilleure capacité à capturer et à reproduire les informations lumineuses, les images HDR offrent des détails plus riches et plus vifs pour l'affichage numérique des images. Comme montré dans la figure 1.1, les images HDR peuvent être synthétisées à partir de LDR avec différentes valeurs d'exposition, qui contiennent plus d'informations qu'une seule image LDR. Cependant, la gamme dynamique de luminance contenue dans les images HDR dépasse souvent les capacités d'affichage des dispositifs d'affichage existants (y compris les écrans HDR). L'image que nous voyons sur un dispositif d'affichage est une image à gamme dynamique réduite et, pour s'adapter aux capacités du dispositif d'affichage, nous devons comprimer et ajuster les parties claires et sombres de l'image, et ce processus modifiera la distribution des informations dans l'image.

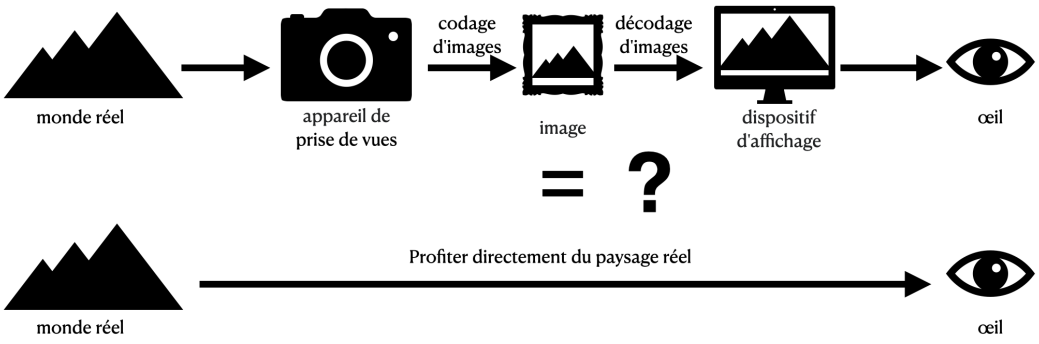


FIGURE 1.3 – Deux manières de percevoir.

Les changements de lumière, de tonalité, de contraste ou d'autres détails peuvent affecter la perception des caractéristiques esthétiques. Par conséquent, nous devons analyser les éléments qui affectent la perception de ces caractéristiques esthétiques à travers les informations de l'image, et ajuster l'image tout en maximisant la préservation des informations esthétiques, de sorte que le résultat que nous percevons à travers le

dispositif d'affichage soit aussi proche que possible de la perception de la scène réelle, comme le montre la figure 1.3.

1.2 Problématique et objectifs

Dans cette thèse, nous cherchons à transposer le concept subjectif de l'esthétique des images en un ensemble de caractéristiques objectifs. Grâce à l'ajustement de ces caractéristiques, nous pouvons restituer autant d'informations sur la scène que possible au dispositif d'affichage.

L'objectif de ce travail consiste à une analyse des informations esthétiques inhérentes aux images, dans le but d'équiper les systèmes informatiques avec la capacité de discernement nécessaire pour identifier et prédire les attributs spécifiques des images qui exercent une influence significative sur la perception esthétique. Parallèlement, elle vise à développer et à affiner des méthodologies de traitement d'image avancées pour justifier les informations de lumière dans les images HDR, destinées à optimiser et à enrichir l'expérience visuelle.

1.3 Organisation du manuscrit

En dehors de l'introduction et la conclusion, cette thèse est organisée en quatre chapitres principaux présentant les travaux de recherche :

Le chapitre 2, Imagerie à grande gamme dynamique, présente les notions associées aux images à grande gamme dynamique, notamment la couleur, la lumière, le système visuel humain, etc. Ce chapitre détaille l'acquisition d'images à grande gamme dynamique, les différentes méthodes de Tone Mapping proposées pour s'adapter au dispositif d'affichage et les critères d'évaluation des méthodes de Tone Mapping.

Le chapitre 3, Traitement d'images HDR, est consacré à deux méthodes permettant d'améliorer l'affichage des images HDR. La section 3.2 explique la méthode de prédiction de la courbe d'adaptation tonale à partir de l'histogramme d'une image, et la section 3.3 propose un ajustement automatique de l'exposition par la construction d'un modèle de réseau de neurones.

Le chapitre 4, Esthétique de l'image, donne un aperçu des concepts fondamentaux de l'esthétique. Il présente dans un premier temps les méthodes existantes d'évaluation et d'analyse de la qualité esthétique et illustre les jeux de données pouvant être utilisés

dans la recherche sur la qualité esthétique de l'image. Ensuite, il est fait de même pour l'analyse des caractéristiques esthétiques des images.

Le chapitre 5, Reconstruction de la composition d'une image : calcul des lignes de force, propose une méthode d'analyse des lignes de force implicites d'une image, qui est un outil important pour comprendre la composition et analyser les qualités esthétiques d'une image. Ce chapitre démontre la faisabilité de l'extraction des lignes de force implicites et propose un algorithme d'extraction correspondant.

Chapitre 2

Imagerie à grande gamme dynamique

Introduction

Dans ce chapitre, nous présenterons les concepts et les méthodes liés aux images à grande gamme dynamique. La lumière et la couleur étant des éléments essentiels de la perception visuelle, nous présentons tout d'abord leurs propriétés naturelles et leur métrique physique. Nous décrivons ensuite le système visuel humain en tant que composante principale de la perception visuelle. Sur cette base, nous expliquons ce qu'est une image à grande gamme dynamique et comment nous l'acquérons, la stockons et l'affichons.

2.1 Lumière

La lumière possède une nature à la fois ondulatoire et corpusculaire. En tant qu'onde, la lumière présente des interférences et des diffractions. En tant que corpuscule, la lumière peut être considérée comme une série de photons, chacun d'entre eux transportant une certaine quantité d'énergie. La vitesse de la lumière dans le vide est d'environ $3\times10^8 m/s$.

Dans l'Antiquité, la compréhension de la lumière par l'homme était très primitive et intuitive. Le philosophe grec Euclide croyait que la vision était le résultat de l'interaction de certains rayons émis par l'œil avec un objet, et comprenait les propriétés fondamentales de la lumière grâce à l'optique géométrique. Au XVIIe siècle, Isaac Newton a proposé la théorie des corpusculaires de la lumière, qui suggérait que la lumière était composée de minuscules corpusculaire. À la même époque, Christian Huygens a proposé la théorie des ondulatoires, qui a ensuite été développée par Augustin Fresnel, expliquant les phénomènes de diffraction et d'interférence de la lumière. Au XIXe siècle, James Clerk Maxwell a réuni l'électricité et le magnétisme grâce à sa série d'équations pour le champ électromagnétique, prouvant que la lumière est un type d'onde électromagnétique.

Au début du XXe siècle, le développement de la théorie quantique a permis de mieux comprendre la dualité onde-corpuscule de la lumière.

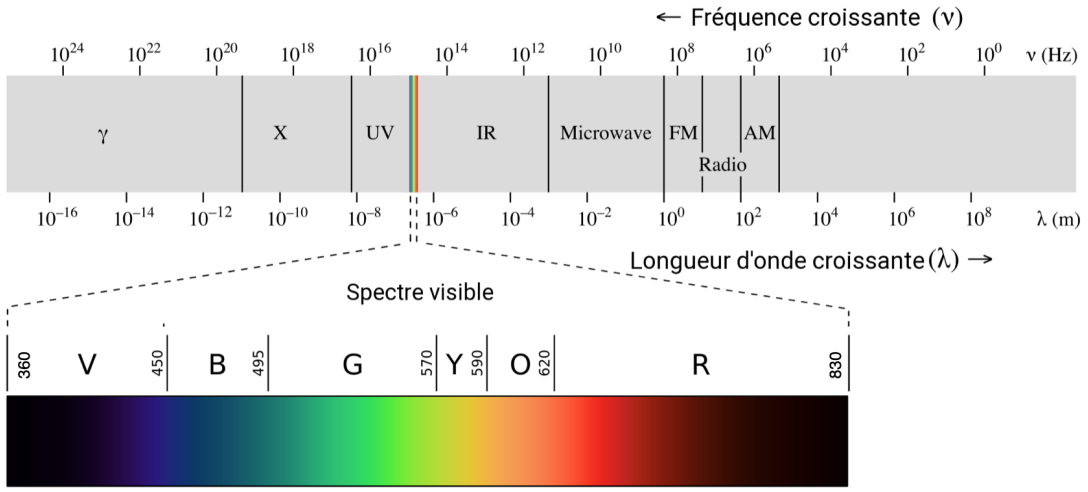


FIGURE 2.1 – Schéma spectral. Figure originale de Gringer (2018), le seuil supérieur et le seuil inférieur pour les longueurs d'onde visibles ont été modifiés conformément aux dernières instructions de la CIE (« CIE-17-21-003 », s. d.).

La distribution de la lumière décomposée en fonction de la longueur d'onde ou de la fréquence constitue un spectre. Le spectre couvre une gamme très large, des rayons gamma aux courtes longueurs d'onde jusqu'aux ondes radio aux grandes longueurs d'onde (voir la figure 2.1). La lumière visible n'est qu'une petite partie de l'ensemble du spectre, dans la gamme de longueurs d'onde d'environ 360 à 830 *nm* (la gamme spectrale de la lumière visible ne présente pas de frontières strictement définies). Cette indétermination résulte de plusieurs facteurs, notamment de l'intensité du flux lumineux parvenant à la rétine et de la sensibilité individuelle de l'observateur. En termes généraux, la limite inférieure du spectre visible est souvent estimée dans une plage allant de 360 à 400 *nm*, tandis que la limite supérieure se situe habituellement entre 760 et 830 *nm* (« CIE-17-21-003 », s. d.)), et elle peut être perçue par les cellules sensibles à la lumière de la rétine de l'œil humain.

2.1.1 Grandeurs photométriques

La figure 2.2 présente les concepts de base de la photométrie et leurs relations, des mesures qui peuvent nous aider à quantifier la lumière.

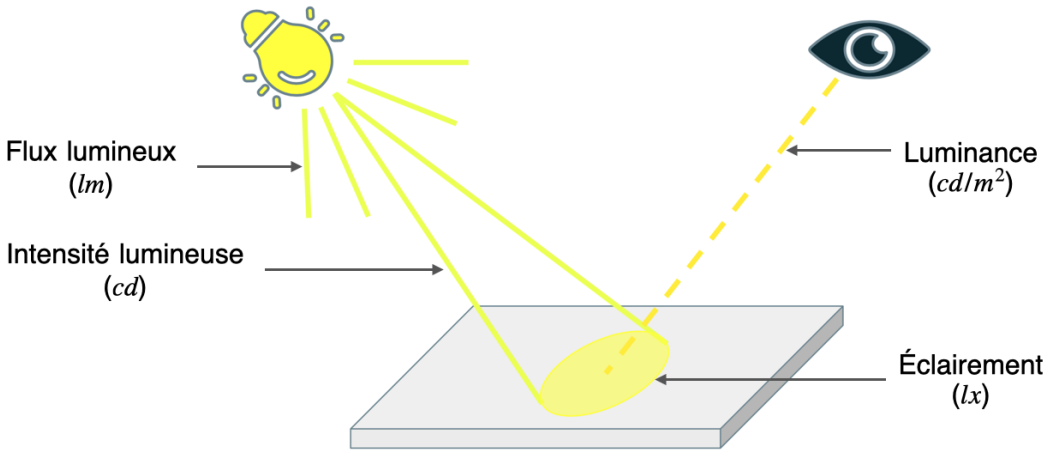


FIGURE 2.2 – Schéma de la relation entre les mesures photométriques.

Le flux lumineux Φ_V est utilisé pour mesurer la variation de l'énergie lumineuse au cours du temps :

$$\Phi_V = \frac{dQ_V}{dt} \tag{2.1}$$

où Q_V est l'énergie lumineuse émise, transférée ou reçue, et t est le temps (« CIE-e-ILV », s. d.). L'unité de Φ_V est lumen (lm). Une ampoule à incandescence traditionnelle de 60 watts produit généralement environ 800 lumens, mais ces mêmes 800 lumens peuvent être émis par une lampe LED d'environ 8 à 12 watts.

L'intensité lumineuse I_V est la densité du flux lumineux par rapport à l'angle solide dans une direction donnée :

$$I_V = \frac{d\Phi_V}{d\Omega} \tag{2.2}$$

où Φ_V est le flux lumineux émis dans une direction donnée et Ω est l'angle solide contenant cette direction (« CIE-e-ILV », s. d.). L'unité de I_V est Candela (cd). Elle est généralement utilisée pour quantifier la quantité de lumière fournie par une source lumineuse directionnelle et est indépendante de la distance d'observation. La candela est l'une des sept unités de base du Système International d'unités (SI), une candela est définie comme le flux lumineux émis par une source ponctuelle uniformément éclairée dans une direction donnée par unité d'angle solide (un degré sphérique) de $1/683\ Watt$ (à une fréquence de $540 \times 10^{12} Hz$) (Newell et al., 2019). Cette fréquence correspond à la longueur d'onde de la lumière verte ($555\ nm$), ce qui est très proche de la longueur

d'onde à laquelle l'œil humain est le plus sensible.

La luminance L_V est la densité de l'intensité lumineuse par rapport à la surface projetée dans une direction déterminée en un point déterminé d'une surface réelle ou imaginaire :

$$L_V = \frac{dI_V}{dA}\frac{1}{cos\alpha} \tag{2.3}$$

où A est la surface sur laquelle le flux lumineux est incident et α l'angle entre la normale à la surface au point spécifié et la direction spécifiée (« CIE-e-ILV », s. d.). L'unité de L_V est la candela par m^2 (cd/m^2). L'ancien nom de cette unité est le nit ($1\ nit = 1\ cd/m^2$), les moniteurs de bureau ont généralement une luminance entre 250 et 350 cd/m^2, tandis que les moniteurs certifiés par « Vesa Certified DisplayHDR » (s. d.) atteignent un pic de luminance de 1 400 cd/m^2.

L'éclairement E_V est la densité du flux lumineux incident par rapport à l'aire en un point d'une surface réelle ou imaginaire :

$$E_V = \frac{d\Phi_V}{dA} \tag{2.4}$$

L'unité de E_V est lux (lx). L'éclairage s'intéresse à la manière dont la lumière atteint un objet, tandis que la luminance s'intéresse à l'intensité lumineuse de la surface d'un objet. L'éclairage est quantifiable, alors que la luminance se concentre davantage sur la perception de l'intensité de la lumière par l'œil humain.

2.2 Couleur

Au cours des premières explorations de la couleur, Aristote pensait que la couleur naissait de l'interaction entre la lumière et l'obscurité, et que la couleur n'était pas une propriété objective d'un objet, mais plutôt un sentiment subjectif de l'homme. Au XVIIe siècle, Newton a confirmé par une série d'expériences sur les prismes que la lumière blanche n'était pas homogène, mais qu'elle était composée de rayons de différentes couleurs. Au XIXe siècle, la théorie trichromatique a été proposée en se fondant sur les recherches de Thomas Young, suggérant que l'œil contient trois récepteurs sensibles à la lumière rouge, verte et bleue. Cette théorie a jeté les bases du modèle moderne de la couleur.

Dans la théorie trichromatique, nous pouvons obtenir toutes les couleurs de lumière en mélangeant la lumière monochromatique rouge, verte et bleue, ce qui constitue la base

du principe de rendu des couleurs d'un écran. Les fonctions de bases trichromatiques $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$ présentées dans la figure 2.3 indiquent la quantité de lumière monochromatique nécessaire pour faire correspondre une unité de puissance à chaque longueur d'onde. Dans cette figure, les trois primaires monochromatiques aux longueurs d'onde de $700,0\ nm$ (rouge), $546,1\ nm$ (vert) et $435,8\ nm$ (bleu). Si la distribution spectrale de puissance d'une couleur est C_λ, les équations généralisées pour le calcul des valeurs tristimulus RGB sont les suivantes :

$$R = \int_\lambda C_\lambda\ \overline{r}(\lambda) d\lambda \tag{2.5}$$

$$G = \int_\lambda C_\lambda\ \overline{g}(\lambda) d\lambda \tag{2.6}$$

$$B = \int_\lambda C_\lambda\ \overline{b}(\lambda) d\lambda \tag{2.7}$$

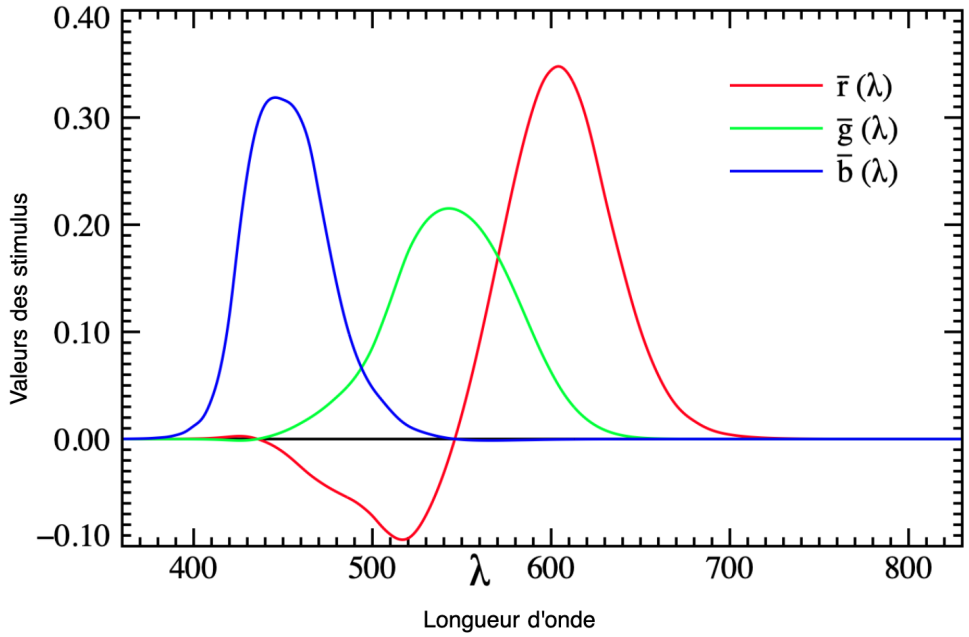


FIGURE 2.3 – Fonctions colorimétriques des couleurs de CIE RGB (« RGBCMF », s. d.).

Pour obtenir ces trois fonctions colorimétriques, les chercheurs (Smith et al., 1931 ; Wright, 1929) ont effectué une série d'expériences. Dans l'expérience, un écran est divisé en deux zones par une cloison opaque, et le côté gauche est éclairé par une lumière

monochromatique d'une longueur d'onde donnée, puis l'intensité des trois sources lumineuses colorées du côté droit est ajustée jusqu'à ce que les couleurs des côtés gauche et droit aient la même apparence. Cependant, toutes les couleurs ne peuvent pas être appariées directement. Lorsque cela n'est pas possible, il est permis d'ajouter une couleur primaire à la couleur testée et de mélanger les deux autres couleurs primaires pour qu'elles s'accordent. Dans ce cas, le nombre de couleurs primaires ajoutées à la couleur test est considéré comme négatif, comme le montre la figure 2.3.

Dans la fonction colorimétriques des couleurs RGB, il y a des valeurs négatives, ce qui rend impossible la caractérisation des couleurs perceptibles par l'œil humain dans la pratique. Pour résoudre ces problèmes, la CIE a développé l'espace colorimétrique XYZ en 1931. Cet espace a été conçu pour couvrir toutes les couleurs perceptibles par l'œil humain. La conversion de l'espace RGB à l'espace XYZ implique une transformation mathématique, où les coordonnées de couleur dans l'espace RGB sont converties en coordonnées XYZ (M. D. Fairchild, 2005). $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$ sont respectivement les fonctions colorimétriques de X, Y et Z, comme le montre la figure 2.4. Alors dans ce nouvel espace de couleur :

$$X = \int_{\lambda} C_{\lambda}\, \overline{x}(\lambda) d\lambda \tag{2.8}$$

$$Y = \int_{\lambda} C_{\lambda}\, \overline{y}(\lambda) d\lambda \tag{2.9}$$

$$Z = \int_{\lambda} C_{\lambda}\, \overline{z}(\lambda) d\lambda \tag{2.10}$$

Dans l'espace XYZ, la valeur du stimulus est positive pour toute couleur visible. Cependant, il n'existe pas de dispositif physique réaliste capable de réaliser les couleurs primaires de la CIE XYZ, de sorte que cette couleur primaire est également appelée « couleur primaire imaginaire » (Reinhard, 2010). Afin de faciliter la représentation des coordonnées 2D de la couleur, nous projetons les points de données dans l'espace 3D XYZ en perspective sur le plan unitaire de cet espace, comme indiqué dans les équations 2.112.122.13.

$$x = \frac{X}{X + Y + Z} \tag{2.11}$$

$$y = \frac{Y}{X + Y + Z} \tag{2.12}$$

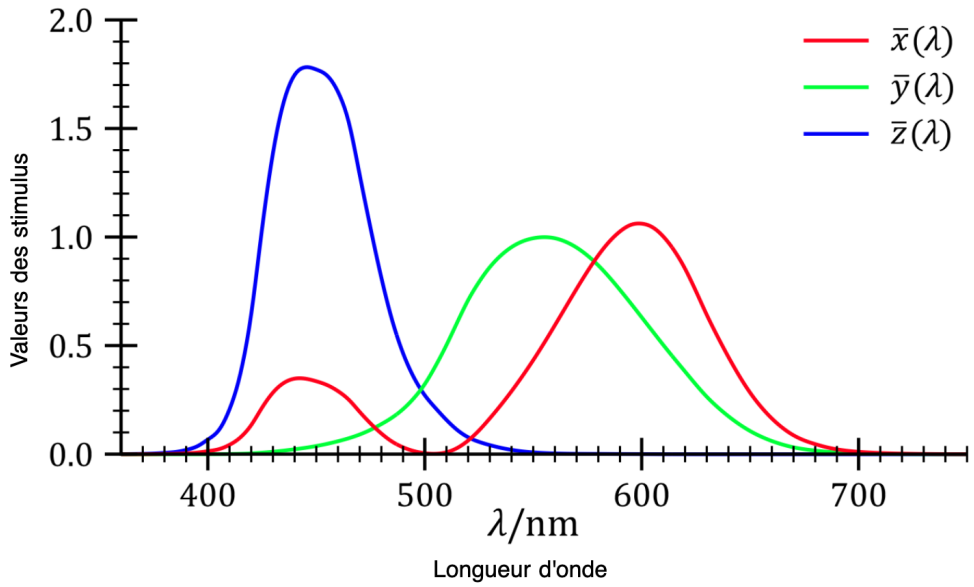


FIGURE 2.4 – Fonctions colorimétriques des couleurs de CIE XYZ (« XYZ », 2009).

$$z = \frac{Z}{X+Y+Z} = 1 - x - y \tag{2.13}$$

La figure 2.5 montre un diagramme de chromaticité avec deux axes tracés à partir des coordonnées de chromaticité. L'ensemble de la zone entourée par la courbe en fer à cheval est appelée gamut de l'espace colorimétrique XYZ, qui comprend toutes les couleurs visibles. Les contours de la courbe sont appelées « locus spectral », avec les nombres violets correspondent aux longueurs d'onde de la lumière monochromatique. La ligne droite au bas de la gamut est appelée « la droite des pourpres ». Il s'agit des couleurs qui, bien que situées à la limite de la gamut, n'ont pas de lumière monochromatique correspondante.

Espace couleur	**Couverture de la CIE 1931**	**Application**
Rec.709	≈ 35,9%	TVHD, Blu-ray
Rec.2020	≈ 75,8%	4K/8K UTVHD
Adobe RGB	≈ 52,1%	Photographie numérique
DCI-P3	≈ 53,6%	Cinéma numérique
ProPhoto RGB	≈ 90%	Photographie haut de gamut

TABLEAU 2.1 – Comparaison de différents espaces colorimétriques.

Un espace de couleur est un modèle mathématique de représentation, de création ou

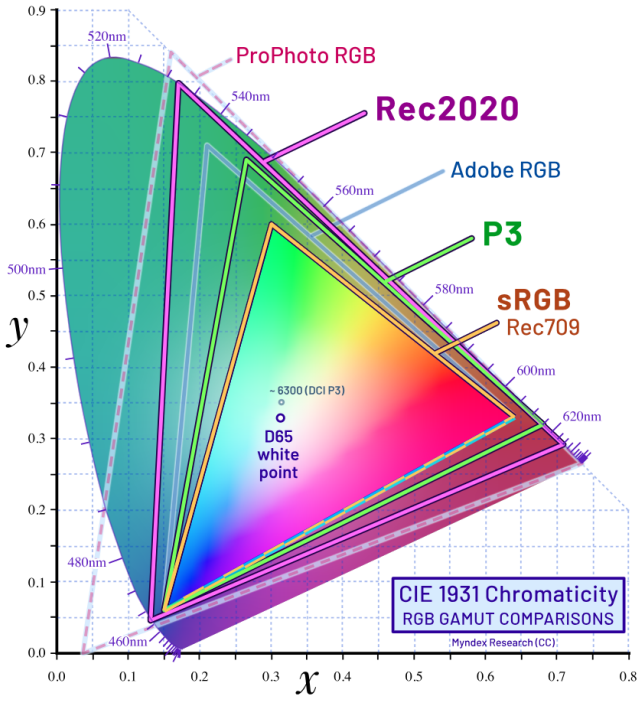


FIGURE 2.5 – Diagramme de chromaticité. Les zones triangulaires de différentes couleurs représentent la gamme de lumière visible qui peut être couverte par différents espaces de couleur (Myndex, 2022).

d'interprétation des couleurs. En fonction des caractéristiques physiques des dispositifs de codage et de décodage des couleurs basés sur différents besoins, les gamuts de couleurs des différents espaces colorimétriques varient (voir la figure 2.5). Actuellement, l'espace de couleur le plus couramment utilisé est l'espace sRGB, basé sur le modèle de couleurs RGB, qui est également couramment utilisé dans l'imagerie HDR. L'espace colorimétrique XYZ est principalement utilisé pour calculer le canal Y en HDR, car il est proche du modèle de la luminance perçue. En ce qui concerne les dispositifs d'affichage, les moniteurs HDR utilisent généralement l'espace Rec. 2020, tandis que les moniteurs standard sont d'utiliser Rec. 709 (voir le tableau 2.1). Il existe d'autres espaces colorimétriques spécialement conçus pour l'imagerie HDR, tels que hdr-CIELAB et hdr-IPT (M. Fairchild et al., 2011), mais ils ne sont pas aussi largement utilisés que les autres modèles existants.

2.3 Perception visuelle

Après avoir découvert les propriétés naturelles de la lumière et de la couleur, nous introduirons dans cette section le système visuel humain (human visual system - HVS) afin de comprendre comment les humains perçoivent la lumière et la couleur. L'HVS est un ensemble complexe et sophistiqué de structures physiologiques et nous nous concentrons ici sur la rétine et le cortex visuel.

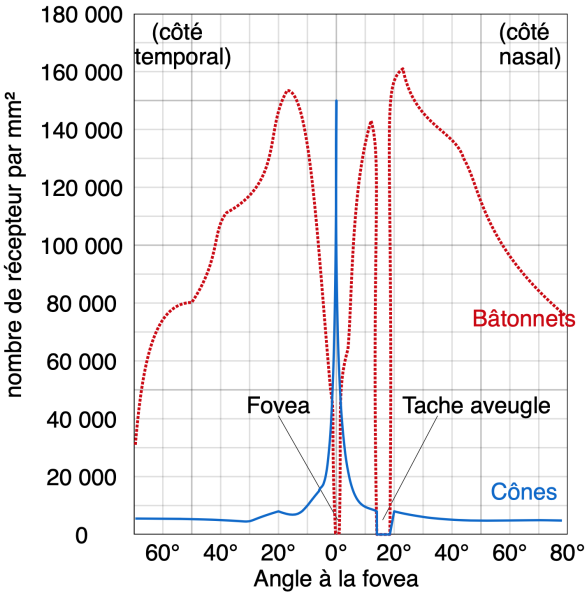


FIGURE 2.6 – Distribution des photorécepteurs de l'œil gauche humain. Les cônes se trouvent principalement dans la fovéa. La densité des bâtonnets se situant principalement entre 10 et 20 degrés. Il n'y a pas de photorécepteurs dans la tache aveugle. Image de (Cmglee, s. d.).

Après avoir traversé la cornée, la lumière pénètre dans l'œil via la pupille. Elle est focalisée sur la rétine qui transforme les signaux lumineux en signaux neuronaux. Ces derniers sont ensuite traités par le cortex visuel du cerveau, résultant en la perception visuelle que nous expérimentons. La rétine est une fine couche de tissu nerveux située à l'arrière de l'œil. Il existe deux types de photorécepteurs dans la rétine : les bâtonnets et les cônes, qui sont responsables de la perception de la lumière dans différentes gammes d'intensité. Chaque œil compte environ 5 millions de cônes et 100 millions de bâtonnets (Wandell et al., 1995). Les cônes sont concentrés dans la fovéa [1]. Par contre, les bâtonnets ne sont pas situés dans la fovéa, mais sont répartis autour de la fovéa, la distribution est

1. Zone centrale de la tache jaune de la rétine où la vision est la plus nette (« Le Robert Illustré », 2018).

illustrée à la figure 2.6.

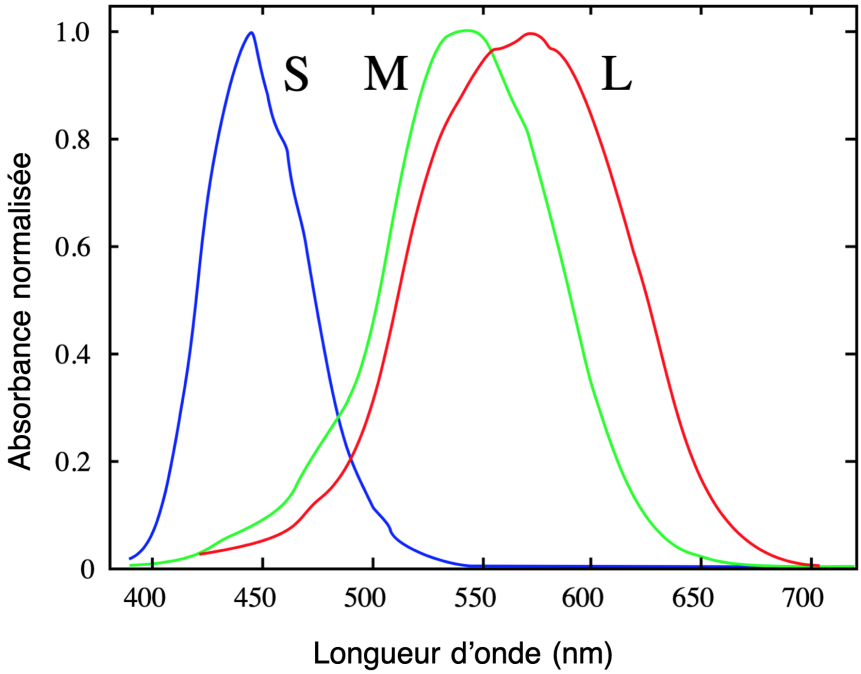


FIGURE 2.7 – Sensibilités spectrales des cônes L, M et S. Image rééditée sur la base de (Cmglee, s. d.).

Les cônes sont divisés en trois types, comme le montre la figure 2.7. Les trois types sont plus sensibles aux ondes lumineuses de différentes longueurs d'onde, longues, moyennes et courtes, et les sensibilités sont représentées par des courbes rouges, vertes et bleues dans la figure. Ces trois types de cônes permettent au système visuel humain de percevoir les couleurs. En revanche, les bâtonnets agissent principalement dans la vision scotopique [2], qui permet de percevoir les changements d'échelle de gris, mais pas les couleurs. Dans la pénombre, les bâtonnets deviennent plus actives pour s'adapter aux environnements peu éclairés, alors que dans les environnements lumineux, la perception des couleurs par les cônes prédomine (M. D. Fairchild, 2005).

Les stimuli visuels provenant de la rétine sont perçus et interprétés par le cortex visuel dans le cerveau. Le cortex visuel est principalement divisé en cortex visuel primaire (zone V1) et en zones de traitement visuel de niveau supérieur. Les informations visuelles atteignent d'abord la zone V1, qui est très sensible aux caractéristiques visuelles de base

2. La vision scotopique correspond à la capacité de l'œil à voir dans des conditions de faible luminosité, voire dans une obscurité partielle.

telles que l'intensité lumineuse, l'orientation des contours et la direction du mouvement. À partir de la zone V1, les informations visuelles sont transmises aux zones de traitement visuel de niveau supérieur, où elles continuent de traiter des caractéristiques visuelles plus complexes, telles que la forme, la couleur et le mouvement.

L'étude de la psychophysique montre que la perception de la lumière par l'œil humain n'est pas linéaire. L'HVS est sensible aux contrastes, mais insensible aux différences de luminance absolues. Comme le décrit la loi de Weber-Fechner (Fechner, 1966), la relation entre l'intensité du stimulus et l'intensité perçue n'est pas linéaire. Si la magnitude d'un stimulus physique est X, et la perception est S, la relation entre eux est logarithmique $S = lnX$. Cependant, cette relation logarithmique ne s'applique pas à tous les scénarios, et dans la recherche de Stevens (Stevens, 1970), la relation entre un stimulus et sa perception est décrite comme une fonction de puissance $S = \alpha X^{\beta}$ plutôt que logarithmique. L'exposant β dépend du type de stimulus ou de la modalité sensorielle, et la constante α est une constante de proportionnalité. Il est donc nécessaire, dans les recherches sur l'image, de tenir compte de cette perception non linéaire.

2.4 Conversion de l'information optique en données d'image

Dans le processus d'imagerie d'un appareil photo numérique, la lumière incidente passe par le système d'objectif de l'appareil photo et entre dans le capteur d'image. Chaque photosite du capteur convertit le signal lumineux en signal électrique par conversion photoélectrique. Dans ce processus, la réponse du photosite à la lumière n'est pas entièrement linéaire, en particulier dans des conditions de lumière très forte ou très faible. Le signal électrique est ensuite converti en signal numérique par des procédés analogiques, d'amplification, etc. et enregistré sous forme de données RAW. Les données RAW sont ensuite traitées et compressées dans d'autres formats d'image plus courants tels que jpg, png, tiff et hdr.

Les données d'image LDR traditionnelles sont codées sur 8 bits, c'est-à-dire que chaque pixel contient 3 canaux et que chaque canal est représenté par un nombre entier de 8 bits (0 - 255), ce qui signifie que les images LDR ne peuvent généralement pas représenter les informations de la grande gamme dynamique. Les images HDR sont généralement codées sous forme de nombres à virgule flottante à précision unique de 16 bits ou à précision double de 32 bits, telle que hdr, tiff et exr. Comme montré dans le tableau 2.2, une plus grande précision d'encodage permet de stocker une gamme dynamique plus large.

Profondeur de Bits	**Niveau de Luminosité ($2^{\text{bit depth}}$)**
8 bits	2^8 = 256 niveaux
10 bits	2^{10} = 1,024 niveaux
16 bits	2^{16} = 65,536 niveaux
32 bits	2^{32} = 4,294,967,296 niveaux

TABLEAU 2.2 – Correspondance entre la profondeur de bits et le niveau de luminosité.

2.5 Méthode d'acquisition de l'imagerie HDR

En considérant le ratio entre la valeur du pixel le plus clair et la valeur du pixel le plus sombre comme une approximation de la gamme dynamique d'une image, la gamme dynamique obtenue par les techniques d'imagerie conventionnelles est généralement très limitée et ne suffit pas à représenter de nombreuses scènes de la vie réelle. Dans les scènes comportant les zones sombres et claires, une seule exposition n'est pas suffisante pour capturer toutes les informations sombres et claires en même temps, en raison des limites physiques du capteur d'image. Les technologies de génération d'images et d'affichage liées au HDR permettent le stockage et l'affichage d'images à gamme dynamique élevée, ce qui rapproche l'expérience visuelle offerte par les appareils électroniques de la scène réelle.

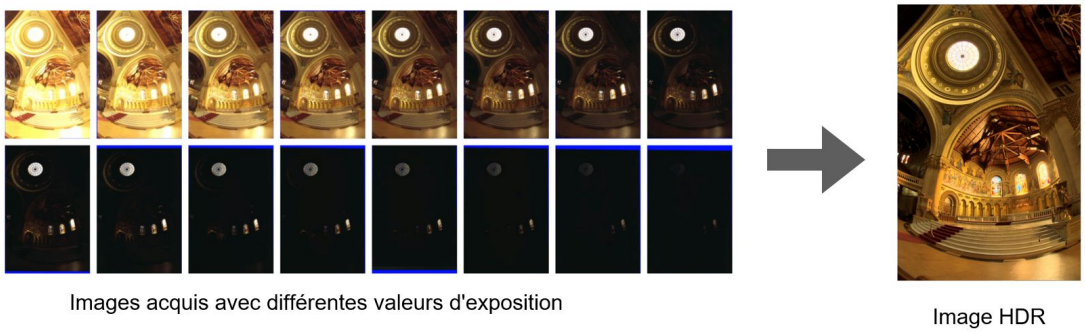


FIGURE 2.8 – Fusion de multi-expositions pour produire une image HDR.

L'acquisition d'images HDR se fait généralement de quatre manières : fusion de multi-expositions, extension de l'exposition unique, senseurs HDR et synthèse d'images. Comme indiqué à la section 2.4, la transformation de la scène réelle en informations d'image implique différents traitements non linéaires, mais nous considérons généralement la transformation globale directement, en supposant que le dispositif de capture est complètement linéaire. Par conséquent, la valeur d'éclairement énergétique peut être récupérée en déduisant la durée d'exposition de la valeur radiance enregistrée (Reinhard,

2010). Dans la fusion multi-exposition, plusieurs images de la même scène sont prises à différentes valeurs d'exposition pour capturer tous les détails, des zones les plus sombres aux plus lumineuses (voir la figure 2.8). La valeur d'éclairement énergétique peut être dérivée en intégrant les calculs pour produire une image HDR :

$$E(i,j) = \frac{\sum_{k=1}^{N_{ev}} w(I_k(i,j)) \frac{I_k(i,j)}{\Delta t_i}}{\sum_{k=1}^{N_{ev}} w(I_k(i,j))} \tag{2.14}$$

où $E(i,j)$ est l'éclairement énergétique à l'emplacement (i,j) du pixel, N_{ev} est le nombre d'images à différentes expositions, I_k est la valeur d'enregistrement de l'image à la k-ième exposition, Δt_i est le temps d'exposition pour I_k, $w(I_k(i,j))$ est une fonction de pondération qui élimine les valeurs extrêmes (Banterle et al., 2018 ; Debevec et al., 1997 ; Reinhard, 2010).

Cependant, le dispositif de capture d'image habituel n'est pas linéaire. Il s'agit généralement d'une courbe de réponse. Nous appelons cette fonction : courbe de réponse de la caméra (CRF). En conséquence, l'équation 2.14 peut être optimisée comme suit :

$$E(i,j) = \frac{\sum_{k=1}^{N_{ev}} w(I_k(i,j)) \frac{f^{-1}(I_k(i,j))}{\Delta t_i}}{\sum_{k=1}^{N_{ev}} w(I_k(i,j))} \tag{2.15}$$

où f^{-1} est le CRF inverse, et l'images HDR peuvent être obtenues en estimant f^{-1} (Debevec et al., 1997 ; Robertson et al., 2000). Outre la déduction des CRF, Hristova et al. (2017) propose une méthode de création d'images HDR à partir d'image capturées avec et sans flash. Ces méthodes sont essentiellement basées sur l'analyse des détails clairs et sombres contenus dans les différentes valeurs d'exposition pour obtenir l'image HDR finale.

L'extension de l'exposition unique, également connue sous le nom de mappage de tonalité inversé, génère des images HDR à partir d'images LDR à exposition unique. L'algorithme de mappage de tonalité inverse est généralement obtenu en expansant le mappage (Banterle et al., 2007) ou en inversant la fonction de compression dans le dispositif de capture (Rempel et al., 2007). Il existe également des techniques d'apprentissage profond qui permettent d'étendre la gamme dynamique en extrayant des caractéristiques profondes de l'image LDR (Eilertsen et al., 2017 ; Santos et al., 2020).

Les capteurs d'image à grande gamme dynamique peuvent capturer des images HDR de manière efficace et directe, mais la gamme dynamique des images capturées reste encore très limitée. De plus, en raison des caractéristiques physiques du matériel, les appareils photo à gamme dynamique élevée ont tendance à avoir une sensibilité plus

faible, et leur qualité d'image dans des conditions de faible luminosité n'est pas très satisfaisante. Les images HDR générées par synthèse d'images sont principalement obtenues à l'aide de moteurs de rendu par des méthodes d'éclairage photo-réalistes. Les logiciels de modélisation numérique, tels que 3D Max, Maya ou Blender, permettent de créer des scènes. Il est possible ensuite de produire des images de ces scènes par l'utilisation d'un moteur de rendu physiquement réaliste (Arnold, Cycle, PBRT, Mitsuba ...). Ceux-ci peuvent réaliser le calcul pour produire directement une image HDR.

2.6 Tone Mapping

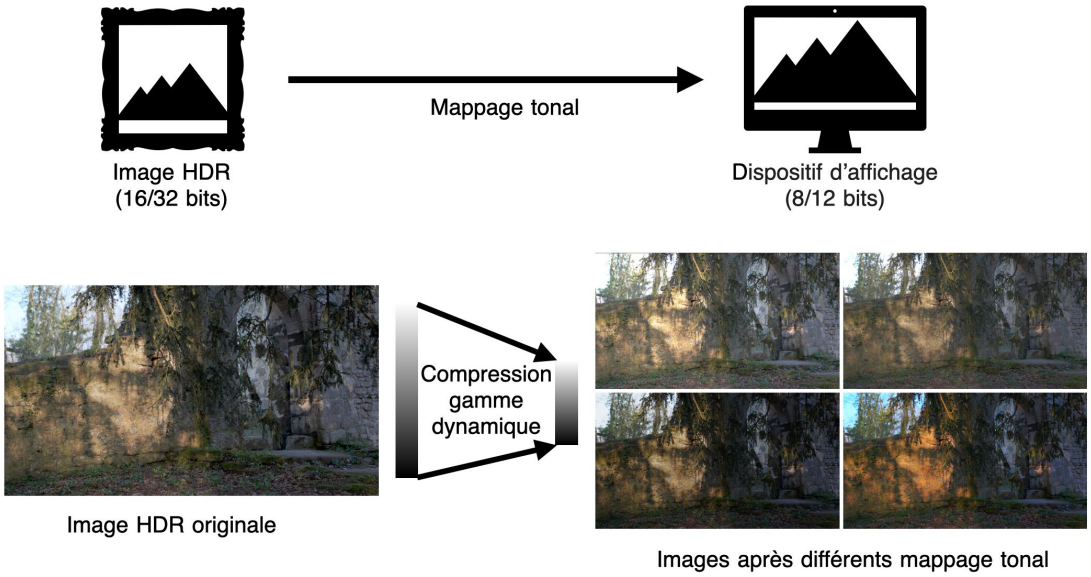


FIGURE 2.9 – Illustration du Tone Mapping. Le Tone Mapping permet aux images HDR d'être affichées plus convenablement dans les appareils dont la plage dynamique est plus limitée.

Les dispositifs d'affichage existants ne peuvent généralement afficher que des données à gamme dynamique limitée, et la gamme dynamique des données HDR dépasse les capacités des dispositifs d'affichage traditionnels. Même les écrans HDR professionnels, dont la gamme dynamique a été considérablement élargie, ne permettent pas de reproduire complètement toutes les informations de la scène des images HDR (voir la figure 2.9). Par conséquent, pour afficher des informations HDR à grande dynamique sur des appareils à dynamique limitée, nous devons appliquer un opérateur de mappage de tonalité (Tone Mapping Operator - TMO) pour comprimer la gamme dynamique. Au cours du

processus de mappage de tonalité, il est inévitable que des données soient perdues. Nous devons donc concevoir une méthode de mappage satisfaisante pour que le résultat final de l'affichage soit aussi proche que possible de la perception humaine en condition naturelle.

Les méthodes de mappage de tonalité existantes peuvent être classées en opérateurs globaux et locaux, l'opérateur global appliquant la même courbe de compression à tous les pixels de l'image, en prenant en compte uniquement les valeurs des pixels et non leur position. Par contre, les opérateurs locaux prennent en compte les informations relatives à la position des pixels et procèdent ensuite à un mappage spécifique des valeurs des pixels.

2.6.1 TMO global

Le TMO global consiste à trouver une fonction de compression basée sur les données statistiques de l'image, puis à appliquer la fonction à l'image entière pour compresser la gamma dynamique. Les fonctions utilisées couramment sont les fonctions linéaires, logarithmiques et exponentielles. Le TMO basée sur ces fonctions présente l'avantage d'être simple et rapide pour effectuer la compression de la gamma dynamique, mais elle a tendance à perdre les détails de l'image.

L'échelonnement linéaire est la méthode de compression de la gamme dynamique la plus directe.Il s'agit de trouver un facteur m qui permette de convertir la luminance de scène à haute dynamique L_w en une luminance à gamme dynamique qui puisse être affichée L_d, comme le montre l'équation 2.16 :

$$L_d = mL_w \tag{2.16}$$

Si nous utilisons directement la réciproque du maximum de luminance $max(L_w)^{-1}$ comme valeur m, l'équation 2.16 équivaut à une normalisation de l'image. En général, les valeurs de luminance élevées enregistrées dans une image HDR sont bien supérieures à la moyenne des pixels, mais elles ne représentent qu'une petite partie de tous les pixels, de sorte que les résultats de cette méthode sont généralement sombres et ne correspondent pas à la perception des scènes réelles par l'œil humain. Par conséquent, Matković et al. (1998) proposent de choisir la valeur m uniquement parmi les pixels situés dans la gamme d'exposition effective, tandis que Ward (1994) considère le seuil de visibilité dans l'affichage.

Comme présenté dans la section 2.3, l'œil humain ne perçoit pas les changements de lumière et d'obscurité de manière linéaire, de sorte que la conception de la fonction de mappage de tonalité doit également prendre en compte les lois de perception du système

visuel. Tumblin et al. (1993) sont considérés comme le premier à faire la proposition de TMO dans le domaine de l'infographie. Cette méthode intègre la relation de loi de puissance entre la luminance du stimulus et la luminance perçue proposée par Stevens pour mapper la luminance affichée à l'écran. La luminance affichée L_d est calculée comme suit :

$$L_d = L_{da}\left(\frac{L_w}{L_{wa}}\right)^{\frac{\gamma(L_{wa})}{\gamma(L_{da})}} \tag{2.17}$$

où L_{da} est la luminance adaptée du dispositif, qui dépend de la capacité d'affichage du dispositif. L_{wa} est la luminance adaptée à l'image, généralement la luminance moyenne logarithmique de l'image (Reinhard, 2010). La fonction γ est une fonction de luminance adaptée de l'œil humain basée sur la loi de Stevens :

$$\gamma(L) = \begin{cases} 2,655 & L > 100\ cd/m^2 \\ 1,855 + 0,4log_{10}(L\ +\ 2,3 \cdot 10^{-5}) & autre \end{cases} \tag{2.18}$$

La fonction logarithmique, en tant que mappage non linéaire, est également utilisée dans TMO (Stockham, 1972). Le mappage sur la base des fonctions logarithmiques peut être exprimée comme suit :

$$L_d = \frac{log(L_w\ +\ 1)}{log(L_{wa}\ +\ 1)} \tag{2.19}$$

Néanmoins, cette méthode de mappage se traduit souvent par une perte de contraste. Drago et al. (2003) proposent un TMO logarithmique adaptatif pour l'affichage de scènes à fort contraste. Afin de renforcer la compression des contrastes, la méthode utilise une fonction log_2 pour les zones sombres et une fonction log_{10} pour les zones plus claires. Pour les autres pixels, le mappage est basée sur des valeurs logarithmiques avec un biais comme indiqué dans l'équation suivante :

$$L_d = \frac{L_{da} \cdot 0,01}{log_{10}(L_{wa}\ +\ 1)} \cdot \frac{log(L_w\ +\ 1)}{log\left(2\ +\ \left(\left(\frac{L_w}{L_{wa}}\right)^{\frac{log(b)}{log(0,5)}}\right) \cdot 8\right)} \tag{2.20}$$

où le biais b est recommandé expérimentalement à $0,85$. Par ailleurs, il existe des mappages globaux qui utilisent l'ajustement de l'histogramme pour mapper la gamme dynamique, en préservant le contraste tout en obtenant des résultats qui correspondent à la perception du système visuel (Larson et al., 1997). Dans le domaine du cinéma et des

jeux vidéo, les courbes de mappage global en S-forme présentées dans la méthode de filmic (« Filmic », 2015) ont également donné de résultats satisfaisants.

Les TMO globaux, qui appliquent généralement une fonction de mappage pour comprimer la gamme dynamique, se distinguent par leur rapidité de traitement et leur facilité d'application. Toutefois, cette approche de traitement à l'échelle globale a tendance à être moins efficace pour restituer le contraste local de manière précise et peut omettre de révéler certaines informations détaillées inhérentes à la scène.

2.6.2 TMO à variation spatiale

Outre les TMO globales, d'autres modèles de TMO prennent en compte les valeurs critiques du pixel et les informations sémantiques connexes afin de préserver davantage de détails de l'image. Certaines approches proposent des méthodes de mappage local ou des algorithmes combinant des mappages globaux et locaux, et une autre partie des TMO prennent en compte des informations telles que la sémantique et le style de couleur. Ces méthodes, appelées TMO non globaux, sont discutées dans cette section.

Le concept TMO d'adaptation locale a été proposé par (Chiu et al., 1993). Considérant que le système visuel humain a des sensibilités différentes aux changements de luminance sous différentes intensités lumineuses, la méthode ajoute un filtre gaussian à chaque pixel pour mapper localement les valeurs de pixel. Cependant, l'utilisation de filtres localisés tend à provoquer un effet de halo. L'effet de halo se produit lorsque le contraste entre les zones adjacentes d'une image est élevé et que l'influence excessive entre les zones de luminosité variée provoque une inversion des couleurs dans les zones de bord, ce qui donne des bords irréalistes ou exagérés autour des objets lumineux.

Afin d'éviter l'effet de halo, Reinhard et al. (2002) proposent une méthode simulant la technique de l'esquive [3] et du brûlage [4] utilisée dans la photographie traditionnelle. Cette méthode emploie deux filtres Gaussiens avec des valeurs de pondération différentes pour chaque pixel. L'objectif est d'identifier la plus vaste zone circulaire autour de chaque pixel, dépourvue de bords à contraste élevé. Cette identification permet ensuite de mapper localement ces zones de manière à éviter la création d'effets de halo, et ainsi produire des images plus naturelles et fidèles à la réalité visuelle. Le TMO proposé par Durand et al. (2002) est également basé sur des filtres Gaussiens. Cette méthode utilise un filtre Gaussien bilatéral qui prend en compte à la fois l'information spatiale et l'information d'intensité de la luminance.

3. L'esquive consiste à filtrer les zones qui doivent être éclaircies pour les rendre plus lumineuses.
4. Le brûlage consiste à filtrer les zones qui doivent être assombries pour les rendre plus sombres.

Au lieu de prendre en compte les valeurs des pixels, la recherche de Tumblin et al. (1999) et la recherche de Fattal et al. (2002) sont basées sur la compression des gradients des pixels pour réduire la gamme dynamique, et Fattal et al. (2002) étendent la méthode pour résoudre l'équation de Poisson sur le champ de gradient afin d'obtenir des valeurs de gamme dynamique réduite. Certaines méthodes (Lauga et al., 2013 ; Yee et al., 2003) consistent à effectuer l'extraction d'informations sémantiques pour le mappage local. Étant donné que les dispositifs d'affichage peuvent avoir des caractéristiques différentes, R. Mantiuk et al. (2008) ont proposé un TMO qui ajuste le contenu de manière adaptative à un dispositif d'affichage donné. Pattanaik et al. (1998) utilisent un modèle multi-échelle de l'espace visuel pour simuler la perception visuelle subjective par le filtrage à différentes échelles. Se basant sur ce modèle, M. D. Fairchild et al. (2002) proposent un modèle d'apparence des couleurs (CAM) et ont étendu les recherches de TMO sur la base du CAM (M. Fairchild et al., 2004 ; Kuang et al., 2007 ; Reinhard et al., 2012). La figure 2.10 illustre les résultats du mappage des TMOs classiques avec des paramètres standard.

En dehors de la conversion HDR-LDR, Mertens et al. (2007) proposent une méthode de fusion de l'exposition qui permet de fusionner plusieurs images LDR avec différentes valeurs d'exposition en une seule image LDR qui se rapproche de l'effet HDR. Cette méthode considérant la saturation et le contraste de l'image, l'application de la décomposition de Laplace de l'image et de la pyramide Gaussienne de la carte de poids pour mélanger des images avec des expositions différentes peut également donner de bons résultats.

Avec le développement de l'apprentissage automatique, certaines recherches ont appliqué des méthodes d'apprentissage automatique aux modèles TMO. Les méthodes d'apprentissage profond, telles que les SVM, sont utilisées pour effectuer l'analyse des points caractéristiques de l'image (A. A. Rana et al., 2017). C. Guo et al. (2021) emploient un CNN pour extraire des caractéristiques profondes des images. DeepTMO (A. Rana et al., 2020) et TMO-Net (Panetta et al., 2021) utilisent des réseaux antagonistes génératifs (GAN) réalisent un mappage des images HDR vers les images LDR grâce à un retour d'information mutuel entre les extracteurs et les discriminateurs.

2.6.3 Évaluation de TMO

Le mappage de tonalité d'une gamme dynamique élevée à une gamme dynamique réduite est un processus de transformation compressive des informations de l'image, dans lequel la perte d'informations est inévitable. Afin de vérifier l'effectivité du modèle TMO, il est nécessaire de posséder un outil de mesure de la qualité du mappage pour trouver la similarité des deux images avant et après le mappage. Dans cette section, nous

Normalisation Drago Filmic

Durand Fattal Ashikhmin

Mantiuk Reinhard Tumblin

FIGURE 2.10 – Résultats des TMOs.

discuterons des méthodes d'évaluation de la qualité applicables.

Évaluations subjectives des TMOs

L'objectif ultime de l'affichage d'une image est d'être perçu par l'œil humain, la comparaison subjective est donc un moyen efficace de mesurer la similarité de deux images. Dans le test de comparaison subjective proposé par Yoshida et al. (2005), les participants ont d'abord été invités à sélectionner le meilleur résultat parmi un ensemble de méthodes TMO, puis à évaluer la qualité des différents résultats TMO en fonction de paramètres spécifiques tels que le naturel, le rendu des détails, etc. Il existe de multiples expériences similaires comparant les effets de différents TMOs, qui se fondent sur différents types de dispositifs d'affichage et d'analyses statistiques afin d'évaluer subjectivement les TMO existants (Cerdá et al., 2018 ; Ledda et al., 2005 ; Urbano et al., 2010).

Pour répondre aux divers besoins et s'adapter à une variété d'appareils, nous pouvons nous inspirer des logiques expérimentales subjectives existantes, en les utilisant comme fondement pour notre propre processus de conception et de vérification.

Évaluations objectives des TMOs

Les expériences subjectives, étant significativement influencées par divers facteurs d'interférence externes, nous incitent à privilégier l'emploi d'évaluation plus objectives pour renforcer la fiabilité et la précision de la recherche. Les critères d'évaluation de la qualité des images couramment utilisés sont classés en évaluation avec référence et évaluation sans référence. L'évaluation sans référence est basée sur une seule image pour donner des scores de qualité ou d'esthétique, ce type d'évaluation peut être consulté à la section 4.2, l'évaluation avec référence est utilisée pour comparer deux images, elle permet de comparer les différences de structure, de bruit, de distorsion et autres entre les deux images.

Les évaluations de référence couramment utilisées sont généralement appliquées aux images LDR non linéaires, telles que MSE, SSIM, PSNR, etc. et ne sont pas entièrement applicables aux images HDR linéaires. Le prédicteur de différence visuelle HDR (HDR-Visual Difference Predictor - HDR-VDP) (R. K. Mantiuk et al., 2023) est un critère qui prédit les différences perçues dans les deux images, il peut être utilisé pour la comparaison des différences entre les images HDR. Cette méthode tient compte de la réponse non linéaire des photorécepteurs humains à la lumière et simule les différences visuelles qui peuvent être perçues par le système visuel sur les images. Le résultat du HDR-VDP est une valeur de similarité et une carte de probabilité de détection des différences, comme le montre la figure 2.11, où la transition du rouge au bleu représente la probabilité de

détection d'une différence entre une valeur élevée et une valeur faible.

FIGURE 2.11 – Critère HDR-VDP. Dans le haut de la droite se trouve une carte des différences de probabilité détectées, les zones rouges indiquent une probabilité plus élevée de percevoir une différence visuelle, telle qu'une perte de détails dans la zone de la lumière dans le cadre rouge.

Considérant la grande différence de gamme dynamique entre l'image HDR originale et celle qui a été mappée, T. Aydin et al. (2008) proposent une méthode qui peut comparer les divergences entre des images ayant des gammes dynamiques différentes. Cette méthode montre les différences entre les deux images comparées par le biais d'une carte de distorsion, où les informations de la carte de distorsion comprennent la perte de caractéristiques visibles, l'amplification de caractéristiques invisibles et l'inversion du contraste polarité.

Une autre méthode consacrée à l'évaluation des TMO est l'indice de qualité des images de mappage de tonalité (TMQI) (Yeganeh et al., 2013). Ce critère fournit un score de qualité globale pour l'image mappé ainsi que des scores de fidélité structurelle et de naturel statistique, et génère des cartes de qualité multi-échelles reflétant les variations de la fidélité structurelle à différentes échelles et dans différents espaces.

Afin de satisfaire aux caractéristiques non linéaires et de gamme dynamique élevée des images HDR, les recherches (Aydın et al., 2008 ; R. K. Mantiuk et al., 2021) proposent les codages perceptuelles uniformes (PU) qui étendent certains critères traditionnels à

l'évaluation des images HDR. Toutes ces méthodes fournissent une base forte et objective pour l'évaluation de la qualité des images HDR et la validation de l'efficacité du modèle de mappage de tonalité.

Résumé

L'acquisition d'images HDR consiste généralement à capturer le maximum la gamme dynamique de la scène, et les méthodes utilisées sont principalement basées sur la fusion de captures ou l'expansion d'images LDR, le rendu virtuel ou l'amélioration de l'architecture matérielle. En raison des capacités d'affichage limitées des appareils existants, il est nécessaire d'utiliser le TMO pour faire convertir les images HDR à haute dynamique en images à gamme dynamique réduite. Les TMO existants visent à reproduire les informations de la scène de la manière la plus réaliste possible sur les appareils dont la gamme dynamique est limitée. La plupart des modèles TMO prennent en compte la caractéristique de haute dynamique des images HDR et intègrent la perception non linéaire de la lumière par l'œil humain afin d'ajuster l'affichage. Les modèles possèdent des courbes non linéaires globales ou des opérateurs de filtrage locaux afin de maximiser la préservation des informations de l'image.

Néanmoins, les images HDR acquises par des méthodes existantes ne représentent pas fidèlement la scène réelle, et les dispositifs d'affichage HDR existants ne peuvent pas reproduire parfaitement toutes les informations sur la couleur et la lumière, de sorte que nous devons encore améliorer l'effet de perception visuelle à l'aide d'ajustements de post-traitement. Au chapitre 3, nous proposons deux méthodes spécifiques pour l'ajustement des images HDR. Au chapitre 5, nous présentons les lignes de force de la composition de l'image qui guident la perception visuelle. La reconstruction des lignes de force est également applicable à l'analyse esthétique des images HDR.

Chapitre 3

Traitement d'images HDR

Introduction

Les logiciels d'édition d'images HDR permettent d'ajuster les valeurs HDR afin d'obtenir un affichage idéal sur les appareils HDR. Dans ce chapitre, nous présentons un jeu d'images HDR avec différents paramètres d'ajustement et nous proposons deux méthodes d'ajustement HDR. Ce jeu de données contient d'une part, des images HDR de différentes scènes et, d'autre part, les paramètres d'ajustement réalisés par un expert, qui permettent d'analyser les caractéristiques esthétiques des images HDR et de prédire automatiquement les paramètres d'ajustement. Grâce à ce jeu de données, nous proposons une méthode d'édition automatique des courbes de tonalité et une méthode de prédiction de la valeur d'ajustement de l'exposition, ces deux méthodes pouvant servir à améliorer l'affichage des images HDR sur les dispositifs d'affichage HDR.

3.1 Données d'analyses

Actuellement, les contenus HDR peuvent-être collectés à partir de différents ensembles de données accessibles au public (M. D. Fairchild, 2007 ; Froehlich et al., 2014 ; Funt et al., 2010). De plus, plusieurs sites web fournissent des images HDR de haute qualité, tels que Poly Haven [1], HDR Maps [2], etc. Une liste détaillée se trouve à l'annexe A.1. Cependant, les images HDR issues de ces sources, sont en nombre très limités, ne contiennent pas de paramètre d'édition et ne sont pas assez complètes en termes de types de scènes (paysage, paysage urbain, etc.). Nous avons donc utilisé un nouvel jeu de données d'images HDR composé de 787 images HDR et de leurs paramètres d'ajustement correspondants.

1. https ://polyhaven.com/hdris
2. https ://hdrmaps.com/

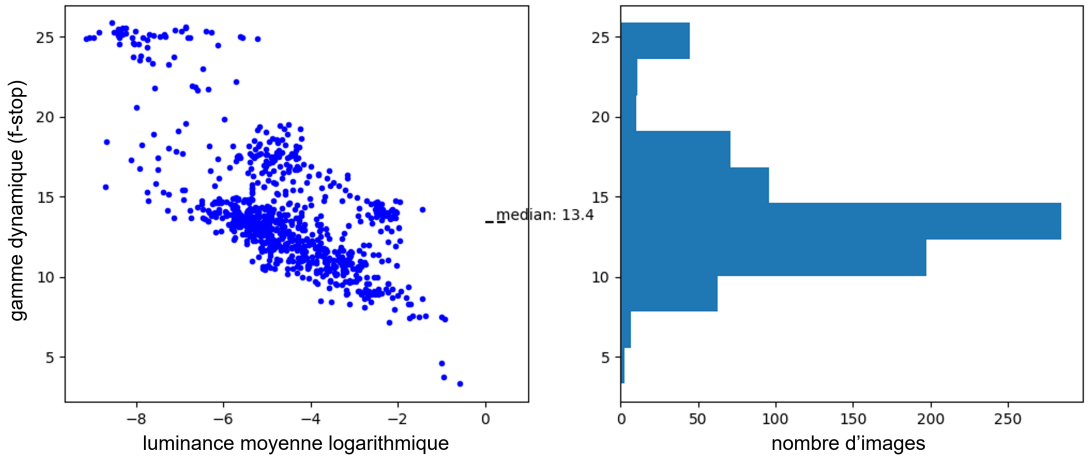


FIGURE 3.1 – Distribution de la gamme dynamique. La partie de gauche illustre la relation entre la gamme dynamique et la valeur moyenne logarithmique de la luminance . La partie de droite est la distribution de la gamme dynamique.

Toutes les images ont été réalisées par un photographe professionnel. Dans ce jeu de données, la gamme dynamique maximale des images HDR est d'environ 25 f-stop, tandis que certaines scènes ont une gamme dynamique d'environ 5 f-stop, et la gamme dynamique médiane est de $13,4$ f-stop. Ici, « f-stop » est une mesure de la dynamique R donnée par l'équation : $R = log_2(Y_{max}) - log_2(Y_{min})$ avec Y_{max} la luminance maximale et Y_{min} la luminance minimale. En photographie, lors de la réduction d'une unité de f-stop, le diamètre de l'ouverture est multiplié par $\sqrt{2}$, et la quantité de lumière entrant dans l'ouverture est multipliée par 2. La figure 3.1 (b) montre la distribution du nombre d'images avec différentes gammes dynamiques. La figure 3.1 (a) montre la relation entre la gamme dynamique et la valeur moyenne du logarithme de la luminance. Nous pouvons observer que les scènes avec une faible gamme dynamique ont généralement une luminance moyenne plus élevée, tandis que les scènes avec une gamme dynamique élevée ont tendance à avoir des valeurs moyennes de luminance plus faibles.

Lors de l'édition des images, le photographe ayant capturé les photographies HDR, devait ajuster manuellement les paramètres de l'image, tels que la valeur d'exposition, le contraste et la courbe de tonalité, en fonction de la scène réelle, de l'esthétique de l'éclairage afin de restituer le plus fidèlement possible l'ambiance de la scène. Durant le processus d'édition de toutes les images HDR, un écran HDR spécifique a été utilisé comme terminal de vérification d'affichage. Si un écran HDR avec configuration légèrement différente (dynamique légèrement différente) est utilisé, la même restitution visuelle peut être obtenue en ajustant les paramètres par une simple mise à l'échelle.

Conditions d'éclairage	Nb d'images	Contenu	Nb d'images
Extérieur lumineux	433	Architecture	477
Soleil dans le cadre	149	Nature	297
Rétro-éclairé	124	Paysage	163
Intérieur moyen	112	Intérieur	188
Ombres profondes	111	Rue	139
Intérieur sombre	84	Paysage urbain	119
Coucher de soleil	80	Bord de mer	74
Extérieur nuageux	72	Véhicule	38
Nuit	54	Portrait	20
Intérieur lumineux	26	Événement	14

TABLEAU 3.1 – Distribution selon les conditions d'éclairage et le contenu de l'image

FIGURE 3.2 – Exemples d'images HDR dans le jeux de données. Un Tone Mapping a été appliqué à des fins de présentation.

Les images contenues dans le jeu de données incluent une variété de scènes, telles

que l'intérieur, l'extérieur, le jour, la nuit, etc. La distribution des images par catégories selon les conditions d'éclairage et le contenu de la scène est montrée dans le tableau 3.1. La figure 3.2 montre une partie des images du jeux de données. Les deux sections suivantes concernant la recherche d'images HDR sont basées sur ce jeux de données.

3.2 Courbe d'ajustement automatique des tonalités

La plupart des méthodes de Tone Mapping mentionnées dans la section 2.6 ont été conçues pour résoudre le problème de l'affichage d'images HDR sur des écrans à faible gamme dynamique. Avec la diffusion rapide des dispositifs d'affichage HDR, nous aimerions disposer d'un moyen d'ajuster directement la courbe de tonalité d'une image HDR sur les dispositifs HDR en fonction de ses caractéristiques HDR (dynamique, profondeur de bits de l'encodage).

La dynamique que les images HDR peuvent enregistrer (jusqu'à 32 f-stops) dépasse la dynamique des écrans HDR grand public (de l'ordre de 12-14 f-stops), de sorte que pour une expérience visuelle de la plus haute qualité, nous devons encore procéder à des ajustements des images HDR pour les adapter aux capacités d'affichage des écrans actuels. Par conséquent, dans cette section, la priorité sera portée sur l'adaptation tonale des images HDR sur les dispositifs d'affichage HDR.

Parmi les paramètres d'ajustement de l'image, la courbe de tonalité est un outil efficace qui permet d'éclaircir ou d'assombrir les zones sombres, d'éclaircir ou d'assombrir les zones claires, de modifier le contraste, etc. Une courbe de tonalité correspond à la définition d'une fonction de transformation des luminances : $Y^* \rightarrow Y'$, où Y^* est la luminance normalisée dans l'intervalle $[0,1]$ ou $[0,100]$.

L'histogramme de l'image est utilisé pour représenter le nombre de pixels à chaque niveau de luminance normalisée de l'image, ce qui permet de montrer la distribution des pixels dans les différents niveaux de luminance normalisée. Nous pouvons modifier la distribution de l'histogramme en ajustant la courbe, ce qui nous permet d'ajuster les tonalités de couleur des différentes zones de l'image. Basé sur le jeu de données HDR présenté dans la section 3.1, nous espérons déterminer la relation entre l'histogramme de l'image et les courbes d'ajustement (voir la figure 3.4 (b)). Pour ce faire, nous construisons un réseau de neurones avec les données de l'histogramme comme entrée et les paramètres de la courbe d'ajustement comme sortie. Cela permet de réaliser une édition automatique à l'aide de la courbe d'ajustement déterminée l'entraînement de réseaux de neurones.

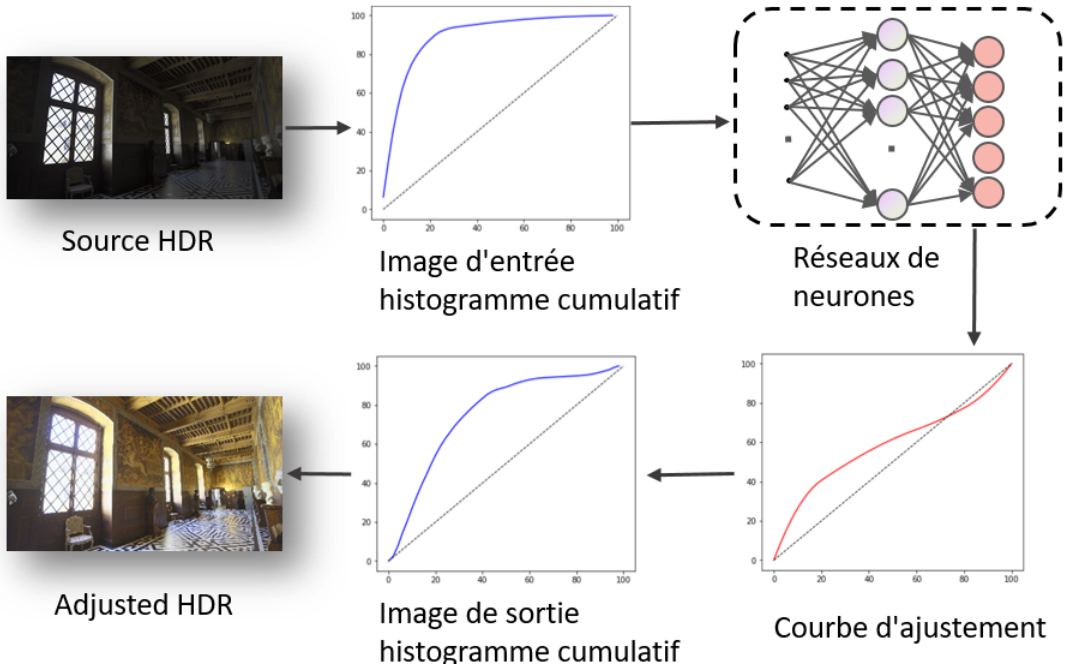
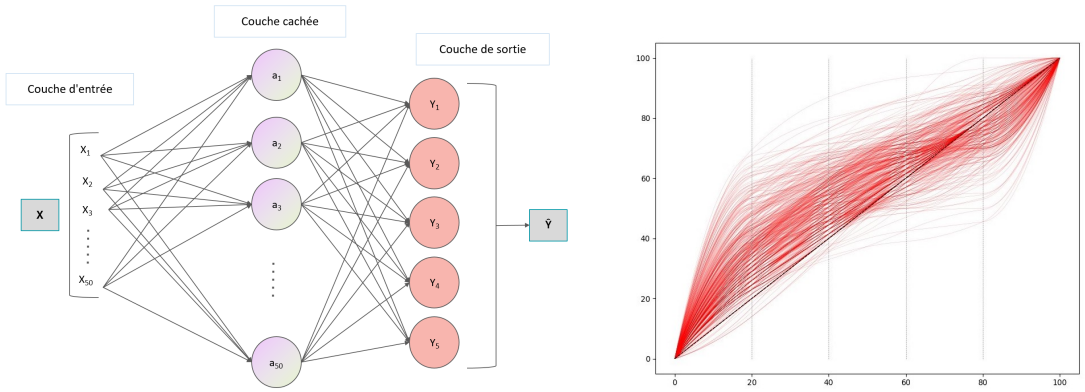


FIGURE 3.3 – Le pipeline de traitement global.

3.2.1 Méthode proposée

Premièrement, nous supposons que les paramètres d'exposition et de contraste de l'image ont été ajustés de manière idéale, dès lors, le dernier paramètre à régler est la courbe d'ajustement des tonalités. Pour les méthodes que nous proposons, l'entrée du modèle est un histogramme cumulatif des luminances de l'image avec 50 points, et la sortie du modèle est une courbe d'ajustement des tonalités déterminée par 5 points clés $P_i(i = 1, 2, ..., 5)$. La valeur initiale des 5 points est de $(10, 10)$, $(30, 30)$, $(50, 50)$, $(70, 70)$ et $(90, 90)$, ce qui permet de contrôler les régions allant des ombres aux hautes lumières. Le point de départ et le point terminal de la courbe sont fixés respectivement à $(0, 0)$ et $(100, 100)$. Les 5 points sont strictement incrémentiels : l'ordonnée du point i est supérieure à l'ordonnée du point $i - 1$. Cela nous permet de garantir une courbe d'ajustement des tons croissante. La sortie P_i inférieure à sa valeur initiale signifie une diminution de la luminance dans cette gamme de luminance. Par contre, la sortie P_i supérieure à sa valeur initiale signifie une augmentation de la luminance dans cette gamme de luminance. Enfin, sur la base de 5 points de contrôle P_i, la courbe est définie par la courbe B-spline définie par ces 5 points de contrôle. La figure 3.3 présente le pipeline de traitement global.

Compte tenu de la relation potentielle entre l'histogramme cumulé et la courbe d'ajustement tonale des images HDR, nous essayons d'apprendre une fonction permettant



(a) architecture (b) courbes

FIGURE 3.4 – Montre l'architecture et les courbes. (a) L'architecture du réseau de neurones. (b) Les courbes ajustées dans le jeu de données HDR.

de déterminer à partir des 50 points de l'histogramme les 5 points de contrôle de la courbe d'ajustement. Nous convertissons l'image HDR normalisée dans l'espace couleur XYZ et utilisons le canal Y comme canal de luminance pour extraire l'histogramme cumulatif. Nous enregistrons l'histogramme sous la forme d'un vecteur de 50 valeurs et la courbe d'édition correspondante sous la forme d'un vecteur de 5 valeurs. Et plus, nous créons un réseau de neurones pour apprendre cette relation de transfert, la figure 3.4 (a) illustre la structure de notre réseau, sa couche cachée comprend une couche entièrement connectée qui est suivie d'une fonction d'activation sigmoïde. Nous utilisons PyTorch pour mettre en œuvre et entraîner cette méthode d'apprentissage. Le processus d'apprentissage se déroule sur un système Windows équipé d'un processeur Intel(R) Core(TM) i9-10885H, le taux d'apprentissage est de $0,1$, avec la fonction de perte de l'erreur quadratique moyenne (MSE), et l'optimiseur Adam (Kingma et al., 2017).

3.2.2 Résultats et évaluations

En testant sur le l'ensemble de validation, les courbes prédites par notre modèle sont proches des courbes professionnelles ajustées manuellement, et les images générées finales sont visuellement très similaires à celles éditées avec les courbes ajustées professionnellement. Comme le montre la figure 3.5, la ligne rouge continue sur le côté droit est le résultat de l'ajustement automatique des courbes, et la ligne rouge en pointillés est le résultat de l'ajustement manuel professionnel. La ligne pointillée bleue et la ligne continue au milieu représentent respectivement les histogrammes des images avant et après l'ajustement. La ligne pointillée rouge est la courbe d'ajustement manuel professionnel, et la ligne continue rouge est la courbe d'ajustement prédite par le réseau

de neurones, et les deux courbes ont une forme similaire, ayant toutes deux éclairci les parties sombres et comprimé les parties lumineuses.

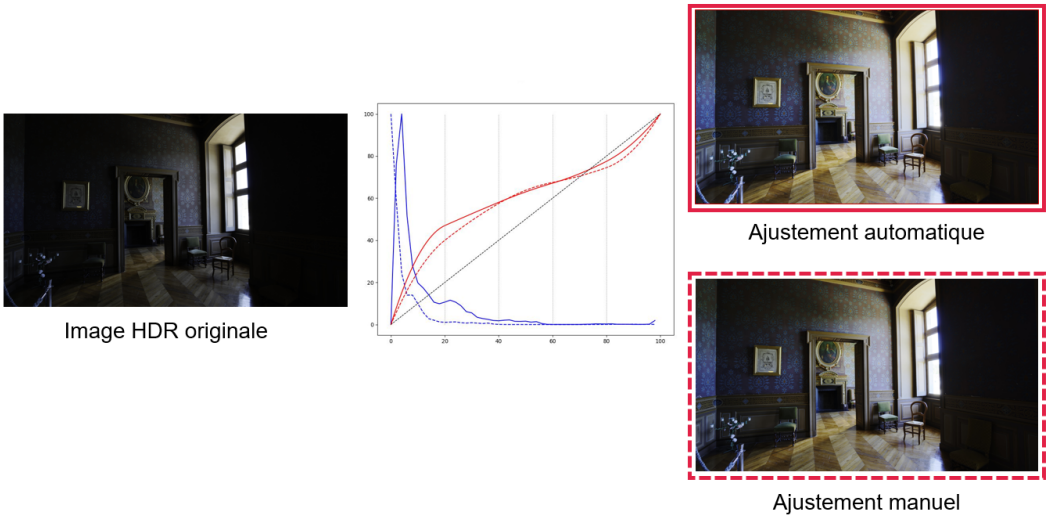
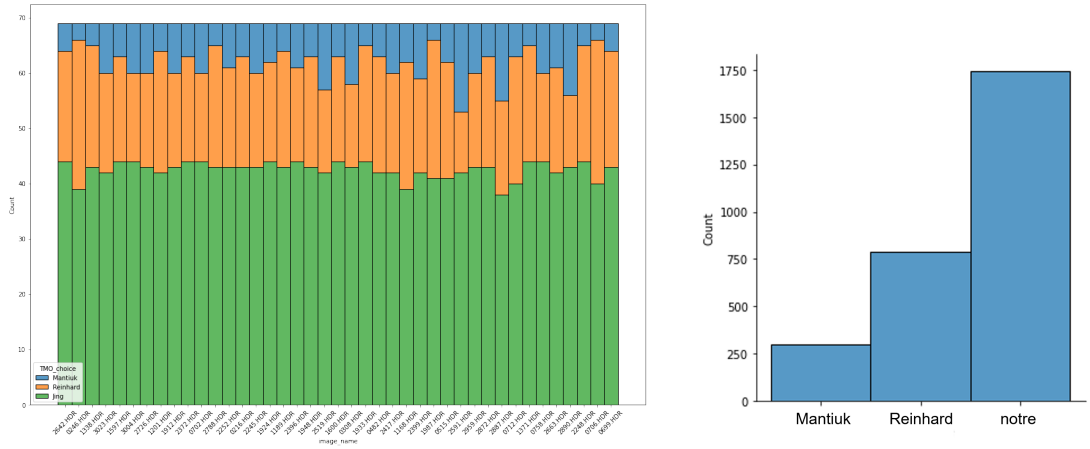


FIGURE 3.5 – Comparaison des courbes automatiques et manuelles. La ligne pointillée rouge est la courbe d'ajustement manuel professionnel, la ligne continue rouge est la courbe d'ajustement de la prédiction du réseau de neurones.

Afin d'évaluer les résultats des courbes prédites par le réseau de neurones, nous avons procédé à une évaluation subjective et à une évaluation objective. Les fonctions de tone mapping de Reinhard (Reinhard et al., 2005) et de Mantiuk (R. Mantiuk et al., 2008) sont deux méthodes couramment utilisées pour l'affichage d'images HDR. Elles ajustent les informations HDR pour obtenir une image adaptée en fonction de l'écran d'affichage, par conséquent, nous avons comparé notre méthode proposée à Reinhard et Mantiuk.

Évaluation subjective

Pour l'évaluation subjective, nous avons utilisé trois écrans HDR du même modèle et avec la même configuration, celui du milieu affichant l'image de référence traitée manuellement et ceux de gauche et de droite affichant de manière aléatoire les résultats des ajustements HDR des différentes méthodes, les participants devant choisir le résultat le plus proche de celui du milieu. Quarante et une images HDR ont été sélectionnées pour l'expérience, et chaque image a été ajustée à l'aide de quatre méthodes d'ajustement : manuelle, Reinhard, Mantiuk, et notre méthode. 26 participants ont participé à cette évaluation, et la figure 3.6 montre que notre méthode produit des résultats plus proches



(a) Résultat pour chaque image test (b) Résultat global

FIGURE 3.6 – Résultats de l'évaluation subjective. (a) La barre verte représente le nombre de personnes pour lesquelles notre méthode a été sélectionnée. (b) Le résultat statistique global des trois méthodes. Nos résultats basés sur l'apprentissage sont davantage sélectionnés que les deux autres méthodes.

de l'ajustement manuel professionnel que les autres méthodes.

Évaluation objective

En plus de l'évaluation subjective, nous avons appliqué quatre critères applicables aux images HDR pour une vérification objective. Premièrement, nous calculons l'entropie de l'information de l'image traitée afin de tester la quantité d'information contenue dans l'image. Ensuite, nous prenons l'image professionnelle ajustée manuellement comme image de référence, et nous utilisons HDR-VDP (R. K. Mantiuk et al., 2023), PU-PSNR et PU-SSIM (Aydın et al., 2008) qui sont les métriques adaptées de l'image HDR pour calculer la différence entre l'image ajustée et l'image de référence. HDR-VDP est utilisé pour indiquer la dégradation de la qualité par rapport à l'image de référence. PU-PSNR sert à évaluer la perception par la vision humaine de la qualité de la reconstruction de l'image, et PU-SSIM indique la similarité structurelle des deux images.

Les scores du tableau 3.2 indiquent qu'en termes de l'entropie de l'information, les résultats de notre courbe automatique contiennent des informations plus riches que les deux autres méthodes. Et pour la qualité de la reconstruction visuelle (PU-PSNR), les résultats entre l'image ajustée par notre courbe automatique a reçu un score de qualité élevé. Les scores de HDR-VDP montrent que les résultats de la méthode Reinhard sont plus proches des résultats réels ajustés par l'expert, mais si la similarité est déterminée

Métrique	**Entropie de l'info**	**HDR-VDP**	**PU-PSNR**	**PU-SSIM**
notre	9,3194	9,0566	20,7981	0,8223
Reinhard	9,0829	9,2969	16,8874	0,8548
Mantiuk	8,990	8,1694	13,1818	0,8906

TABLEAU 3.2 – Scores d'évaluation objectifs.

directement par la perception visuelle, comme le montre la figure 3.7, notre méthode est plus proche de l'image de référence en termes de perception visuelle. Par contre, en termes de similarité structurelle (PU-SSIM), les trois méthodes présentent une grande similarité structurelle avec l'image de référence. La méthode de Mantiuk a obtenu le meilleur score par rapport aux deux autres méthodes, étant donné que la méthode prend en compte la distorsion visible de l'image pendant le processus d'ajustement et préserve plus d'informations structurelles.

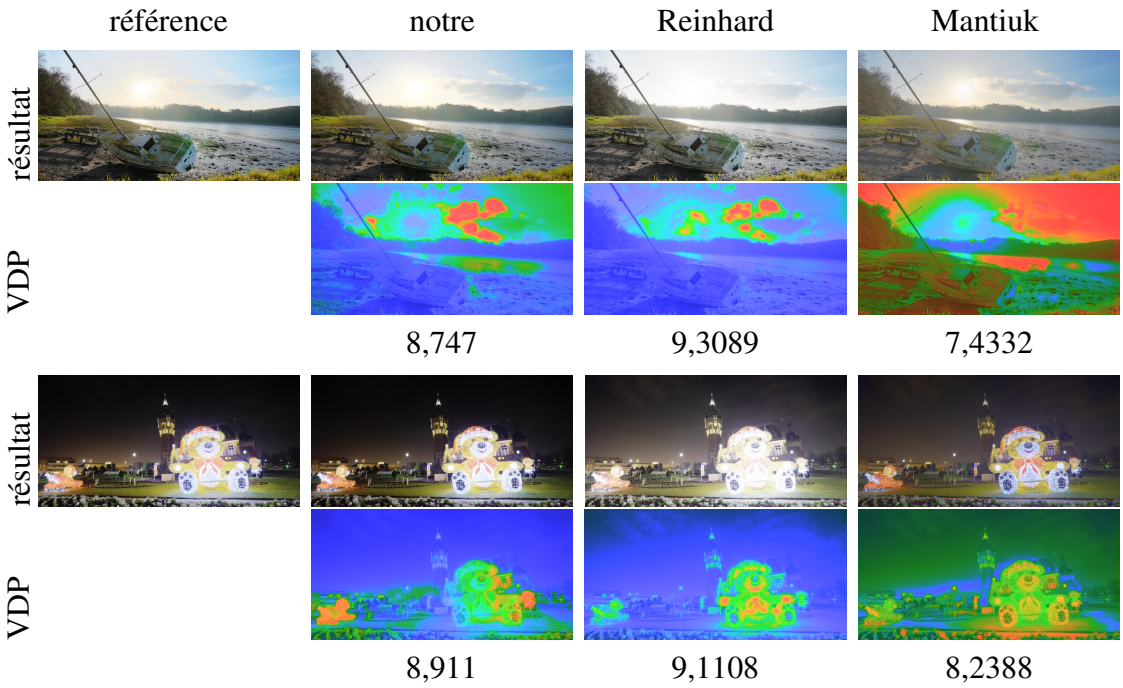


FIGURE 3.7 – VDP résultats. Les résultats du VDP montrent la différence entre les résultats des différentes méthodes et l'image de référence. Dans les deux scénarios, la perception visuelle globale de notre méthode est plus proche de l'image de référence, même si le score de qualité VDP est inférieur à celui de la méthode Reinhard.

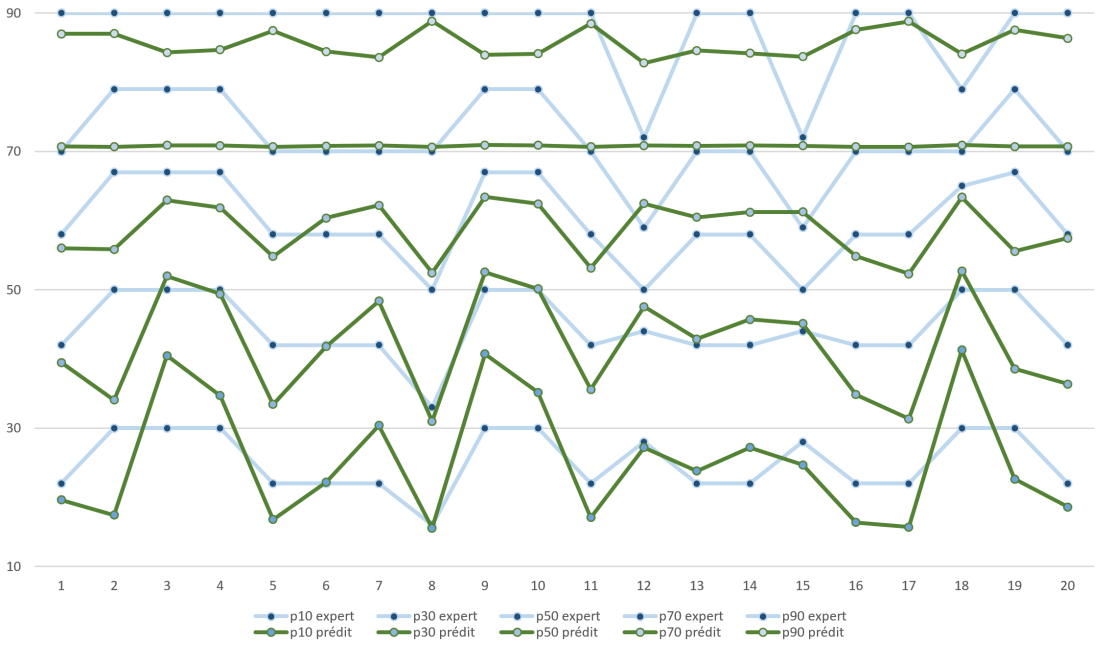


FIGURE 3.8 – Comparaison des valeurs prédites et des valeurs d'ajustement d'expert. L'axe vertical indique les cinq points clés de la courbe et l'axe horizontal indique les numéros des 20 images.

Les courbes de la figure 3.8 illustrent l'écart entre les courbes ajustées par d'expert et nos résultats prédits sur 20 images testées sélectionnées aléatoirement. La moyenne des écarts en cinq points est respectivement de $4,69$. Actuellement, un écart de moins de 10 est tolérable, au sens où il ne produit d'effet visuellement significatif, pour un seul point clé. Par conséquent, notre courbe automatique permet une édition simple, avec les résultats proches de ceux ceux de l'ajustement d'un expert.

3.2.3 Conclusions

Dans cette partie, nous avons mis en œuvre l'ajustement automatique des images HDR en construisant un réseau de neurones simple. La base de données HDR de la section 3.1 est utilisée comme source de données, les paramètres d'édition des courbes sont extraits comme données d'entraînement, et le réseau de neurones est entraîné pour déterminer les 5 points de contrôle de la courbe d'édition en fonction de l'histogramme cumulé de la luminance de l'image HDR. Les résultats expérimentaux montrent que cette méthode permet de réaliser efficacement un ajustement automatique des courbes HDR, facilitant ainsi le résultat d'affichage des images HDR.

3.3 ExposureCNN : adaptation automatique de l'exposition

Dans les scènes photographiques réelles, les conditions d'éclairage sont généralement complexes. Pour préserver la perception des conditions d'éclairage, tout en préservant les détails de l'image, le réglage de l'exposition est essentiel dans le processus d'acquisition et d'édition/retouche de l'image. Contrairement à une image LDR, une image HDR stocke une grande dynamique de luminance, des zones très sombres aux zones très lumineuses. Par conséquent, les méthodes d'ajustement de l'exposition existantes pour les images LDR, dont l'approche est de maximiser le nombre de pixels dans un intervalle limités de luminance, ne s'adaptent généralement pas bien aux images HDR. Pour s'adapter aux caractéristiques des images HDR, de nouvelles méthodes d'ajustement de l'exposition HDR sont nécessaire.

En raison des conditions d'éclairage, de la configuration de l'équipement ou d'autres facteurs externes, les images capturées peuvent être surexposées ou sous-exposées, ce qui donne des images désagréables trop claires ou trop sombres. Pour résoudre ces problèmes, de nombreuses méthodes proposées se concentrent sur les images sous-exposées (X. Guo, 2016 ; Y. Zhang et al., 2019), utilisent la carte d'intensité reconstruite ou une méthode basée sur l'apprentissage pour améliorer l'image sous-exposé.

Des méthodes plus générales (Afifi et al., 2021 ; Hu et al., 2018 ; Yuan et al., 2012 ; Q. Zhang et al., 2019) visent à la fois des problèmes de surexposition et de sous-exposition pour proposer une solution globale au problème de l'ajustement de l'exposition. Cependant, toutes ces méthodes sont basées sur la recherche dans le domaine du LDR et ne prennent pas en compte les caractéristiques de l'image HDR.

En général, une image LDR est prise à partir d'une seule exposition, tandis qu'une image HDR nécessite la capture de plusieurs images de la même scène avec différentes valeurs d'exposition et leur fusion en une carte de luminance, comme indiqué dans la section 2.5. Ce processus ne garantit pas que toutes les images HDR finales soient bien exposées. L'ajustement de l'exposition est donc une opération importante dans le traitement des images HDR.

Concernant le codage des images, une image LDR utilise 8 bits par canal de couleur et ne peut donc stocker qu'une gamme dynamique limitée, alors que l'image HDR utilise 16 à 32 bits pour stocker la luminance de la scène réelle. Par conséquent, les résultats de la recherche basés sur les informations LDR ne sont pas adaptés au traitement direct des images HDR. Prenant en compte les images HDR, notre travail vise à extraire les caractéristiques HDR et à ajuster automatiquement les valeurs d'exposition sur la base

(a) original (b) notre

FIGURE 3.9 – Exemple d'utilisation d'ExposureCNN.

de ces caractéristiques HDR.

Dans la section suivante, nous présentons notre travail de recherche sur le problème de l'ajustement de l'exposition des images HDR. Nous propose une nouvelle approche intitulée ExposureCNN. Il s'agit d'une méthode basée sur l'apprentissage pour l'ajustement automatique de l'exposition dans le domaine HDR. ExposureCNN est basé sur un réseau de neurones convolutif (CNN) pour extraire les caractéristiques d'une image HDR et prédire une valeur d'ajustement de l'exposition. Des expériences sur de multiples images montrent que notre méthode est efficace pour l'ajustement automatique de l'exposition. La figure3.9 montre le résultat de notre méthode.

3.3.1 Méthode proposée

Dans notre méthode d'ajustement automatique de l'exposition, nous commençons par construire un jeu de données d'images HDR et nous y appliquons une augmentation des données. Ensuite, nous introduisons une architecture CNN conçue pour extraire les caractéristiques HDR et prédire la valeur d'exposition. Notre architecture est illustrée dans la figure 3.10, elle prend en entrée une image HDR, et le réseau prédit automatiquement la valeur d'exposition la plus appropriée en sortie.

Prétraitement des données

Le jeu de données HDR introduite dans la section 3.1 contient l'ajustement des valeurs d'exposition, nous nous appuyons donc sur cette base de données pour l'entraînement de notre modèle de réseau de neurones. Cependant, la quantité de données (images d'entrée et valeurs d'exposition correspondantes) n'est pas suffisante pour entraîner un CNN profond. En conséquence, nous avons employé une technique d'augmentation des

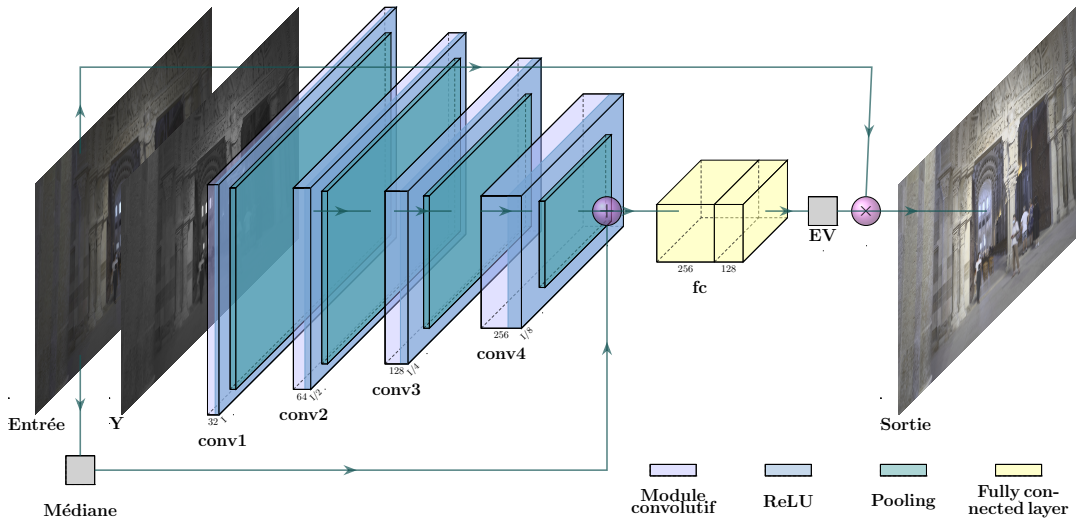


FIGURE 3.10 – Modèle ExposureCNN. *Y* indique le canal de luminance, *Conv* indique le module convolutif, *fc* indique les couches fully connected.

données pour accroître la taille de le jeu de données d'entraînement.

Les méthodes courantes d'augmentation des données telles que la modification des couleurs, le recadrage et la rotation ne conviennent pas à la prédiction de l'exposition car elles modifient la présentation de l'image et affectent la perception visuelle réelle. Il convient de noter que le jugement esthétique d'un photographe professionnel peut varier considérablement lorsque les couleurs et la disposition (par exemple, recadrage, rotation, etc.) varient. Le photographe, en tant qu'artiste, doit transmettre un message basé sur la mise en page du contenu.

Par conséquent, nous avons décidé de n'appliquer qu'une augmentation de la valeur d'exposition à notre jeu de données. Cette augmentation met à l'échelle l'image HDR et ses valeurs d'exposition étiquetées en fonction d'un facteur. L'objectif est que si la luminance globale de l'image augmente ou diminue, le photographe procédera aux ajustements d'exposition correspondants. Nous introduisons un facteur de groupe g qui est un paramètre indiquant le nombre d'augmentation des données, $g = n$ indique l'augmentation de jeu de données n fois. Par exemple, $g = 0$ indique qu'il n'y a pas d'augmentation des données ; $g = 1$ indique un ensemble de données double avec l'ajout d'une valeur d'exposition supplémentaire, chaque image a une valeur de changement stochastique dans l'intervalle $[-1,0;1,0]$; $g = 2$ indique un triple ensemble de données, la valeur de changement stochastique s'ajoutera à la valeur précédente. Finalement, lors

de notre test, nous avons fixé $g = 3$ pour obtenir 2 324 images HDR comme données d'apprentissage.

Architecture du réseau

De nombreuses techniques sophistiquées utilisent une structure CNN pour extraire les caractéristiques HDR, telles que la reconstruction HDR basée sur l'apprentissage (Eilertsen et al., 2017 ; Santos et al., 2020) et l'évaluation de la qualité HDR (Banterle et al., 2020), ces approches nous montrent l'intérêt d'une structure CNN dans l'imagerie HDR. La figure 3.10 montre notre modèle appelé ExposureCNN, il est composé de 4 modules convolutifs et de 2 couches entièrement connectées. Pour réduire les coûts de calcul, les images d'entraînement HDR sont redimensionnées à une résolution de $(128, 256)$. Ensuite, considérant que l'information sur la lumière est la plus importante dans l'image HDR et qu'elle affecte directement la valeur d'exposition, nous extrayons le canal de luminance Y de l'image HDR d'entrée I à l'aide de l'équation 3.1 (ITU, 1990) :

$$Y = 0,2126 * I_R + 0,7152 * I_G + 0,0722 * I_B \tag{3.1}$$

où I est dans l'espace couleur RGB linéaire, et $\{I_R, I_G, I_B\}$ sont les trois canaux de couleur de I. En raison de l'encodage des images HDR dans une large gamme de valeurs, nous transférons Y dans le domaine logarithmique comme l'équation 3.2, ε est une valeur très petite pour éviter le zéro. Nous normalisons ensuite Y_{log} dans l'intervalle de $[-6, 8]$ (ce qui correspond à l'écart d'unités logarithmiques de perception du système visuel humain) comme entrée d'apprentissage du CNN.

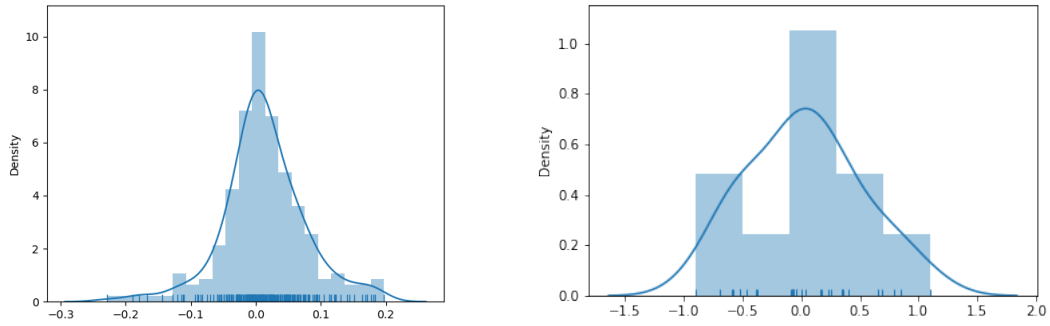
$$Y_{log} = \log_{10}(Y + \varepsilon) \tag{3.2}$$

En plus de l'information normalisée sur la luminance, nous supposons que l'information concernant la gamme dynamique originale est également cruciale pour la prédiction de l'exposition. D'après notre analyse statistique, la valeur médiane du canal de luminance présente une corrélation subtile avec la valeur d'ajustement de l'exposition. Nous calculons donc cette valeur médiane au début de l'apprentissage et la concaténons avec les premières couches entièrement connectées, comme le montre la figure 3.10.

Mise en oeuvre

Nous avons mis en œuvre et entraîné notre méthode à l'aide de PyTorch. Nous utilisons un taux d'apprentissage fixe de 10^{-4}, avec la fonction de perte de l'erreur quadratique moyenne (MSE) et l'optimiseur d'Adam (Kingma et al., 2017). La machine d'apprentissage et d'évaluation est un ordinateur Windows équipé d'un GPU NVIDIA Quadro RTX 3000. Le temps d'apprentissage pour 100 époques est d'environ 9 heures. La figure 3.11 (a) présente l'histogramme de l'erreur dans l'ensemble d'évaluation à la

fin d'apprentissage, la figure 3.11 (b) présente l'histogramme de l'erreur dans le jeu de test avec les poids final.



(a) (b)

FIGURE 3.11 – Histogrammes de l'erreur. (a) est l'histogramme de l'erreur évaluée sur le jeu d'évaluation. (b) est l'histogramme de l'erreur évaluée sur le jeu de test. Notons que l'erreur est centrée autour de 0 et que, dans la majorité des cas, une erreur de ±1 d'exposition est tolérable.

3.3.2 Expérimentations

Pour évaluer notre méthode, nous procédons d'abord à une étude d'ablation afin de tester notre modèle, puis nous effectuons des évaluations subjectives et objectives afin de comparer notre méthode à d'autres méthodes. Les expériences ont été réalisées sur deux jeux de tests, le premier est un jeu HDR que nous collectons de la même manière que les données d'entraînement, il comprend l'image HDR originale et la vérité terrain, le second est le jeu de données Funt [3], nous l'utilisons pour évaluer la performance de généralisation croisée des jeux de données de notre ExposureCNN.

Étude d'ablation

Nous avons essayé différents modèles lors de la conception de l'architecture du réseau, comme l'utilisation de ResNet-18 (K. He et al., 2015) pour extraire les caractéristiques HDR, ou l'entraînement sans concaténation de la valeur médiane d'origine. En outre, pour améliorer la performance de l'apprentissage, nous avons essayé d'appliquer l'apprentissage par transfert, qui consiste à utiliser les poids des couches du réseau entraîné. Cette méthode accélère la convergence du réseau en chargeant les fichiers de poids

3. https ://www2.cs.sfu.ca/~colour/data/funt_hdr

de tâches d'apprentissage similaires. Ici, nous avons chargé les poids de NoR-VDPNet (Banterle et al., 2020), une métrique sans référence utilisée pour l'évaluation de la qualité HDR/SDR, dont les poids des couches convolutives ont déjà permis l'extraction de caractéristiques potentielles d'images HDR.

Métrique	**notre**	**IW**[3]	**NM**[4]	**ResNet-18**
m_e[1]	0,314	0,382	0,359	0,406
δ_e[2]	0,281	0,293	0,304	0,341
HDR-VDP3	9,4037	9,3556	9,3594	9,3559

[1] indique la moyenne de l'erreur absolue ;
[2] indique l'écart-type de l'erreur ;
[3] indique les résultats avec les poids initiaux ;
[4] indique les résultats sans concaténation de la valeur médiane originale.

TABLEAU 3.3 – Scores métriques pour l'étude d'ablation.

(a) Org (b) ours (c) IW (d) NM (e) Res18

FIGURE 3.12 – Résultats de l'étude d'ablation. (a) L'image HDR originale. (b) Le résultat de notre ExposureCNN. (c) Avec les poids initiaux et entraînement avec les mêmes époques. (d) Sans médiane. (e) Emploi ResNet-18.

Les résultats des tests dans le tableau 3.3 montrent que notre réglage actuel est meilleur que les trois autres. La figure 3.12 illustre les résultats visuels de notre étude d'ablation, elle démontre que le modèle sans valeur médiane de luminance (la figure 3.12 (d)) et l'utilisation de ResNet-18 (la figure 3.12 (e)) sont incapables de traiter correctement l'éclairage HDR, et le modèle avec des poids initiaux (la figure 3.12 (c)) est proche de notre ExposureCNN. Au cours de notre test, l'apprentissage avec les poids initiaux converge plus rapidement, mais après un nombre suffisant d'époques d'apprentissage, le résultat sans les poids initiaux est légèrement meilleur.

Comparaisons subjectives

Actuellement, la plupart des méthodes d'ajustement de l'exposition sont basées sur des images LDR. Les méthodes (Afifi et al., 2021 ; Nsampi et al., 2021) sont utilisées pour l'ajustement de l'exposition des images LDR, et ces deux méthodes ne prennent pas en charge les images à haute résolution. Après avoir réduit la résolution des images HDR testées, les résultats ont montré que les deux méthodes étaient incapables de traiter correctement les zones de haute luminance des images HDR, et ont introduit des artefacts dans les zones sombres. De ce fait, les deux méthodes ne sont pas adaptées à nos objectifs actuels d'ajustement d'images HDR.

FIGURE 3.13 – Exemple de comparaison entre notre méthode, BrightHist, White-Box et HDRNet. Nos résultats (c) et les résultats de BrightHist (d) sont plus proches de la référance (b) et ne présentent pas d'aberration chromatique. La plupart des résultats de White-Box (e) sont sous-exposés. Les résultats de HDRNet (f) sont sous-exposés et présentent une aberration chromatique.

Par conséquent, nous avons sélectionné trois méthodes qui sont les plus proches de

(a) original (b) notre (c) BrightHist (d) WhiteBox (e) HDRNet

FIGURE 3.14 – Exemple de comparaison sur le jeu de données de Funt. La perception visuelle de nos résultats (b) et le résultat de BrightHist (c) sont meilleure que les deux autres, elles montre plus de détails et plus de lumière naturelle.

notre objectif de prédiction d'exposition appropriée. La première est Schulz et al. (2007), qui transforme le problème de l'ajustement de la valeur d'exposition en un problème de recherche de la valeur intégrale maximale sur l'histogramme de la gamme dynamique de l'image. Nous le nommerons par la suite « BrightHist ». La deuxième est WhiteBox (Hu et al., 2018), qui est un cadre pour l'édition d'images, WhiteBox utilisent des images RAW comme entrée et prédisent une série d'opérations de post-traitement, qui comprend l'ajustement de l'exposition. La troisième méthode est HDRNet (Gharbi et al., 2017), qui est une méthode d'amélioration d'image largement utilisée. HDRNet applique la grille bilatérale dans l'architecture du réseau pour traiter les images haute résolution.

Nous comparons toutes les images HDR sur le moniteur HDR, les images présentées dans ce document ont été converties en tone mapped par la méthode de Reinhard et al. (2005) à des fins de présentation. La figure 3.13 présente un résultat de comparaison dans quatre scénarios différents (rue lumineuse de la ville avec un contraste élevé, paysage avec le soleil dans le cadre, intérieur sombre avec une vue sur un extérieur lumineux, architecture au coucher du soleil). Nous pouvons remarquer que notre résultat est plus proche de la référence que les autres méthodes. Nos résultats sont visuellement similaires à ceux de BrightHist, et les différences spécifiques peuvent être observées dans la figure 3.15 et dans la section de vérification objective. Nous avons également effectué la même comparaison avec les images du jeu de données de Funt. Étant donné que ce jeu de données n'inclut pas la référence de vérité terrain, nous comparons uniquement leur effet de perception. La figure 3.14 montre que notre méthode produit une image HDR mieux exposée.

Évaluation objective

Concernant l'évaluation objective, nous avons choisi trois mesures appropriées pour le contenu HDR. Toutes sont des métriques à référence complète, y compris l'évaluation de la qualité de l'image HDR la plus couramment utilisée HDR-VDP (R. K. Mantiuk et al., 2023). Les métriques standard pour les images LDR fonctionnent dans le domaine non linéaire et ne sont pas adaptées pour être appliquées directement aux images HDR linéaires, telles que PSNR (Peak Signal to Noise Ratio) et SSIM (Structural Similarity) sont couramment utilisées comme métriques LDR. Pour les adapter au domaine HDR, nous appliquons PU-PSNR et PU-SSIM (R. K. Mantiuk et al., 2021), qui combinent le codage PU (Perceptually Uniform) et le transfert PQ (Perceptual Quantizer) pour l'adaptation des images HDR. Le tableau 3.4 indique les résultats statistiques de ces métriques, un score plus élevé signifie une meilleure qualité, la méthode proposée présente un avantage par rapport aux autres. La figure 3.15 montre la différence précise entre la méthode proposée et la méthode de BrightHist.

Métrique	**ExposureCNN**	**BrightHist**	**WhiteBox**	**HDRNet**
HDR-VDP3	8,8569	8,6146	6,2775	5,1396
PU-SSIM	0,8505	0,7856	0,4563	0,4495
PU-PSNR(dB)	23,0850	20,7701	6,4400	6,3420

TABLEAU 3.4 – Notes métriques pour notre méthode et d'autres.

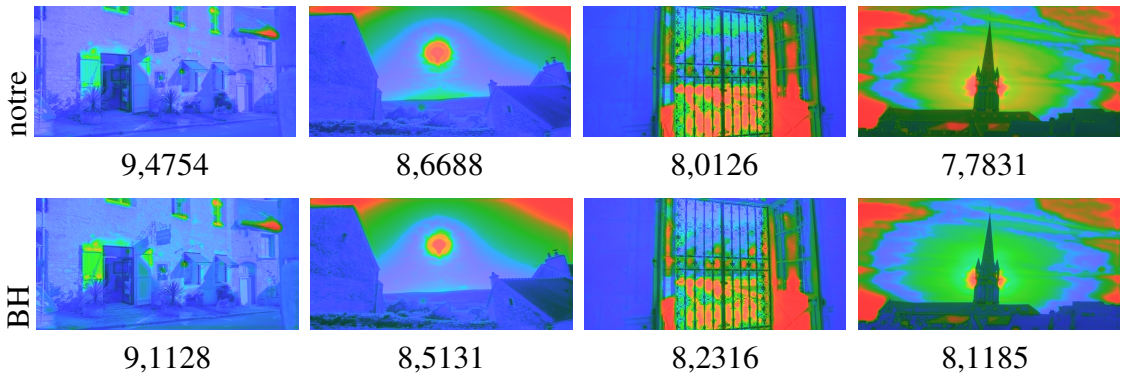
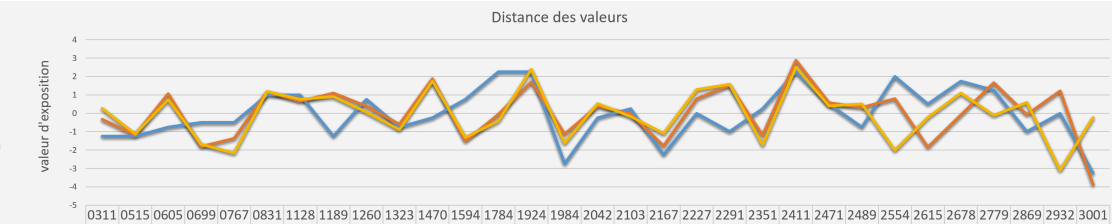


FIGURE 3.15 – VDP visualisation. La ligne ci-dessus montre la différence entre nos résultats et les images de référence. La ligne suivante montre la différence entre les résultats de BrightHist et les images de référence. Les chiffres représentent les valeurs VDP correspondantes.

Étant donné que la méthode proposée par BrightHist peut donner des valeurs d'expo-

sition précises et que les résultats du traitement sont visuellement très ressemblants à notre méthode, nous avons comparé les valeurs d'exposition données par celle-ci avec les valeurs ajustées par l'expert. Comme le montre la figure 3.16, la courbe bleue représente les valeurs d'exposition dérivées de la méthode BrightHist, la courbe orange représente les valeurs prédites par notre ExposureCNN, la courbe jaune représente les valeurs réelles ajustées par l'expert, et sous les courbes se trouvent les trois valeurs d'ajustement de l'exposition pour des images de numéros de série différents. Les courbes de la figure montrent que la trajectoire de la courbe orange correspond davantage à celle de la courbe jaune. Au travers des statistiques, la moyenne de l'erreur entre la valeur de BrightHist et la valeur d'expert est $1,3434$, et la moyenne de l'erreur entre notre ExposureCNN et la valeur d'expert est $0,7540$. Cela montre que les valeurs prédites de notre ExposureCNN sont plus proches des valeurs réelles ajustées par l'expert.



	0311	0515	0605	0699	0767	0831	1128	1189	1260	1323	1470	1594	1784	1924	1984	2042	2103	2167	2227	2291	2351	2411	2471	2489	2554	2615	2678	2779	2869	2932	3001
HB	-1.3	-1.3	-0.8	-0.5	-0.5	1	1	-1.3	0.75	-0.8	-0.3	0.75	2.25	2.25	-2.8	-0.3	0.25	-2.3	0	-1	0.25	2.25	0.5	-0.8	2	0.5	1.75	1.25	-1	0	-3.3
EVCNN	-0.3	-1.2	1.07	-1.8	-1.4	1.17	0.66	1.1	0.38	-0.6	1.89	-1.5	-0.1	1.72	-1.1	0.41	-0	-1.8	0.79	1.52	-1.2	2.88	0.57	0.32	0.8	-1.8	-0.1	1.65	-0.1	1.22	-3.9
expert	0.27	-1.1	0.74	-1.7	-2.2	1.18	0.75	0.91	0.07	-0.9	1.74	-1.3	-0.4	2.39	-1.6	0.52	-0.2	-1.1	1.28	1.57	-1.7	2.5	0.4	0.5	-2	-0.2	1.09	-0.1	0.59	-3.1	-0.2

FIGURE 3.16 – Courbes statistiques des trois méthodes.

En résumé, si l'on tient compte des résultats expérimentaux, le modèle ExposureCNN que nous proposons est simple mais efficace, et ses performances sont supérieures à celles d'autres méthodes, tant au niveau de la comparaison subjective que de l'évaluation objective.

3.3.3 Conclusion

Dans cette section, nous avons présenté ExposureCNN, une nouvelle méthode d'ajustement de l'exposition basée sur l'apprentissage. Cette méthode utilise le réseau CNN pour extraire les caractéristiques HDR, puis, sur la base de ces caractéristiques, ajuster les valeurs d'exposition des images HDR. Les résultats expérimentaux ont montré que l'architecture de l'ExposureCNN est efficace pour traiter les images HDR. Lors des comparaisons subjectives et de l'évaluation objective, elle surpasse les autres méthodes dans divers scénarios. Dans nos travaux futurs, nous aimerions étendre notre approche aux vidéos HDR, en tenant compte de la cohérence temporelle des valeurs d'exposition.

Résumé

Ce chapitre présente la possibilité d'explorer les caractéristiques potentielles des images HDR grâce à des méthodes d'apprentissage automatique. Nous commençons par introduire un jeu de données d'images HDR créée par un photographe professionnel. Cette base de données contient une variété de scènes et les paramètres d'édition associés, fournissant ainsi des données d'entraînement riches pour le modèle d'apprentissage automatique. Nous établissons tout d'abord la relation de correspondance entre l'histogramme et la courbe d'ajustement des tonalités des images HDR par un modèle de réseau de neurones simple. Cela permet d'ajuster différentes zones de l'image HDR par la prédiction de courbes. Ensuite, nous avons construit un modèle CNN pour prédire la valeur de correction de l'exposition en extrayant les caractéristiques de l'image HDR. Les expériences montrent que les deux méthodes proposées peuvent prédire des paramètres d'ajustement proches de ceux des photographes professionnels, ce qui permet d'améliorer automatiquement l'expérience visuelle des images HDR.

En plus de la gestion des gammes de luminance, dimension importante quant à l'esthétique, nous souhaitons également trouver des méthodes pour analyser d'autres caractéristiques/dimensions esthétiques contenues dans les images HDR. Par conséquent, au chapitre 4, nous résumons les concepts et les méthodes impliqués dans l'analyse esthétique, et au chapitre 5, nous proposons une approche de reconstruction des lignes de force pour comprendre l'esthétique de la composition.

Chapitre 4

Esthétique de l'image

Introduction

Des caractéristiques telles que la lumière, la couleur et la composition constituent l'image entière. La perception de ces caractéristiques affecte l'expérience visuelle de l'esthétique de l'image. Dans ce chapitre, nous présenterons la notion de qualité esthétique, d'analyse des caractéristiques esthétiques, les méthodes d'évaluation de la qualité de l'esthétique des images mais aussi les jeux de données liés à l'étude de la qualité esthétique, ainsi que les méthodes et les significations de l'analyse des caractéristiques esthétiques de l'image.

4.1 Notion d'esthétique de l'image

Le thème de l'esthétique de l'image est un vaste domaine qui fait appel à diverses disciplines, notamment l'histoire de l'art, la philosophie, la psychologie, les arts visuels et l'informatique. Par exemple, le philosophe du XVIIIe siècle Emmanuel Kant pensait que l'esthétique décrit notre appréciation du monde extérieur par l'intermédiaire de nos sens (« Le Robert Illustré », 2018). Dans cette thèse, nous nous concentrerons sur les qualités esthétiques et les caractéristiques esthétiques des images numériques. Il s'agit d'analyser les éléments spécifiques qui affectent la qualité esthétique des images et, sur cette base, d'améliorer l'expérience visuelle.

Lorsque nous discutons des qualités esthétiques des images numériques, l'objectif habituel est d'évaluer la beauté de l'image ou de trouver un moyen exécutable par ordinateur d'améliorer l'affichage des images numériques sur l'appareil concerné. Il s'agit d'utiliser des méthodes informatiques pour comprendre, évaluer et créer des images esthétiquement agréables. L'exploration de ce domaine intègre des connaissances en matière de vision par ordinateur, d'apprentissage automatique, de traitement d'images et

de principes esthétiques. Nous faisons l'hypothèse que certains aspects de l'esthétique peuvent être calculés et perçus par un ordinateur, tout comme le définit dans Hoenig (2005). L'esthétique informatique est la recherche de méthodes informatiques qui peuvent prendre des décisions esthétiques applicables de manière similaire à celle des humains.

En termes d'application pratiques, les créateurs utilisent souvent des règles éprouvées pour améliorer la qualité esthétique des images (voir la figure 4.1). Cela inclut l'application de principes dans la composition, ainsi que l'harmonisation des couleurs et des choix judicieux en termes de contraste, de saturation et d'exposition. La règle des tiers, la division en section d'or et les lignes de force peuvent améliorer l'attrait visuel. Ces éléments sont relatifs au style de l'image que nous appelons « caractéristiques esthétiques ». Différentes combinaisons de couleurs peuvent évoquer différentes émotions et ambiance, et un choix approprié des couleurs peut renforcer le thème de l'image et le message transmis.

FIGURE 4.1 – Améliorer la qualité esthétique des images grâce à des règles empiriques.

En revanche, d'un point de vue subjectif, la perception de la qualité esthétique est influencée par des facteurs tels que les préférences personnelles, le contexte culturel, la situation et le contexte. Une personne peut aimer un style ou un thème particulier, tandis qu'une autre peut ne pas l'apprécier. Les standards de compréhension de la beauté varient selon les différents contextes culturels. De plus, même pour une même image, le même spectateur peut avoir des perceptions esthétiques différentes à des moments différents.

Dans l'analyse informatique, les attributs esthétiques d'une image se divise en carac-

téristiques d'ordre inférieur et d'ordre supérieur. Les caractéristiques d'ordre inférieur englobent des éléments tels que la couleur, l'illumination, la texture et les lignes physiques. En revanche, les caractéristiques d'ordre supérieur comprennent des aspects plus complexes tels que la composition spatiale, le contexte thématique et le style artistique. La recherche de la qualité esthétique des images vise principalement à établir un lien entre les caractéristiques visuelles de bas niveau, qui sont calculables, et les sémantiques de haut niveau liées à la perception humaine.

Au départ, la recherche s'est surtout attachée à identifier les caractéristiques visuelles de base qui influencent la qualité esthétique d'une image, telles que la couleur, la ligne, la texture, la forme, le contraste, etc. En se basant sur l'extraction de caractéristiques, le chercheur tente de construire un modèle d'évaluation esthétique calculable (Birkhoff, 2013 ; Henry, 1885). Dans les dernières années, grâce à l'amélioration de la technologie de l'apprentissage profond, des modèles d'apprentissage profond ont également été appliqués à la recherche sur la qualité esthétique des images.

4.2 Critères esthétiques de l'image

Le mathématicien George David Birkhoff a introduit une théorie de la mesure esthétique, il s'agit d'une formule mathématique permettant de quantifier la valeur esthétique. La formule de base est $M = \frac{O}{C}$, où M est la mesure esthétique, O représente l'ordre, et C représente la complexité. Si nous l'appliquons à la qualité esthétique des images, l'ordre peut concerner la symétrie, la répétition ou la composition et l'harmonie des couleurs, etc. La complexité peut concerner le nombre d'éléments, la complexité du motif ou la variété des couleurs et des formes, etc. Le modèle de mesure esthétique de Birkhoff constitue une tentative de quantification de la qualité esthétique, mais son application pratique peut être affectée par un certain nombre de facteurs, notamment le jugement subjectif de l'évaluateur et les caractéristiques de l'œuvre d'art elle-même (Birkhoff, 2013 ; Douchová, 2016). Puis, Max Bense a combiné le concept de la mesure esthétique de Birkhoff avec la théorie de l'entropie de l'information de Claude Shannon (Shannon, 1984), ce qui a donné naissance à la qualité esthétique de l'information.

Ces modèles de mesure de la qualité esthétique, applicables dans divers domaines tels que les images, la musique, ou l'architecture, ouvrent des perspectives innovantes en matière d'évaluation esthétique. Leur concept, reposant sur la possibilité de modéliser et d'analyser la qualité esthétique, constitue un pilier fondamental pour les recherches futures dans ce domaine. Toutefois, l'utilisation de ces modèles pour les jugements de la qualité esthétique quantitatifs et absolus ne tient pas compte de la nature variable et dépendante du contexte des jugements esthétiques visuelles.

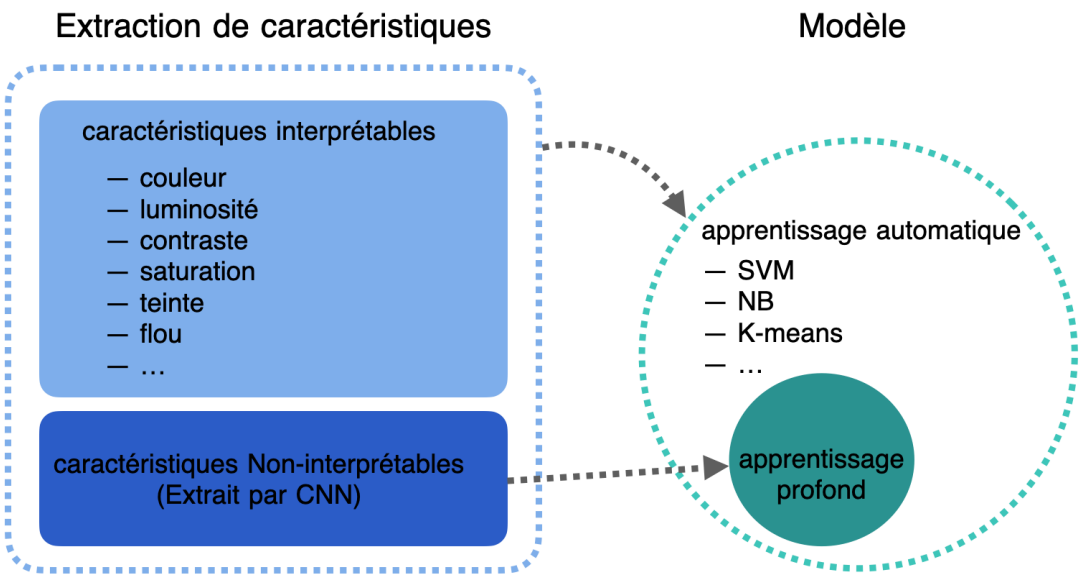


FIGURE 4.2 – Aperçu de la prédiction de la qualité esthétique des images.

4.2.1 Modèles traditionnels d'apprentissage automatique

Au début de la recherche sur la qualité esthétique des images, les chercheurs espéraient que la qualité esthétique des images pourrait être évaluée à partir des caractéristiques d'ordre inférieur de l'image. La recherche de Tong et al. (2004) construit un classificateur pour distinguer les photographies prises par des photographes professionnels et celles prises par des utilisateurs ordinaires, l'idée principale est d'utiliser l'analyse en composantes principales (PCA) pour extraire certaines caractéristiques significatives de l'image et de les transmettre aux machines à vecteurs de support (SVM) ou à un classificateur bayésien pour la classification. De plus, Yan Ke et al. (2006) présentent un modèle de classification, qui prend en compte à la fois les propriétés d'ordre supérieur des images, telles que la carte des contours spatiaux, la distribution des couleurs, la teinte et le flou, et les propriétés d'ordre inférieur, telles que le contraste et la luminance, utilise le classificateur de Bayes naïf pour distinguer les images professionnelles de haute qualité esthétique et les images ordinaires de faible qualité esthétique.

L'approche de Datta et al. (2006) utilise des SVM pour prédire les scores esthétiques des images. Cette méthode distingue les images à haute qualité esthétique de celles en manquant, en se basant sur 15 caractéristiques visuelles les plus précisément définies par le modèle. Certaines études ultérieures ont amélioré la précision de la prédiction

en ajoutant des caractéristiques locales ou des corrélations entre les caractéristiques locales (Joshi et al., 2011). Datte a ensuite développé ACQUINE (Datta et al., 2010), un site d'analyse esthétique en ligne qui évalue automatiquement la qualité esthétique des images importées par les utilisateurs (actuellement indisponible).

Bien d'autres recherches sur la qualité esthétique des images utilisent également SVM, elles sont généralement basées sur différentes approches de l'extraction des caractéristiques. Nishiyama et al. (2011) utilisant l'harmonie des couleurs comme référence principale pour la notation esthétique. Marchesotti et al. (2012) introduisent les descripteurs génériques d'images, telles que le sac-de-mots (BOV) et le vecteur de Fisher (FV), pour la classification. Le FV (Perronnin et al., 2007) capture la distribution des caractéristiques locales dans une image sur la base d'un modèle statistique. Le BOV (Csurka et al., 2004) est couramment utilisé dans les tâches de reconnaissance et de classification d'images. Il s'appuie sur le modèle des sacs de mots du traitement du langage naturel, où les caractéristiques extraites de plusieurs images sont converties en une collection de « vocabulaire visuel ».

Certains travaux de recherche intègrent également la reconnaissance des objets saillants (Dhar et al., 2011) et du contenu des images (Wei Luo et al., 2011) en plus de l'extraction classique des caractéristiques visuelles. Cependant, les caractéristiques esthétiques extraites sur la base des conceptions algorithmiques traditionnelles présentent généralement certaines limites. Elles sont souvent basées sur la connaissance empirique des attributs existants et décrivent le concept abstrait de l'esthétique avec une combinaison de paramètres approximatifs. Ces méthodes fournissent donc des modèles interprétables pour la qualité esthétique de nos images cognitives, mais elles ne sont pas encore assez complètes pour analyser une perception visuelle de dimension supérieure.

4.2.2 Modèles d'apprentissage profond

L'apprentissage en profondeur est une méthode d'apprentissage utilisant des réseaux neuronaux profonds, qui est un type d'apprentissage automatique. La caractéristique principale de l'apprentissage profond est sa capacité à construire et à former des modèles de réseaux complexes qui capturent des caractéristiques plus avancées et plus abstraites des données. Les concepts et méthodes fondamentaux de l'apprentissage profond sont présentés dans « fidle » (s. d.) et l'ouvrage de Geron (2019).

L'histoire de la modélisation des réseaux neuronaux trouve ses racines dans les années 1940 et 1950, période durant laquelle les systèmes nerveux biologiques ont servi de source d'inspiration fondamentale. C'est à cette époque que Warren McCulloch et Walter Pitts ont introduit le modèle de neurones McCulloch-Pitts (Chandra, 2022), jetant

ainsi les bases théoriques des réseaux neuronaux. Toutefois, l'essor de ces réseaux neuronaux artificiels a été fortement entravé par les capacités de calcul limitées de l'époque, retardant ainsi leur développement et leur application pratique.

Actuellement, l'augmentation de la capacité de calcul et l'essor du *big data* ont donné une forte impulsion au développement des réseaux neuronaux profonds. La victoire écrasante d'AlexNet (Krizhevsky et al., 2012) au concours de reconnaissance visuelle à grande échelle ImageNet (ILSVRC) en 2012 a démontré le grand potentiel des réseaux neuronaux convolutifs (CNN) dans la gestion de tâches visuelles complexes.

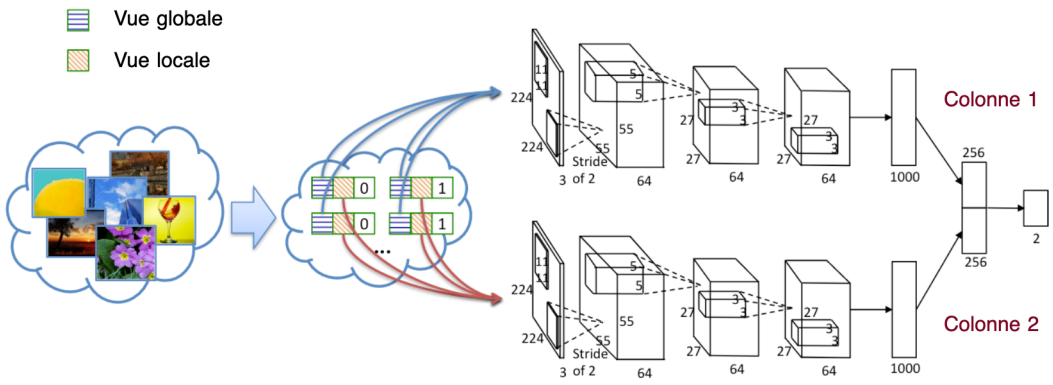


FIGURE 4.3 – Illustration de RAPID. La première colonne est utilisée pour extraire les caractéristiques globales et la deuxième colonne est utilisée pour extraire les caractéristiques locales (Lu et al., 2014).

Basé sur le travail de AlexNet, RAPID (Lu et al., 2014) applique un modèle de CNN à l'évaluation de la qualité esthétique des images. RAPID construit une structure CNN à deux colonnes (voir la figure4.3), où la première colonne du CNN est utilisée pour extraire les caractéristiques globales de l'image entière et la deuxième colonne est utilisée pour extraire les caractéristiques locales de la segmentation aléatoire. Enfin, un résultat binaire pour la qualité esthétique de l'image est prédit en entraînant conjointement les couches entièrement connectées.

DMA-Net (Deep Multi-Patch Aggregation Network) (Lu et al., 2015) divise l'image d'entrée en cinq patchs de détails fins pour l'opération de convolution afin d'obtenir des informations plus détaillées dans l'image, et effectue une évaluation de la qualité esthétique en intégrant les caractéristiques extraites par les 5 CNN. A-Lamp (Ma et al., 2017) est aussi un modèle multi-patch, mais il y ajoute des mécanismes d'attention et des caractéristiques de disposition afin d'améliorer la précision de la prédiction. Il existe

également des modèles multi-patchs similaires qui intègrent des mécanismes d'attention, combiné à l'architecture ResNet (K. He et al., 2015) pour l'extraction de caractéristiques d'images (Sheng et al., 2018). Cette combinaison a démontré une précision accrue dans les résultats obtenus lors des expériences. Outre le modèle multi-patch, Kao et al. (2017) proposent un modèle CNN multi-tâche qui prédit les scores de qualité esthétique tout en donnant une classification sémantique des images.

L'opération de convolution est au cœur des CNN, ce qui entraîne inévitablement une perte d'informations structurelles spatiales lors de l'extraction des caractéristiques des images. Les CNN initiaux étaient principalement utilisés pour des tâches telles que la classification d'images et la reconnaissance d'objets, où les modifications de la taille et de la composition des images n'avaient pas d'impact majeur sur les résultats de ces tâches. Cependant, dans les tâches d'évaluation de la qualité esthétique, la réduction de la taille des images et les changements dans leur composition peuvent souvent influencer directement le score de qualité esthétique. Pour aborder ce problème, Mai et al. (2016) et Hosu et al. (2019) ont respectivement proposé la couche de pooling spatial adaptatif (ASP) et les blocs d'activation multi-niveaux à pool spatiale (MLSP). Le concept clé de leur approche consiste à moduler la dimension du champ récepteur, tout en conservant une taille de sortie constante lors de l'opération de pooling. Cette technique permet d'adapter le traitement à des images de dimensions et de proportions diverses (voir la figure 4.4).

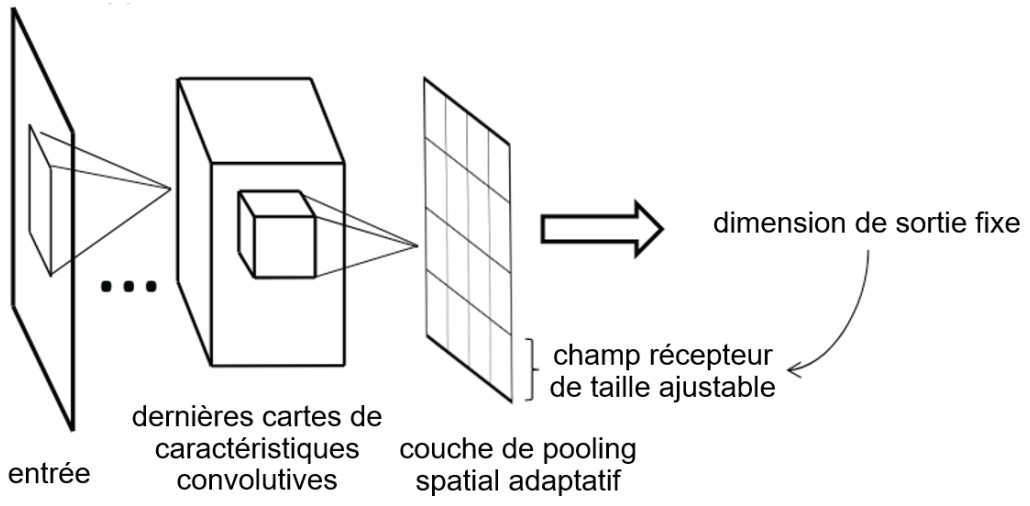


FIGURE 4.4 – Structure de la couche de pooling spatial adaptatif (Mai et al., 2016).

Généralement, les modèles d'évaluation de la qualité esthétique sont des sorties

directes de scores esthétiques prédits, et la référence réelle dans l'entraînement est également la moyenne des notes de tous les participants dans la base de données, de telle sorte que les prédictions ne présentent pas les préférences subjectives du groupe de participants. Par conséquent, NIMA (Talebi et al., 2018) adapte la sortie du réseau convolutif à une distribution de scores prédits, une distribution qui montre les différences possibles dans la perception de la qualité esthétique de l'image par différents spectateurs, fournissant ainsi des informations plus riches et plus nuancées sur la qualité de l'esthétique de l'image.

NIMA a mis à l'essai trois architectures CNN pour l'extraction de caractéristiques esthétiques à partir d'images, à savoir VGG16 (Levina et al., 2001), Inception-v2 (Szegedy et al., 2015) et MobileNet (Howard et al., 2017), et a constaté expérimentalement que la précision d'Inception-v2 était légèrement supérieure à celle des deux autres architectures pour la tâche d'extraction de caractéristiques esthétiques.

Étant donné que les prédictions du modèle NIMA sont des distributions notées sous forme d'histogrammes, au lieu d'utiliser une fonction de perte conventionnelle dans le modèle, la distance normalisée de Earth Mover's (EMD) (Levina et al., 2001) 4.1 est appliqué pour calculer l'erreur de la distribution.

$$EMD(p,\widehat{p}) = \left(\frac{1}{N}\sum_{k=1}^{N}\left|CDF_p(k) - CDF_{\widehat{p}}(k)\right|^2\right)^{1/2} \tag{4.1}$$

où p est la fonction de masse de probabilité vérité, $\widehat{p}$ est la fonction de masse de probabilité estimées, CDF est la fonction de distribution cumulative, N indique le nombre total d'unités de mesure (dans le jeu de donnée AVA $N = 10$).

De même, pour prédire la distribution des scores, le modèle de X. Jin et al. (2017) intègre les concepts de la divergence cumulative symétrique discrète de Jensen-Shannon et le kurtosis pour mesurer la fiabilité dans la fonction de perte $loss^{RS-CJS}$, voir la fonction de 4.24.3, où $r^{kurtosis}(p)$ est le facteur de fiabilité.

$$loss^{RS-CJS}(p,\widehat{p}) = r^{kurtosis}(p)CJS(p,\widehat{p}) \tag{4.2}$$

$$\begin{aligned} CJS(p,\widehat{p}) = \frac{1}{2}\Bigg(&\sum_{k=1}^{N} CDF_p(k) log \frac{CDF_p(k)}{\frac{1}{2}CDF_p(k) + \frac{1}{2}CDF_{\widehat{p}}(k)} \\ &+ \sum_{k=1}^{N} CDF_{\widehat{p}}(k) log \frac{CDF_{\widehat{p}}(k)}{\frac{1}{2}CDF_p(k) + \frac{1}{2}CDF_{\widehat{p}}(k)}\Bigg) \end{aligned} \tag{4.3}$$

La recherche de Chambe et al. (2019) a affiné le modèle NIMA en rassemblant une

collection de six types d'images photographiques professionnelles afin d'améliorer la précision de la prédiction. Néanmoins, l'évaluation de la qualité esthétique basée sur une seule image est relativement compliquée. Elle peut être perturbée par le scénario de l'environnement ou l'état de l'évaluateur. Par conséquent, Kong et al. (2016) proposent un modèle de rangement esthétique qui prend en compte tant les paramètres que le contenu. Ce modèle est basé sur l'architecture siamoise (Chopra et al., 2005) et accepte une paire d'images comme entrée et donne une prédiction du rang de la qualité esthétique.

L'équipe de Google Research a proposé en 2019 la famille de modèles EfficientNet (Tan et al., 2020), un modèle CNN plus efficace qui permet une classification d'images performante avec des coûts de calcul et de paramètres plus réduits. À partir du modèle EfficientNet, X. Jin et al. (2022) proposent un modèle de réseau multitâche qui comprend un réseau principal EfficientNet pour l'extraction de caractéristiques profondes, un sous-réseau de régression pour prédire les scores globaux, un sous-réseau de classification des caractéristiques esthétiques pour guider le sous-réseau de régression, et trois sous-réseaux génériques pour prédire les scores d'attributs. Ce modèle prédit une note esthétique pour l'image entière tout en donnant des notes esthétiques pour trois attributs : l'espace, la couleur et la composition.

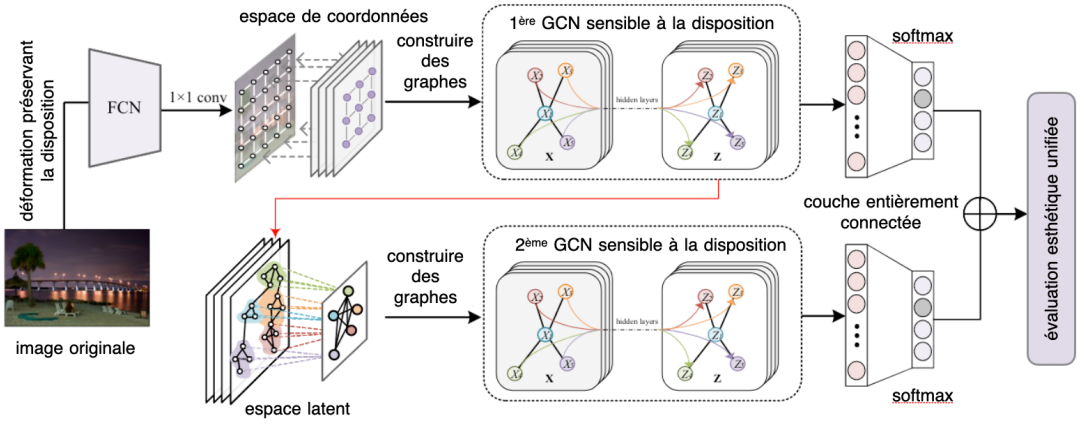


FIGURE 4.5 – pipeline de HLA-GCN (She et al., 2021) .

HLA-GCN (She et al., 2021) construit un réseau convolutif graphique à deux couches hiérarchique tenant compte de la mise en page, basé sur un modèle de réseaux neuronaux graphiques(GCN) (Kipf et al., 2017) . La première couche du réseau est utilisée pour construire un graphe esthétiquement pertinent dans l'espace des coordonnées. La deuxième couche est chargée d'effectuer l'inférence du graphe dans l'espace latent. Enfin les caractéristiques des deux couches sont intégrées par une couche entièrement

connectée pour la prédiction de la qualité esthétique (voir la figure 4.5).

Dans le domaine plus spécialisé de l'évaluation de la qualité esthétique des couleurs, S. He et al. (2023) identifient les points d'intérêt grâce au transformateur de délégué et intègre un module statistique capable de diviser rationnellement l'espace colorimétrique pour prédire la qualité esthétique des couleurs d'une image.

Les images numériques dont nous discutons englobent une variété de genres et ne se limitent pas à la photographie. Yi et al. (2023) proposent un modèle d'évaluation de la qualité esthétique qui utilise une combinaison du modèle VGG-19 et du modèle ResNet-50 pour extraire les caractéristiques stylistiques et esthétiques des images, afin d'évaluer la qualité esthétique des images artistiques.

La plupart des modèles d'évaluation de la qualité esthétique actuels donnent une note binaire, bonne ou mauvaise, ou une note esthétique spécifique, mais cette évaluation est uniquement numérique et ne comprend pas d'évaluation linguistiques spécifiques. Le modèle de Kuang-Yu Chang et al. (2017) est la première tentative de mesurer la qualité esthétique d'une image sous la forme d'un texte. La méthode Fusion d'aspects proposée ci-dessous génère avec succès des évaluations linguistiques esthétiques d'images en optimisant l'architecture CNN-LSTM (Hochreiter et al., 1997 ; Xu et al., 2016). Par la suite, de manière similaire, X. Jin et al. (2019) proposent un réseau multi-attributs esthétique pour générer des évaluation linguistiques et des scores de caractéristiques esthétiques individuels.

4.3 Données esthétiques de l'image

Grâce à la démocratisation de l'internet, il est facile de partager et de rechercher différents types d'images. Certaines plateformes en ligne, qui offrent un espace aux amateurs et aux professionnels de la photographie pour présenter et améliorer leurs compétences photographiques, permettent aux utilisateurs de télécharger des photographies et d'évaluer d'autres travaux. Ces plateformes ouvertes fournissent également de riches données de recherche aux chercheurs en image.

Photo.net (« PhotoNet Home », 2024) est un site web sur lequel les photographes notent leurs photos entre eux. Les utilisateurs peuvent déposer et afficher leurs photographies dans des galeries personnelles. Ces galeries sont ouvertes au public et permettent aux autres utilisateurs de les visualiser, de les commenter et de les évaluer. Les photographies sont notées sur la qualité esthétique avec des notes allant de 1 à 7, plus la note est élevée, plus la qualité esthétique et l'originalité sont élevées.

DPChallenge (« DPChallenge - A Digital Photography Contest », s. d.) est une communauté de photographes et une plateforme de compétition qui vise à promouvoir l'art de la photographie et les compétences des photographes. En participant au concours, les photographes peuvent élargir leurs horizons et améliorer leurs compétences créatives. Le concours sera évalué par les pairs sur une échelle de 1 à 10 pour la qualité globale du travail.

Flickr (« flickr », 2024) est une plateforme polyvalente de partage de photos qui permet non seulement de stocker et de partager des photos, mais aussi de créer une communauté photographique active qui permet aux utilisateurs d'apprendre les uns des autres et de s'inspirer mutuellement. Flickr permet aux utilisateurs d'ajouter des tags à leurs photos, ce qui rend les images plus facilement consultables par d'autres personnes. Elle dispose d'une fonction de recherche puissante qui vous permet de rechercher des photos en fonction des tags, de l'heure, du lieu, etc.

FIGURE 4.6 – Méthodes d'évaluation différentes pour les deux ensembles de données. Dans l'AVA, les notes esthétiques sont données sur une échelle de 1 à 10. L'AADB évalue en émettant des jugements binaires sur 8 attributs (Kong et al., 2016 ; Murray et al., 2012).

AVA (Murray et al., 2012) La base de données AVA (Aesthetic Visual Analysis) est

une grande base de données largement utilisée pour l'évaluation de la qualité esthétique des images. La base de données contient plus de 250 000 images, chacune étant associée à un score esthétique, à des étiquettes de classification dans plus de 60 catégories et à des étiquettes liées au style photographique (voir la figure 4.6). Toutes les images de AVA proviennent de DPChallenge.

AADB (Kong et al., 2016) La base de données AADB (Aesthetic and Attributes Database) est similaire à AVA, chaque image dans AADB est accompagnée d'une évaluation de la qualité esthétique par un évaluateur humain, et chaque image contient 8 attributs esthétiques et une valeur binaire pour cet attribut (bon ou mauvais).

FIGURE 4.7 – Exemples de AMD-A. S indique la note globale, C indique la note de couleur, L indique la note de lumière et CM indique la note de composition (X. Jin et al., 2022).

AMD-A (X. Jin et al., 2022) AMD-A (Aesthetic mixed dataset with attributes) contient 16 924 images, Chaque image contient un score esthétique global et trois scores d'attributs esthétiques, lumière, couleur et composition (voir la figure 4.7).

PCCD (Kuang-Yu Chang et al., 2017) PCCD (Photo Critique Captioning Dataset) contient 4 235 images, chaque image inclut un note global et les commentaires linguistiques sur les attributes.

DPC-Captions (X. Jin et al., 2019) DPC-Captions (DPC est tiré de DPChallenge.com) contient 154 384 images. Chaque image inclut des commentaires linguistiques pour un maximum de cinq attributs esthétiques d'images.

RPCD (Nieto et al., 2022) RPCD (Reddit Photo Critique Dataset) contient 73 965 images à haute résolution (2993×2716 pixels en moyenne). Chaque image contient une note esthétique globale et un paragraphe de commentaires informatifs.

BAID (Yi et al., 2023) BAID (Boldbrush Artistic Image Dataset) est une base de données dédiée à l'évaluation de la qualité esthétique des images d'art, qui contient 60 337 images d'art. La plupart des autres base de données pour l'évaluation de la

qualité esthétique des images concernent des images photographiques et n'incluent pas beaucoup d'images de créations artistiques. BAID comble cette lacune (voir la figure 4.8).

FIGURE 4.8 – Exemples de BAID. Note de 0 à 10, plus la note est élevée, plus la qualité esthétique est grande (Yi et al., 2023).

D'autres bases de données pour l'évaluation esthétique des images, telles que CUHK (Yan Ke et al., 2006), CUHKPQ (Wei Luo et al., 2011), LIVE (Ghadiyaram et al., 2016), ICAA17K (S. He et al., 2023), TID2013 (Ponomarenko et al., 2013) et EVA (Kang et al., 2020), peut également être appliquée pour résoudre des problèmes spécifiques, mais elles ne contiennent pas assez d'informations sur la notation esthétique par rapport à AVA et AADB. La richesse des bases de données d'images offre la possibilité d'appliquer des méthodes d'apprentissage automatique à la prédiction de la qualité esthétique des images. Cependant, ces bases de données sont généralement des images LDR, qui sont des données non linéaires avec une gamme dynamique limitée. Les caractéristiques esthétiques qu'elles contiennent ne peuvent pas être directement utilisées pour analyser la qualité esthétique des images HDR. Nous introduisons donc un dataset destiné à l'analyse des images HDR dans le chapitre 3. Puis au chapitre 5, nous proposons une méthode de reconstruction des lignes de force, qui est applicable à la fois aux images LDR et aux images HDR, pour analyser les caractéristiques esthétique des images.

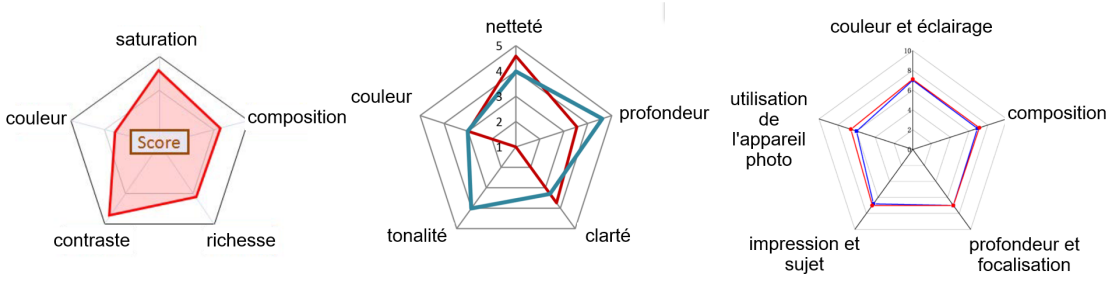
4.4 Analyse et renforcement esthétique de l'image

4.4.1 Analyse des caractéristiques esthétiques

En discutant de la qualité esthétique des images, on fait souvent l'hypothèse que la notion d'esthétique est un sentiment subjectif qui ne peut pas être facilement quantifié. Cependant, l'image elle-même, en tant qu'objet mathématique, possède ses propres caractéristiques, tels que la composition, la couleur et la lumière. Les modifications de ces attributs affectent directement la qualité esthétique d'une image. Nous espérons donc contribuer à comprendre la perception de l'esthétique en analysant les attributs inhérents à une image.

Dans la conception moderne de l'esthétique, les objets qui possèdent certaines caractéristiques intrinsèques admises par les observateurs sont plus susceptibles d'être considérés comme beaux. Toutefois, le jugement final reste basé sur la préférence subjective de l'individu (Valenzise et al., 2022). Par conséquent, dans l'analyse esthétique des images, nous souhaitons expliquer la perception de la qualité esthétique en analysant les caractéristiques des images. Nous essayons également de déterminer l'influence de ces caractéristiques sur le jugement final de la qualité esthétique.

La recherche de Kong et al. (2016) prend en compte les caractéristiques de l'image dans l'étude en enregistrant les évaluations de huit caractéristiques de l'image. La prédiction de la qualité esthétique globale a été guidée par les caractéristiques pertinentes des images que sont la couleur, la lumière, la composition et le contenu. De plus, Kang et al. (2020) résument les propriétés de l'image en termes de « lumière et couleur », de « composition et profondeur », de « qualité technique » et de « sémantique ». Cette recherche examine la difficulté d'évaluer la qualité esthétique de différentes images et analyse l'importance des attributs de l'image dans l'évaluation de la qualité esthétique. Les analyses expérimentales ont confirmé l'existence d'une relation linéaire entre les caractéristiques des images et la qualité esthétique.



(a) (b) (c)

FIGURE 4.9 – Diagrammes radar pour les caractéristiques de l'image (T. O. Aydin et al., 2015 ; X. Jin et al., 2019 ; Lo et al., 2013).

Les caractéristiques de l'image sont multidimensionnelle et les diagrammes radar sont utilisés dans les recherches pour démontrer l'analyse des caractéristiques de l'image. Lo et al. (2013) conçoivent une application intelligente pour analyser les scores perceptuels d'une image sur cinq caractéristiques et donner un score esthétique global au centre du diagramme radar (voir la figure 4.9 (a)). Le diagramme radar (voir la figure 4.9 (b)) donné par T. O. Aydin et al. (2015) montre les changements dans les scores des caractéristiques des images avant et après l'édition, et l'analyse peut également être utilisée pour analyser les résultats de Tone Mapping des images HDR. En comparant les

résultats de différents TMOs, nous pouvons observer la similitude de certaines opérations de différents TMOs dans le traitement des caractéristiques de la même image HDR. La figure 4.9 (c) est un diagramme radar donné par X. Jin et al. (2019), qui divise les mots-clés d'une étiquette d'image en cinq grandes catégories de caractéristiques de l'image et analyse les scores esthétiques et les descriptions textuelles correspondants.

D'autres méthodes d'analyse esthétique basées sur des caractéristiques individuelles de l'image existent, telles que l'analyse de l'harmonie des couleurs (Nishiyama et al., 2011), l'analyse du contenu de l'image (Dhar et al., 2011 ; Wei Luo et al., 2011), etc., comme mentionné dans la section 4.2. Du point de vue de l'influence sur la perception visuelle, nous pouvons analyser l'esthétique des images en trois catégories principales : la lumière, la couleur et la composition.

Lumière

La lumière est l'une des principales caractéristiques esthétiques pour les photographes et les cinéastes (B. Brown, 2016 ; Wisslar, 2012). En tant que premier élément déclencheur de la perception visuelle, sans lumière, il n'y aurait pas de perception de la couleur, de la composition ou d'autres caractéristiques esthétiques (Arnheim, 2004). Les caractéristiques liées à la lumière doivent décrire la lumière dans une image : l'image est-elle claire ou sombre, le contraste est-il faible ou élevé ? Dans la recherche de Bist et al. (2017), les styles de lumière sont classés en cinq catégories en fonction de la luminance et du contraste de l'image : sombre, basse, moyenne, claire ou haute (voir la figure 4.10). La sombre représente un éclairage à faible contraste et à faible luminance. Elle est souvent utilisée dans les films d'horreur ou de meurtre pour créer la peur ou le suspense. Cela peut créer des sentiments de peur ou de suspense. La basse est généralement caractérisée par des éclairages à fort contraste et à faible luminance. Ce type d'éclairage peut créer une ambiance tendue et intense. La moyenne a un éclairage avec un contraste moyen et une luminance modérée. Elle est plus quotidien et a un plus large éventail d'applications. La claire possède un éclairage très contrasté et une forte luminance. Ce style est plus courant dans les scènes de nature en extérieur. La haute comporte un éclairage avec un faible contraste et une forte luminance. Cette combinaison peut être utilisée pour transmettre une ambiance détendue et douce.

Par conséquent, différents choix de styles de lumière peuvent aider les auteurs à transmettre différents messages d'ambiance. Les changements dans la gamme dynamique de la luminance au sein d'une image peuvent entraîner une modification du style de lumière, apportant ainsi une perception visuelle différente aux spectateurs. En particulier, lors du traitement d'images à grande gamme dynamique, les ajustements liés à la lumière, tels que l'exposition et la distribution de la lumière, ont un impact plus marqué sur

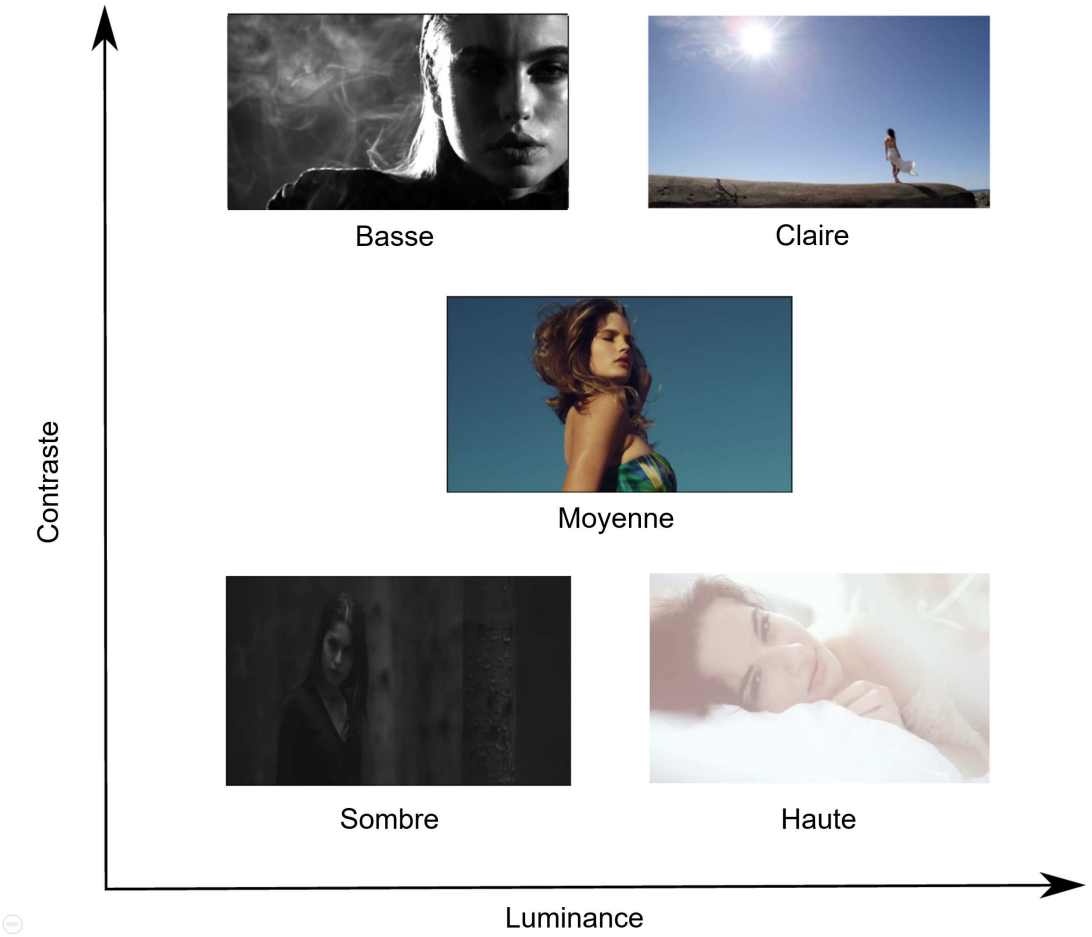


FIGURE 4.10 – Cinq styles de lumière (Bist et al., 2017).

l'expérience visuelle. Le chapitre 3 en donne une illustration détaillée.

Couleur

L'harmonie des couleurs est souvent utilisée dans l'analyse esthétique des images. Des schémas de couleurs courants tels que les couleurs complémentaires, les couleurs analogues et les couleurs triade [1] sont également fréquemment utilisés dans l'évaluation de la qualité esthétique (Arnheim, 2004 ; S. He et al., 2023 ; Shamoi et al., 2022) et l'amélioration de la qualité des images (Cohen-Or et al., 2006). En analysant les couleurs contenues dans une image, nous pouvons générer une palette d'images. L'utilisation d'une palette de couleurs peut nous aider à comprendre et à ajuster l'harmonie des

1. cette technique consiste à choisir trois couleurs dans la roue chromatique de façon à ce qu'elle forment un triangle équilatéral.

couleurs de l'image (Chang et al., 2015). La palette de couleurs nous révèle les règles de l'harmonie des couleurs de l'image, et différentes règles de couleurs peuvent créer des expériences visuelles variées. Comme dans la figure 4.11, l'image de droite utilise une combinaison de couleurs chaudes pour créer une sensation de douceur et de confort, tandis que celle de gauche avec sa palette de couleurs froides évoque une esthétique tranquille et profonde. Le choix des couleurs complémentaires au centre met en évidence le contraste entre mouvement et calme dans l'image.

FIGURE 4.11 – Exemple de palette de couleurs. L'image de gauche montre une image composée de couleurs chaudes analogues, celle du milieu une image composée de couleurs complémentaires et celle de droite une image composée de couleurs froides analogues.

Composition

La composition décrit la disposition générale des éléments d'une image, montrant les relations entre différents éléments. En tant que composante importante des caractéristiques esthétiques, la composition peut affecter directement la perception esthétique d'une image. Les principes de composition les plus couramment utilisés sont la règle des tiers, la règle d'or et la règle diagonale, qui sont toutes ajustées par la composition pour renforcer le centre visuel et influencer la perception d'une image.

L'ajustement de la composition d'une image peut améliorer sa qualité esthétique. La recherche menée par F.-L. Zhang et al. (2013) vise à améliorer la qualité esthétique des images en réarrangeant les objets au premier plan dans les photographies. Cette étude commence par analyser la sémantique et les relations de dépendance des objets dans l'image, puis déplace ces objets et leurs zones associées vers des emplacements qui correspondent le mieux à la règle des tiers ou à la composition en diagonale. La figure 4.12 illustre les résultats d'amélioration.

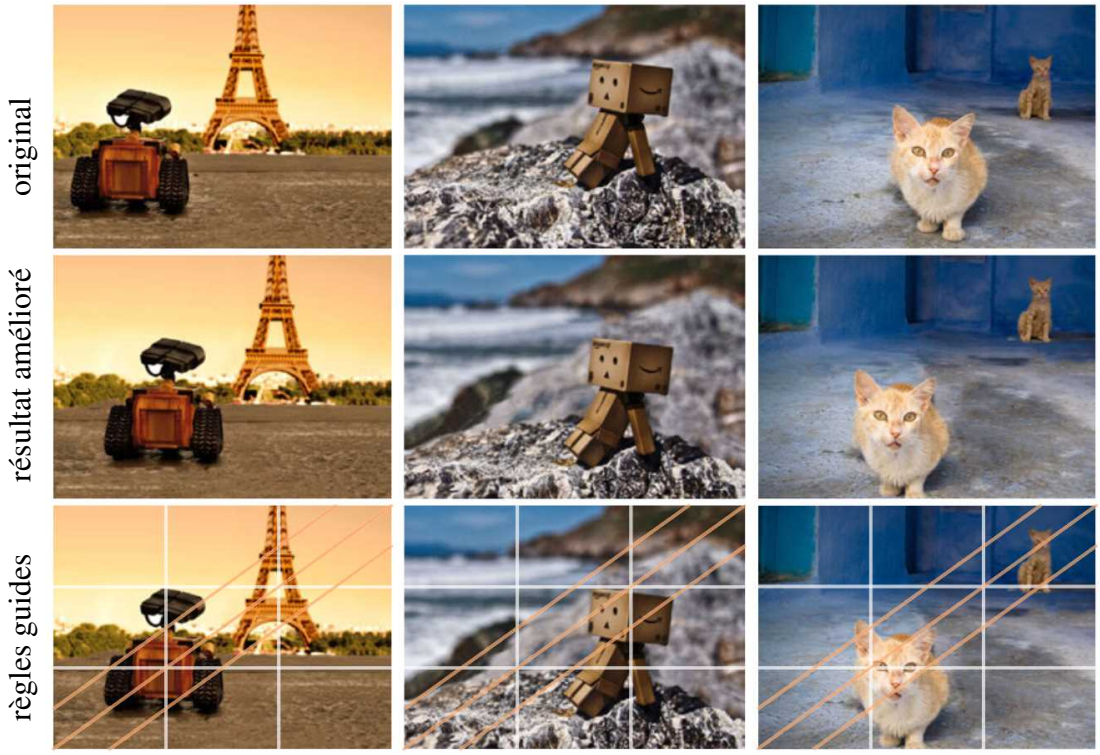


FIGURE 4.12 – Exemples d'améliorations de la composition (F.-L. Zhang et al., 2013). Les lignes blanches sont basées sur la règle des tiers ; les lignes orange sont basées sur le diagonal.

L'analyse de la composition des images peut également appliquer des méthodes d'apprentissage basées sur les données, en utilisant des réseaux d'apprentissage profond pour évaluer les règles de composition des images. Par exemple, le modèle proposé par Celona et al. (2021) prédit les étiquettes de composition déjà présentes dans les données, telles que centrée, horizontale, diagonale, etc. B. Zhang et al. (2021) proposent une méthode qui divise la composition en huit modèles principaux, puis prédit les modèles de composition de l'image en combinant les caractéristiques potentielles analysées par le réseau d'apprentissage en profondeur et la carte de saillance de l'image. Il existe

également de nombreuses méthodes pour améliorer les résultats du recadrage d'une image grâce à l'analyse de la composition (Abeln et al., 2016 ; Li et al., 2018 ; W. Wang et al., 2017).

Tout en analysant les règles de composition des images, nous souhaitons également assister l'analyse de la composition en marquant explicitement certaines formes ou lignes de l'image. Lorsque nous apprécions une image, l'auteur utilise des lignes conçues pour guider notre chemin d'interprétation de l'image. Ces lignes sont des lignes de force de la composition, elles ne sont pas nécessairement des lignes physiquement visibles, et nous espérons mieux comprendre la composition de l'image en identifiant ces lignes. Dans le chapitre 5, nous avons mené une étude détaillée sur les lignes de force.

4.4.2 Renforcement de la qualité esthétique

L'analyse des caractéristiques esthétiques des images peut nous aider à comprendre comment les caractéristiques objectives des images affectent l'expérience visuelle. En outre, elle peut aussi aider les ordinateurs à effectuer une sélection automatique des images, comme la sélection des couvertures d'albums de téléphones portables et le classement des résultats de recherche d'images. La compréhension des caractéristiques esthétiques peut également aider les algorithmes à renforcer l'esthétique des images. Par exemple, EnhanceGAN (Deng et al., 2018) ajoute une évaluation esthétique au discriminateur du réseau de neurones GAN, ce qui permet d'utiliser les caractéristiques esthétiques pour guider l'amélioration de l'image.

Le travail de Bychkovsky et al. (2011) nous permet de constater l'impact des préférences individuelles pour les caractéristiques de l'image sur les résultats finaux de renforcement de l'image. Par conséquent, lors de la retouche d'image, certains méthodes (Bianco et al., 2020 ; Fischer et al., 2020) prennent en compte le choix subjectif des caractéristiques de l'image.

Dans le cadre de cette thèse, nous espérons également mieux comprendre la perception visuelle en analysant les caractéristiques esthétiques des images : comprendre la composition d'une image à travers ses lignes de force implicites (le chapitre 5), ajuster les courbes tonales en analysant l'histogramme de l'image (la section 3.2) ou encore extraire la caractéristique de profondeur de l'image pour ajuster l'exposition (la section 3.3).

Résumé

L'essence de la recherche en esthétique de l'image est d'étudier la relation entre les informations numérisées sur les images et l'expérience visuelle humaine. Dans les premiers modèles d'esthétique informatique, la qualité esthétique des images était analysée par l'extraction de caractéristiques esthétiques interprétables. Les réseaux d'apprentissage profond analysent et prédisent la qualité esthétique des images en calculant des caractéristiques latentes profondes. Pour différents objectifs d'analyse esthétique, les chercheurs ont établi une variété de jeux de données afin de soutenir l'étude de l'esthétique des images. Comprendre les caractéristiques esthétiques des images nous aide à obtenir une expérience visuelle de meilleure qualité. Dans le chapitre 3, nous avons extrait les caractéristiques latentes des images à haute gamme dynamique à l'aide d'un modèle de réseaux de neurones et nous avons utilisé ces caractéristiques pour améliorer l'effet d'affichage. Dans le chapitre 5,nous souhaitons comprendre les caractéristiques esthétiques des images à travers une méthode explicite et interprétable de reconstruction des lignes de force.

Chapitre 5

Reconstruction de la composition d'une image : calcul des lignes de force

Introduction

La composition est une caractéristique importante qui affecte la perception esthétique d'une image. Dans ce chapitre, nous présenterons les lignes de force de la composition de l'image. Nous présentons d'abord le concept des lignes de force, puis nous démontrons expérimentalement la possibilité de définir des lignes de force explicites. Ensuite, nous proposons un algorithme pour reconstruire automatiquement les lignes de force d'une composition d'image. En même temps, nous proposons un critère pour comparer la similarité de deux ensembles de lignes.

5.1 Introduction des lignes de force

Les images servent de moyen pour raconter des histoires. Joly (2015) propose que, mis à part la scène représentée et sa mise en scène, divers éléments esthétiques et artistiques aident les créateurs à transmettre le message qu'ils ont voulu. Ces éléments englobent des aspects tels que la palette de couleurs, l'ambiance de l'éclairage, le cadrage, le point de vue et la composition. Par conséquent, la reconstruction de ces composants esthétiques peut faciliter l'analyse et la compréhension de l'image. La reconstruction de caractéristiques esthétiques spécifiques a fait l'objet de nombreuses publications, y compris le calcul d'une palette de couleurs (Cohen-Or et al., 2006 ; Y. Wang et al., 2019), qui décrit les couleurs primaires d'une image, et la détermination du style d'éclairage (Bist et al., 2016) associé à l'esthétique d'une image. En ce qui concerne les modèles de composition d'image, le principal défi découle des diverses approches utilisées pour décrire la composition de l'image. Certains auteurs utilisent des modèles basés sur des formes (Bang, 2000), décrivant la composition en termes de formes primaires ou leur

FIGURE 5.1 – Lignes de force. Notre méthode calcule automatiquement les lignes de force probables qui sous-tendent la composition de l'image. Selon la complexité de la composition de l'image, notre méthode peut reconstruire soit une seule ligne de force, soit plusieurs lignes de force. Le haut représente l'image originale et le bas est le résultat de reconstruction.

juxtaposition, tandis que d'autres utilisent des modèles basés sur des lignes (Dykinga, 2014). Les modèles de composition d'image basés sur la forme intègrent également des lignes de force pour élucider la disposition des formes (voir la figure 5.1), fournissant une justification esthétique de ces lignes de force. Par exemple, Molly Bang (Bang, 2000) explique que les lignes directrices diagonales ascendantes impliquent un mouvement ou une tension.

La première recherche sur l'évaluation de l'esthétique de l'image utilisant des caractéristiques conçues à la main a souligné le rôle central de la composition dans l'évaluation de l'esthétique. Dans ces études, les auteurs ne tentent pas de reconstruire les lignes de force de la composition, mais s'appuient plutôt sur des règles de composition classiques, telles que la règle des tiers ou des diagonales vers le haut et vers le bas (Freeman, 2007). Ils évaluent l'alignement des objets principaux le long des lignes verticales et horizontales qui divisent l'image en tiers (Kong et al., 2016 ; Li et al., 2018). Bien que la détection des lignes de force soit cruciale pour la compréhension du contenu de l'image et l'évaluation esthétique, il n'existe actuellement aucune méthode dédiée pour récupérer ces lignes. Debnath et al. (2022) reconnaissent l'impact des lignes de force sur les scores esthétiques et proposent un réseau de neurones convolutifs (CNN) pour la reconnaissance des lignes de force. Cependant, leur méthode estime simplement l'existence de lignes de force évidentes dans une image sans préciser leurs emplacements précis. La règle des tiers est également largement utilisée dans le recadrage automatisé d'images (Mai et al., 2016 ; W. Wang et al., 2017).

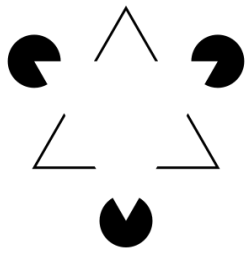


FIGURE 5.2 – La figure de Kanizsa est une illusion d'optique qui démontre comment notre cerveau perçoit les lignes invisibles. Même s'il n'y a pas de lignes réelles reliant certaines formes, notre cerveau crée la perception des lignes cachées en raison de notre capacité innée à combler les lacunes et à interpréter des informations incomplètes.

Le rôle des lignes dans le système visuel humain est fondamental pour percevoir et interpréter le monde environnant. Les lignes sont des éléments visuels essentiels que le système perceptif utilise pour construire des formes, des objets et des scènes. Ce processus de perception visuelle et d'organisation a été montré par la psychologie de la Gestalt, qui fournit un contexte historique pour notre compréhension de la façon dont le cerveau traite l'information visuelle (Abbasov, 2021). L'importance des lignes dans la perception est évidente dans le concept de « contours illusoires », où les bords, les lignes ou les formes semblent exister dans une scène visuelle même lorsqu'ils ne sont pas physiquement présents dans le stimulus. En d'autres termes, notre système visuel remplit les informations manquantes pour créer l'illusion de contours ou de frontières qui ne sont pas physiquement présents. L'un des exemples les plus connus est la figure de Kanizsa ou « configuration Pac-Man » (voir la figure 5.2). Des recherches récentes ont indiqué que certaines structures neuroanatomiques sont spécialisées dans la détection des lignes. La perception de la ligne implique à la fois un traitement de niveau inférieur dans le cortex visuel primaire (V1) et un traitement de niveau supérieur dans les zones du cortex d'association visuelle (González-Casillas et al., 2018).

Les lignes de force et les lignes sémantiques sont deux concepts utilisés dans l'analyse d'images, mais ils ont des fonctions et des objectifs différents, bien qu'ils partagent certains aspects. Les lignes de force sont des éléments de composition visuelle qui guident le regard du spectateur à travers une image. Elles sont souvent utilisées pour créer un équilibre ou une tension dans une œuvre d'art ou une photographie. Les lignes de force peuvent être des lignes réelles, telles que des routes ou des chemins, ou des lignes imaginaires créées par des motifs, des ombres ou des contrastes. Elles servent principalement à orienter l'œil du spectateur et à créer un équilibre visuelle.

D'un autre côté, les lignes sémantiques sont des éléments qui délimitent les zones d'une image en fonction de leur contenu sémantique (Lee et al., 2017), c'est-à-dire leur signification ou leur catégorie. Par exemple, la ligne d'horizon dans une photographie sépare souvent le ciel de la terre, deux catégories distinctes. Les lignes sémantiques sont utilisées pour identifier et délimiter les différentes parties d'une image en fonction de leur contenu et de leur signification.

Les deux sont principalement basés sur des éléments visuels, tels que des lignes et des formes, pour communiquer des informations aux spectateurs. En résumé, bien que les lignes de force et les lignes sémantiques soient toutes deux des concepts importants dans l'analyse d'images, elles servent des objectifs différents : les lignes de force guident la composition visuelle tandis que les lignes sémantiques délimitent les zones en fonction de leur contenu sémantique (voir la figure 5.3).

FIGURE 5.3 – Comparaison entre les lignes sémantiques (représentées par des lignes bleues dans les images de gauche, calculées selon (D. Jin et al., 2022)) et les lignes de force (illustrées par des lignes rouges dans les images de droite) calculées avec notre méthode. Bien que certaines lignes sémantiques puissent également servir de lignes de force, elles sont insuffisantes pour une représentation complète de la composition de l'image. En effet, il existe des lignes de force de composition qui ne s'alignent pas sur les lignes sémantiques.

À notre connaissance, aucune méthode de calcul n'a été introduite pour tracer les lignes de force de l'image. De plus, il n'y a pas de vérité terrain prédéfinie pour les lignes de force de la composition de l'image, car ce sont des caractéristiques dérivées de l'analyse de l'image (Joly, 2015). Le consensus parmi les experts en image dans la définition des grandes lignes de la composition est une question doit être vérifiée, car les interprétations individuelles peuvent varier. Des études subjectives montrent un accord substantiel entre les experts sur la plupart des images, soutenant la faisabilité de

la détection automatisée de la ligne de force.

Dans cette section, nous présentons une méthode qui identifie automatiquement les lignes de force probables dans la composition d'une image (voir la figure 5.1). Notre méthode comprend plusieurs étapes : (1) le calcul de la carte de contraste de l'image ; (2) la génération de lignes de force potentielles pondérées selon la carte de contraste et (3) le regroupement des lignes de force potentielles pour extraire les lignes de force finales.

Nos principales contributions sont les suivantes :

— Tout d'abord, nous montrons qu'il y a un fort consensus parmi les experts sur l'endroit où se trouvent les lignes de force de la composition d'une image. Par conséquent, cela confirme que la modélisation de la composition de l'image par un ensemble de lignes de force est une approche intéressante.

— Deuxièmement, nous proposons un ensemble de données de quarante images et composition sur lesquelles des lignes de force ont été dessinées par quatre experts. L'ensemble de données comprend des photographies, des peintures et des dessins.

— Troisièmement, nous proposons une méthode non supervisée pour calculer la composition probable des lignes de force d'une image. Nous avons conçu la méthode aussi simple que possible pour avoir une méthode compréhensible.

Dans la section 5.2, nous donnons un bref aperçu des méthodes existantes conçues pour détecter les lignes dans les images. L'objectif principal de ces méthodes est la détection des arêtes réelles (présence physique) dans les images plutôt que des lignes sous-jacentes (présence perçue). Dans cette section, nous résumons également les approches liées à la détection des lignes sémantiques. Dans la section 5.3, nous introduisons une mesure pour mesurer la distance entre deux ensembles de lignes et évaluons le consensus parmi les experts dans l'identification de la composition des lignes de force dans les images. Les résultats de cette étude soulignent la validité de la proposition d'un algorithme automatique pour reconstruire les lignes de force de la composition. La section 5.4 explique la méthode utilisée pour calculer les lignes de force de la composition, tandis que la section 5.5 se plonge dans l'analyse des résultats. De plus, nous démontrons que cet algorithme peut être un outil précieux pour guider la capture d'images dans la section 5.6. Enfin, la section 5.7 conclut ce travail et prépare le terrain pour les travaux futurs.

5.2 Travaux connexes aux lignes

De nombreuses méthodes ont été proposées pour la détection des lignes dans les images numériques. La transformée de Hough (Duda et al., 1972) est l'un des algorithmes

les plus couramment utilisés pour la détection de ligne. Il transforme la tâche de détection des lignes droites dans l'espace d'image en celle de détecter des points dans un espace de paramètres. Cependant, la transformation de Hough peut prendre beaucoup de temps. Pour atténuer ce problème, la transformée probabiliste de Hough a été introduite (Kiryati et al., 1991). D'autres approches innovantes, telles que celles présentées dans (Fernandes et al., 2008 ; Princen et al., 1989), appliquent un noyau elliptique-gaussien et une structure pyramidale pour améliorer la transformation originale de Hough. Dans la détection de ligne par la transformée de Hough, l'étape initiale typique implique la détection des bords. Outre les détecteurs de ligne basés sur la transformée de Hough, plusieurs approches ont proposé des détecteurs de segments de ligne basés sur les algorithmes classiques, telles que Suárez et al. (2021), M. Brown et al. (2015), et Grompone Von Gioi et al. (2012).

En outre, il existe des approches basées sur l'apprentissage, telles que celles que l'on trouve dans (Teplyakov et al., 2022), (H. Zhang et al., 2021), (Huang et al., 2018), et (Dai et al., 2022), qui tirent parti des CNNs pour prédire les segments de ligne dans les images. Par exemple, Teplyakov et al. (2022) utilisent une architecture U-net pour prédire les masques de segments [1] et les champs tangents [2], appliquant ensuite un algorithme de regroupement pour les convertir en segments de ligne finaux. D'autre part, H. Zhang et al. (2021) introduisent une représentation des segments de ligne en utilisant le centre, l'angle et la longueur, combinée à une architecture de caractéristiques partagées pour la prédiction des segments de ligne. Les objectifs de ces méthodes sont de récupérer des segments de ligne (caractéristiques locales) dans les images pour extraire la structure filaire des objets ou pour estimer les poses humaines (Baumgartner et al., 2023 ; J. Zhang et al., 2024).

Les lignes de force de la composition ne sont pas nécessairement constituées des contours des objets, nous ne nous appuierons donc pas sur ces approches. Les contours des objets ou les limites des zones sémantiques à grande échelle sont des courbes significatives dans les images qui structurent les images. Le calcul de ces structures géométriques dans une image numérique sans aucune information a priori a conduit à de nombreuses publications. Certains méthodes (Cao et al., 2005 ; Desolneux et al., 2000) ont calculé les limites maximales en utilisant le contraste local et le principe de Helmholtz [3]. La structure à grande échelle donnée par ces méthodes n'est pas l'ensemble des lignes de force de la composition, mais se rapproche d'une esquisse décrivant les limites des objets de l'image.

1. L'ensemble des pixels de la région qui peuvent contenir des segments de ligne.

2. Il est représenté ici comme l'ensemble des tangentes à l'angle entre le segment de ligne et la ligne horizontale.

3. Le principe de Helmholtz indique que tout écart important par rapport à l'hypothèse de l'uniformité aléatoire est perceptible. Ce principe peut être appliqué à la détection des contours dans la vision par ordinateur, en se concentrant sur le fait que les structures significatives dans une image, telles que les contours et les lignes, ont une importance statistique et ne peuvent pas apparaître de manière aléatoire.

Les lignes sémantiques, définies comme les lignes primaires et significatives qui délimitent diverses régions sémantiques au sein d'une image (Lee et al., 2017), ont également reçu une attention importante. Le travail de Lee et al. (2017) a marqué le début de la détection des lignes sémantiques, avec l'introduction d'un apprentissage multitâche CNN pour prédire les lignes sémantiquement importantes. Pour faciliter l'apprentissage du réseau, ils ont assemblé un ensemble de données de lignes sémantiques comprenant 1 750 images et ont utilisé la métrique d'intersection moyenne sur union (mIoU) pour la mesure de la distance de ligne. S'appuyant sur le travail de Lee et al. (2017) et D. Jin et al. (2022), D. Jin et al. (2021) ont conçu un réseau d'harmonisation et un graphique complet pour déterminer les lignes sémantiques finales. Ils ont également introduit une métrique d'intersection sur union (HIoU) basée sur l'harmonie pour mesurer le score global de correspondance de deux ensembles de lignes. En outre, Zhao et al. (2021) ont intégré Hough Transform dans un réseau d'apprentissage profond pour prédire les lignes sémantiques dans l'espace paramétrique, transformant ainsi la détection des lignes en détection de points. Ils ont organisé un ensemble de données contenant 6 500 images à travers diverses scènes et ont proposé une métrique de distance qui prend en compte à la fois les distances euclidiennes et angulaires pour mesurer les distances de ligne. Un fil conducteur entre ces travaux est leur adoption d'approches d'apprentissage supervisé, qui nécessitent la disponibilité d'ensembles de données appropriés.

Cependant, ces méthodes traditionnelles de détection de ligne ne sont pas en mesure de reconstruire les lignes de force de l'image. Les détections de lignes droites se concentrent sur les lignes physiques réelles de l'image, les détections de bord se concentrent sur la représentation des détails des bords de l'objet plutôt que sur une ligne qui a des effets de guidage. Les lignes sémantiques sont approximatives des lignes de force dans des scénarios spéciaux, mais sont incomplètement équivalentes. Les détections de lignes sémantiques mettent l'accent sur la détection des lignes qui segmentent les régions sémantiques d'une image plutôt que sur l'analyse des caractéristiques de composition de l'image. Par conséquent, nous proposons une méthode visant à reconstruire les lignes de force de l'image.

5.3 Lignes de force vers la composition de l'image du modèle

Nous proposons de modéliser la composition de l'image, englobant divers supports visuels tels que la peinture, la gravure, le dessin et la photographie, en utilisant le concept de lignes de force. Dans cette section, la caractérisation complète d'une ligne englobe un

arrangement infini de points, mais sa manifestation sur une image 2D se matérialise sous la forme d'une collection de pixels provenant du périmètre de l'image.

Les lignes de force structurent l'ensemble de l'image, elles ne décrivent pas la structure d'une partie de l'image, par conséquent, une ligne de force traverse l'image : en commençant sur un bord d'image et en terminant sur un autre bord d'image. Les lignes de force structurent le contenu à l'intérieur de l'image tandis que les bords définissent le cadre. Par conséquent, nous supposons que les bords de l'image ne sont pas des lignes de force. L'évaluation de la viabilité de l'utilisation de lignes de force pour modéliser la composition de l'image implique d'examiner le consensus d'experts dans la définition de la ligne de force lors de la récupération de la composition de l'image. Sans vérité de base fournie par l'auteur, l'identification des lignes de force devient subjective, en s'appuyant sur l'interprétation du spectateur.

Nous montrons que l'accord d'experts justifie la détection automatisée des lignes de force. En effet, nous avons mené une expérience subjective où les experts en images ont décrit les grandes lignes en fonction de leur interprétation de la composition et ont comparé leurs similitudes. À cette fin, nous avons adapté la mesure introduite par Zhao et al. (2021) dans la section 5.3.1.

5.3.1 Distance entre deux ensembles de lignes

Dans leur travail Zhao et al. (2021) ont introduit un score de similitude connu sous le nom de score EA (distance Euclidienne et distance Angulaire) pour quantifier la similitude entre deux lignes. Ce score est défini comme suit :

$$\mathcal{S}_{EA}(l_i, l_j) = \left(\left(1 - \frac{\theta(l_i, l_j)}{\pi/2}\right) \times \left(1 - D(l_i, l_j)\right)\right)^2 \tag{5.1}$$

Ici, l_i et l_j dénote deux lignes, et $\theta(l_i, l_j)$ représente l'angle entre ces deux lignes, $D(l_i, l_j)$ représente la distance euclidienne entre les points médians de deux lignes (les deux lignes sont recadrées pour s'adapter au cadre de l'image). Il est essentiel de noter que l'image est d'abord transormée en une image carrée (de ratio 1).

Maintenant, notre objectif est d'évaluer l'accord entre les ensembles de lignes de force de composition définies manuellement. Par conséquent, nous avons besoin d'une mesure pour évaluer la similitude entre deux ensembles de lignes. Tirer parti du score de similitude S_{EA}, notre première étape consiste à définir la distanced d_{LS} entre une ligne l_i et un ensemble $\mathcal{G}$ comprenant N lignes. Nous adoptons la méthode classique de calcul de la distance entre un élément et un ensemble d'éléments en fonction de la distance entre

deux éléments individuels. Ceci est exprimé comme suit :

$$d_{LS}(l_i, \mathcal{G}) = \min_{n \in \{1..N\}} (d_{EA}(l_i, g_n)) \tag{5.2}$$

Ici, $d_{EA}(l_i, l_j) = 1 - \mathcal{S}_{EA}(l_i, l_j)$ et g_n dénote une ligne appartenant à l'ensemble $\mathcal{G}$.

Par la suite, nous définissons la distance entre deux ensembles de lignes de force $\mathcal{D}_{LS}(\mathcal{F}, \mathcal{G})$ comme la moyenne de deux distances moyennes. La première distance calcule la moyenne de $d_{LS}(f_i, G), i = 0, 1, \ldots, N_F$, alors que la deuxième distance calcule la moyenne de $d_{LS}(f_i, F), i = 0, 1, \ldots, N_G$, l'équation pour exprimer cette distance est la suivante :

$$\mathcal{D}_{LS}(\mathcal{F}, \mathcal{G}) = \frac{1}{2}\left(\left(\frac{1}{N_\mathcal{F}} \sum_{i=1}^{N_\mathcal{F}} d_{LS}(f_i, \mathcal{G})\right) + \left(\frac{1}{N_\mathcal{G}} \sum_{i=1}^{N_\mathcal{G}} d_{LS}(g_i, \mathcal{F})\right)\right) \tag{5.3}$$

Dans les équations ci-dessus, $\mathcal{F}$ et $\mathcal{G}$ répresentent des ensembles de $N_\mathcal{F}$ et $N_\mathcal{G}$ lignes de force, respectivement. Cette formulation garantit que nous obtenons une mesure de distance symétrique, ce qui signifie que $\mathcal{D}_{LS}(\mathcal{F}, \mathcal{G})$ est équivalent à $\mathcal{D}_{LS}(\mathcal{G}, \mathcal{F})$, plus de détails à l'annexe B.1.1. Ceci est cohérent avec la logique de comparaison de similitude, lors de la comparaison des résultats des lignes de force dans deux images, l'ordre des images n'affecte pas le résultat de la comparaison.

5.3.2 Étude préliminaire sur les accords d'experts

Notre étude préliminaire visait à évaluer le consensus parmi les experts en création et en analyse d'images concernant les lignes de force de la composition, en tenant compte de la nature subjective de leur perception. Dans l'expérience, quatre experts ont été chargés de délimiter les lignes de force sur un ensemble diversifié de quarante images. À l'aide de la mesure décrite à la section 5.3.1, nous avons évalué quantitativement l'accord entre les experts, qui n'avaient aucune contrainte de temps pour définir les lignes.

La distance médiane entre les lignes de force de la composition délimitées par les experts est relativement faible, ce qui signifie un degré élevé de consensus entre eux. Les distances varient entre $0,03$ et $0,42$, comme l'illustre la figure 5.4. Cependant, sur le côté droit de la figure, nous observons une disparité plus prononcée entre les experts. Cet écart est particulièrement perceptible dans certaines images aux compositions complexes, où même les évaluations des experts peuvent différer (voir la figure 5.5 et la figure 5.6).

À l'exception de onze images, les distances inter-experts sont inférieures ou égales à $0,2$. Pour la moitié des images, les distances sont inférieures à $0,14$. La distance médiane

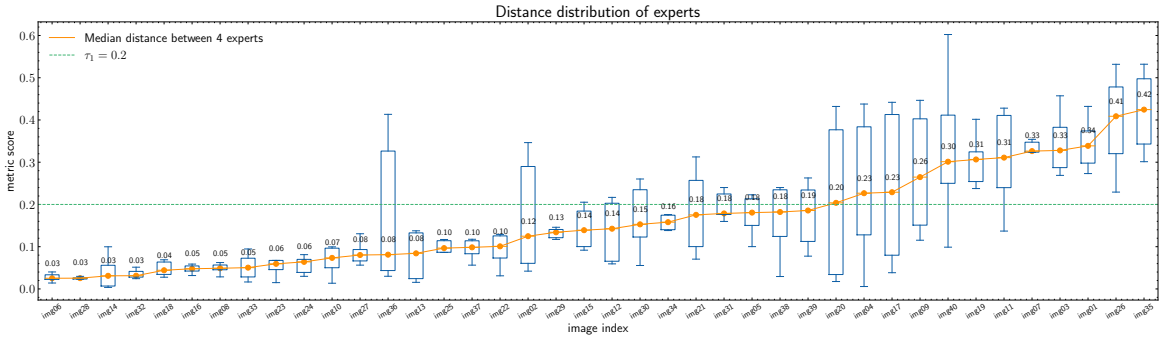


FIGURE 5.4 – Analyse des distances entre les ensembles de lignes de force de la composition tracés par des experts. Les images sur l'axe des x sont disposées par ordre croissant de distance médiane. Chaque image est représentée par un graphique en boîte montrant la distribution des scores de distance entre les experts. La ligne orange dans chaque tracé de boîte représente les distances médianes entre les lignes de force de la composition identifiées par les quatre experts.

dépasse $0,4$ pour seulement deux images (images 26 et 35). Cela suggère un niveau substantiel d'accord entre les experts.

Les résultats de cette étude indiquent un fort consensus parmi les experts lors de l'identification de la composition des lignes de force dans les images. La concordance substantielle dans leurs évaluations met en évidence la fiabilité et la cohérence de leurs jugements. Compte tenu du grand accord entre ces experts, il est justifié de proposer un modèle pour détecter les lignes de force de la composition.

En résumé, les résultats de cette étude préliminaire soutiennent l'idée que les experts font preuve d'un fort consensus dans leur identification de la composition des lignes de force dans les images. Cela ouvre la voie au développement d'un modèle dans ce domaine.

5.4 Méthode : calcul des lignes de force de la composition

Dans cette section, nous présentons un algorithme pour détecter les lignes de force de la composition dans les images. La figure 5.7 donne un aperçu d'approche, qui consiste en deux étapes principales : premièrement, pondérer toutes les lignes de force potentielles, et deuxièmement, regrouper les lignes pour identifier les lignes de force finales.

(a) expert A (b) expert B

(c) expert C (d) expert D

FIGURE 5.5 – Meilleur accord entre les experts : les images de a à d montrent les lignes de force définies par chaque expert. Chaque expert a défini une seule ligne de force qui est confondue avec l'horizon, la distance médiane entre quatre ensembles de lignes de force est de $0,03$.

La première étape comprend trois composantes :

1. Nous utilisons l'interpolation spline pour redimensionner l'image de sa résolution d'origine à $s \times s$. Ce redimensionnement conserve la structure générale de l'image.
2. Nous calculons une carte de contraste de l'image redimensionnée.
3. Nous attribuons des pondérations à toutes les lignes potentielles en fonction de la somme des valeurs de contraste pour chaque pixel croisé.

Dans la deuxième étape, nous avons développé un algorithme de regroupement pour déterminer les lignes de force finales à partir de l'ensemble des lignes potentielles. Chaque itération implique de regrouper des lignes de force potentielles étroitement positionnées pour former des groupes. La ligne centrale de chaque groupe devient l'une des dernières lignes de force.

(a) expert A (b) expert B (c) expert C (d) expert D

FIGURE 5.6 – Pire accord entre les experts. Les images de a à d montrent les lignes de force définies par chaque expert. Dans ce cas, deux experts ont défini deux lignes de force tandis que les deux autres ont défini 3 lignes de force, la distance médiane entre quatre ensembles de lignes de force est de $0,42$, néanmoins, nous pouvons noter que tous les experts ont tracé une diagonale vers le bas comme ligne de force.

5.4.1 Génération de l'ensemble de lignes potentielles

Nous commençons par générer toutes les lignes de force possibles dans l'image. Une ligne de force commence à partir d'un bord de l'image et se termine par un autre bord. Par conséquent, le nombre total de lignes de force potentielles est donné par : $s^2 \times \frac{4\times3}{2} = s^2 \times 6$. Pour chaque couple $(i, j) \in s \times s$, il produit 3 lignes à partir du pixel i le long de la bordure b_n et se terminant au pixel j le long de la frontière b_m, avec n et m dans $\{0, 1, 2, 3\}$, et $n \neq m$. Comme il y a quatre frontières de départ possibles, chaque couple (i, j) donne 12 lignes. Il est important de noter que le couple (i, j) produit les mêmes lignes que (j, i) (voir la figure 5.8). Pour améliorer la robustesse de nos résultats, nous excluons les lignes très courtes et celles qui sont à proximité des bords de l'image, comme l'illustre la figure 5.8. Cela réduit l'espace de recherche à $(s - \delta)^2 \times 6$. Une valeur typique pour δ est $s/10$.

5.4.2 Carte de contraste : dérivée discrète du gradient L_1 norme

Après avoir généré chaque ligne de force potentielle, nous calculons un poids pour chaque ligne. Un poids plus élevé signifie une plus grande signification visuelle, ce qui indique une plus grande probabilité d'être une ligne de force. Diverses formes de valeurs de contraste de pixels sont utilisées pour détecter les pixels contrastés dans une image (Ming-Ming et al., 2011 ; Yang et al., 2013 ; Yun et al., 2022). Nous proposons de calculer

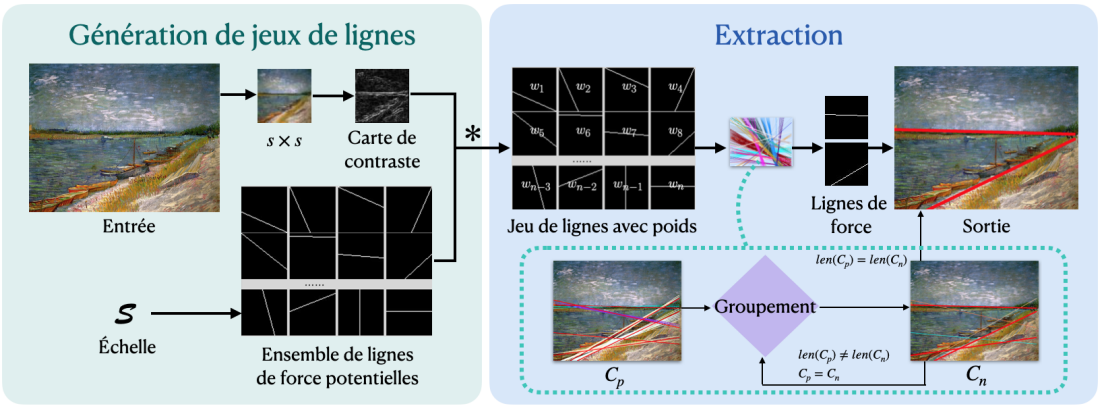


FIGURE 5.7 – Aperçu du pipeline de notre méthode. L'opérateur « * » entre « génération de jeux de lignes » et « extraction » désigne le calcul du poids d'une ligne (voir Équation 5.4).

la carte de contraste $\mathcal{M}_{i,j}$ comme suit :

$$\begin{aligned}\mathcal{M}_{i,j} = &|P_{i,j+1} - P_{i,j}| + |P_{i+1,j-1} - P_{i,j}| \\ &+ |P_{i+1,j} - P_{i,j}| + |P_{i+1,j+1} - P_{i,j}|\end{aligned} \tag{5.4}$$

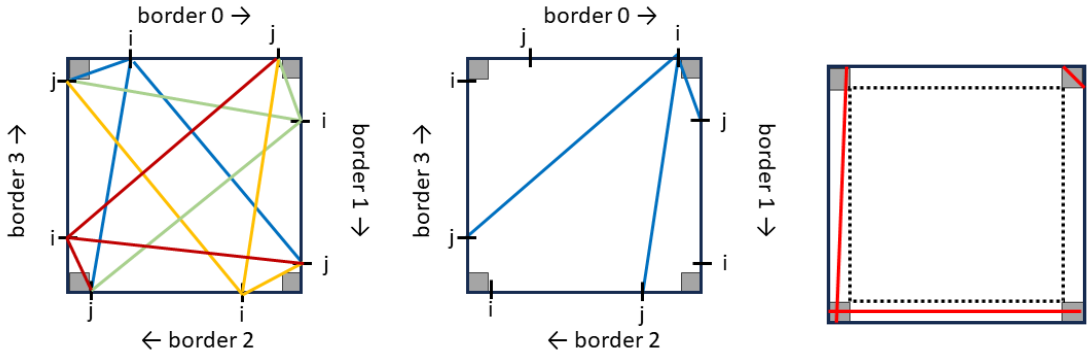
Ici, $P_{i,j}$ représente la valeur du pixel Y de la ligne i et de la colonne j dans le XYZ espace colorimétrique. Notre valeur de contraste ressemble beaucoup à la fois à la dérivée discrète du $L1$ norme du gradient et au Laplacien discret de l'image.

Par la suite, le poids d'une ligne est calculé comme la valeur cumulative de l'élément des valeurs de contraste des pixels :

$$\mathcal{W}_i = \sum_{p \in l_i} \mathcal{M}_p \tag{5.5}$$

Ici, p représente tous les pixels en ligne l_i, et $\mathcal{M}_p$ est la valeur de contraste du pixel p. Comme le montre la figure 5.9.

Nous ne normalisons pas le poids des lignes en fonction de leur longueur pour donner la priorité aux lignes plus longues. En fait, nous supposons que les lignes de force englobent l'ensemble de l'image.



(a) Nous générons 12 lignes de chaque couple (i, j), où i est la coordonnée du point de départ à la frontière k et j la coordonnée de fin à la frontière $l \neq k$. Le couple (i, j) est en $s \times s$. Notez que le couple (j, i) produit le même ensemble de lignes.

(b) Les lignes proches des bords, indiquées en rouge, ne sont pas considérées comme des lignes de force potentielles.

FIGURE 5.8 – Génération de toutes les lignes de force possibles.

5.4.3 Extraction de lignes de force

Après avoir obtenu toutes les lignes de force potentielles pondérées, nous procédons à un processus d'extraction (voir Algorithm 1) pour identifier les lignes de force finales. Les lignes de force finales sont représentées par la ligne centrale de chaque groupe. L'algorithme d'extraction fonctionne de manière itérative. À chaque itération, un nouvel ensemble de groupes est généré, avec un nombre de groupes égal ou inférieur à celui de l'itération précédente. Le processus se termine lorsque le nombre de groupes reste inchangé entre deux itérations consécutives. Chaque itération suit les étapes suivantes :

- une itération est commencée avec un ensemble de groupes de lignes : $\mathcal{C} = \{C^i\}$ avec $C^i = \left(l^i, W^i, \left[l^i_j\right]\right)$. i désigne l'index du groupe alors que j désigne l'index de la ligne. Chaque groupe est défini par :
  - sa ligne centrale l^i est la ligne du groupe dont le poids est le poids médian des lignes l^i_j appartenant au groupe.
  - les lignes l^i_j appartenant au groupe, est également noté : $l_j \in C^i$
  - poids W^i est égal au poids maximal des poids des lignes de force dans le groupe : $W^i = \max_{l_k \in C_i} W_k$
- un nouvel ensemble de lignes $\mathcal{L} = \{l_p\}$ est composé des lignes centrales des groupes et du poids associé. Puis un nouveau groupe ensemble est commencé à un ensemble vide : $\mathcal{C} = \{\}$
- pour chaque ligne l_p, par ordre décroissant de poids W_p, nous évaluons si l_p

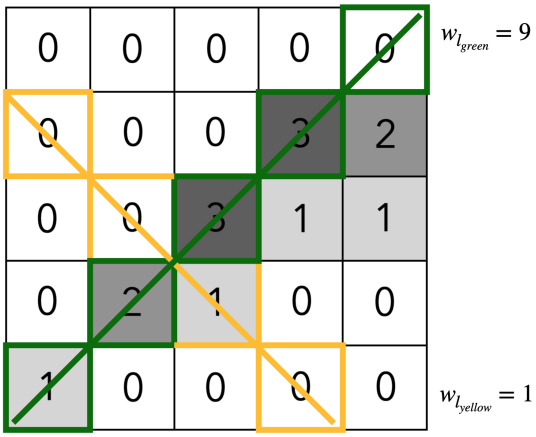


FIGURE 5.9 – Calculer les poids. Le poids d'une ligne est la somme des valeurs de contraste des pixels qu'elle couvre.

appartient à un groupe existant C^q. Les critères d'inclusion pour l_p dans C^q sont les suivants : $(d_{EA}(l_p, l^q) < \delta_d)$ et $(|W_p - W^q| < \delta_W)$. Si les critères sont validés, alors la ligne l_p est mis en C^q, et nous passons à la ligne suivante. Si l_p est dans le cardre C^q, c'est-à-dire $(d_{EA}(l_p, l^q) < \delta_d)$ mais avec $(|W_p - W^q| > \delta_W)$ alors la ligne l_p est jetée et nous passons à la ligne suivante. Enfin, si l_p n'a pas été écartée et n'appartient à aucun groupe existant, un nouveau groupe $C^k = (l^k = l_p, W^k = W_p, [l_p])$ est construit.

— à la fin de l'itération, nous mettons à jour les seuils. La ligne médiane pondérée a été utilisée comme ligne centrale du groupe. Le seuil de distance est affiné en ajoutant $1/\mathcal{S}$, tandis que le seuil de poids est affiné en ajoutant 1.

Le processus d'initialisation de l'algorithme est le suivant : (1) définissez les valeurs initiales pour $\delta_W = 3$ et $\delta_d = 8/s$ où s est la taille en pixel de l'image redimensionnée, et (2) l'ensemble initial de lignes sont toutes les lignes d'attaque possibles calculées à l'étape précédente. En pratique, nous limitons l'ensemble initial de lignes aux deux premiers pour cent des lignes de l'étape précédente, par ordre décroissant, ce qui accélère l'algorithme et n'a pas d'impact sur les résultats finaux. L'algorithme se termine lorsque le nouvel ensemble de groupes est égal à celui de l'itération précédente.

Dans la section suivante, nous présentons quelques résultats, en particulier les lignes de force de la composition calculées par notre algorithme, puis nous nous plongeons dans une analyse détaillée de ces résultats.

Algorithme 1 : groupement de lignes de force

Data : les groupes des lignes de force
$\mathcal{C} = \{C^i\}$ avec $C^i = \left(\text{centralLine} : l^i, \text{weight} : W^i, \text{lines} : \left[l^i_j\right]\right)$

while $\mathcal{C}^{old} \neq \mathcal{C}^{new}$ **do**
  $\{l^i\} \leftarrow$ obtenir les lignes centrales de $\mathcal{C}^{new}$ et les trier
  $\mathcal{C}^{old} \leftarrow \mathcal{C}^{new}$
  $\mathcal{C}^{new} \leftarrow \{\}$
  créer un nouveau groupe $C^0 = \left(\text{centralLine} : l^0, \text{weight} : W^0, \text{lines} : \left[l^0\right]\right)$
  add C^0 to $\mathcal{C}^{new}$
  `/* boucle de lignes */`
  for l^p *in* $\{l^j\}$ **do**
    `/* boucle de lignes */`
    for C^q *in* $\mathcal{C}^{new}$ **do**
      if $(d_{EA}(l^p, l^q) < \delta_d)$ **then**
        if $(|W^p - W^q| < \delta_W)$ **then**
          add l^p to C^q lines
        else
          jeter l^p
        end
      end
    end
    if l^p*ne sont pas mis au rebut et ne sont en aucun cas* $\mathcal{C}^{new}$ **then**
      créer un nouveau C^p : (centralLine : l^p, weight : W^p, lines : $[l^p]$)
      ajouter C^p à $\mathcal{C}_{new}$
    end
  end
  mettre à jour les groupes
  $\delta_d \leftarrow \delta_d + 1$
  $\delta_{weight} \leftarrow \delta_{weight} + 0,5$
end

5.5 Résultats et discussion

Dans cette section, les lignes de force sont d'abord reconstruites à partir de divers types d'images, y compris les peintures et la photographie. Ensuite, la performance de notre modèle est alors évaluée au moyen d'études subjectives. Enfin, nous comparons nos résultats avec les données de vérité terrain.

5.5.1 Résultats

La figure 5.10 présente une sélection de résultats obtenus à l'aide des paramètres d'algorithme suivants : $s = 64$ pixels, $\delta_d = 8/s$, et $\delta_W = 3$. Globalement, les lignes de force de la composition reconstruite semblent correspondre aux observations. Notre algorithme produit en moyenne $2,8$ lignes de force sur un ensemble de données de 40 images, allant d'une seule ligne de force pour des compositions simples à jusqu'à 5 pour les cas plus complexes.

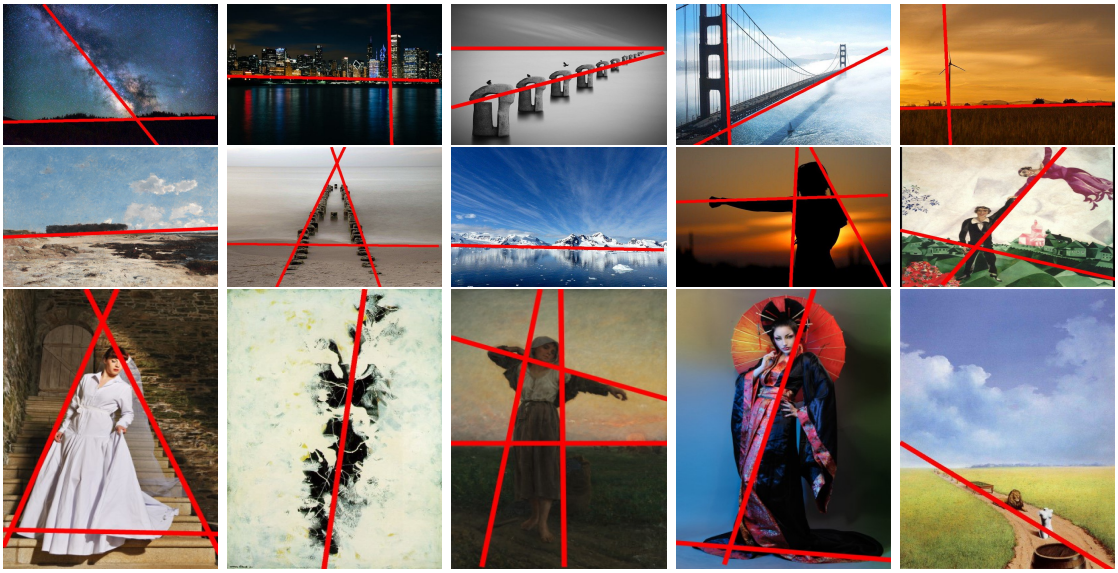


FIGURE 5.10 – Résultats des lignes de force. Les images de test comprennent à la fois des peintures et des photographies de diverses compositions, notre méthode donne des résultats qui paraissent visuellement cohérents avec la perception humaine des lignes de force.

La figure 5.11 présente les lignes sémantiques obtenues à partir des 15 mêmes images en utilisant l'approche de Zhao et al. (2021). Si, dans certains cas, certaines lignes sémantiques s'alignent sur les lignes de force, l'approche de la ligne sémantique ne révèle

pas autant la composition que les lignes de force. Cette observation soutient l'intérêt de l'algorithme de reconstruire les lignes de force de la composition.

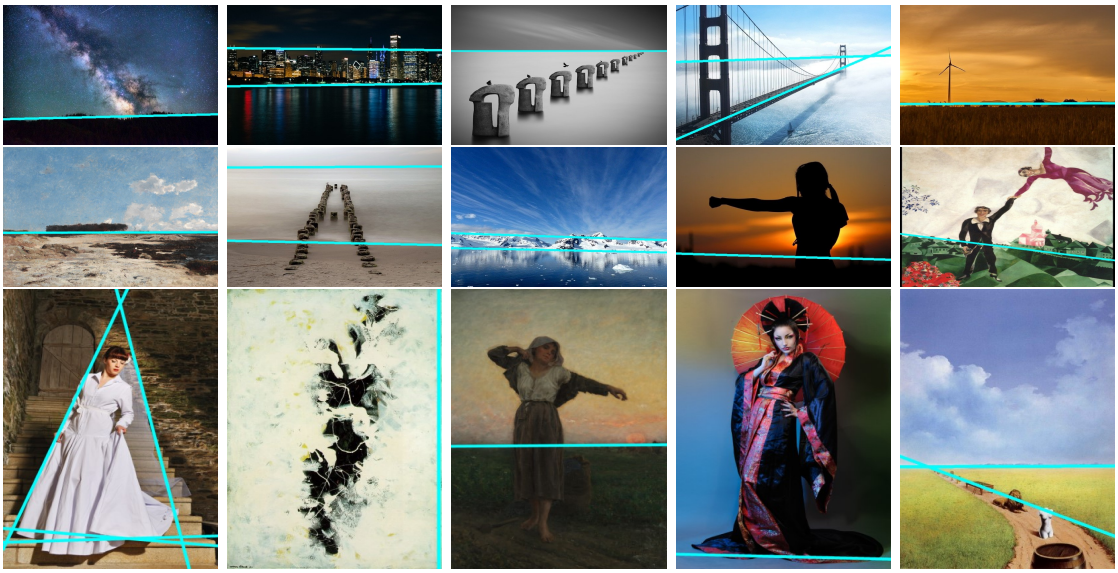


FIGURE 5.11 – Résultats de la détection des lignes sémantiques (Zhao et al., 2021).

5.5.2 Étude perceptive

Dans cette section, nous mesurons les performances du modèle pour les observateurs « naïfs ». Pour ce faire, les lignes de force prédites par le modèle sont présentées à un groupe de sujets sans connaissances spécifiques. Après une brève présentation de ce qui constitue une ligne de force de la composition, les observateurs ont été invités à choisir, par le biais d'une procédure 2AFC procedure (Two Alternative Forced Choice), deux choix forcés alternatifs en Français, les résultats du modèle et une autre version.

Pour évaluer la pertinence du modèle pour décrire les lignes de force, une ensemble de données comprend 40 images diverses sélectionnées pour leur variété. Pour chaque image, trois versions de la ligne de force sont calculées (voir la figure 5.12) : l'une avec la méthode décrite ci-dessus, l'autre avec le même nombre de lignes disposées au hasard sur l'image et la dernière correspond aux lignes choisies par l'un des experts. La version des lignes choisies pour représenter les experts est, pour chaque image, celle qui minimise la distance par rapport aux autres experts. Cette version est considérée comme l'opinion la plus représentative de l'expert pour l'image considérée.

Les comparaisons suivantes sont proposées :

— méthode proposée par rapport à la méthode aléatoire
— méthode proposée par rapport à la version experte.

méthode proposée aléatoire expert

FIGURE 5.12 – Un exemple pour la comparaison.

Au total, quinze volontaires ont participé à l'expérience (11 hommes et 4 femmes). L'âge moyen est $23,4$ années avec un écart-type de $11,1$. Tous les participants ont une vision normale ou une vision corrigée à la normale, quatre des douze portent des lunettes.

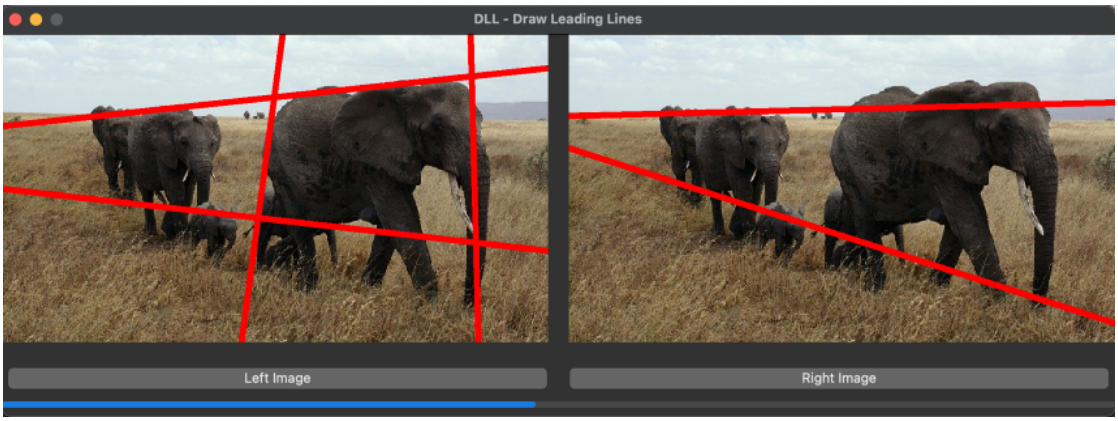


FIGURE 5.13 – Interface pour l'expérience. Les participants doivent choisir l'image pour laquelle le jeu de lignes affichée correspond le plus près possible aux lignes de force de l'image.

Les images sont affichées sur un moniteur d'ordinateur portable standard (résolution Full HD). Les sujets reçoivent une brève présentation de ce que sont les lignes de force avec quelques exemples. On leur demande ensuite de choisir la version qui, selon eux, représente le mieux les lignes de force sur les images, comme le montre la figure 5.13. Afin d'éviter le biais de l'ordre et de la position, l'ordre de présentation du $40 \times 2 = 80$ les images sont aléatoires (ordre et présentation latérale).

TABLEAU 5.1 – Résultats de l'expérience 2AFC.

Condition	Choix du modèle	Autres choix	Pourcentage
méthode proposée v.s. aléatoire	559	41	93,2
méthode proposée v.s. expert	192	408	32,0

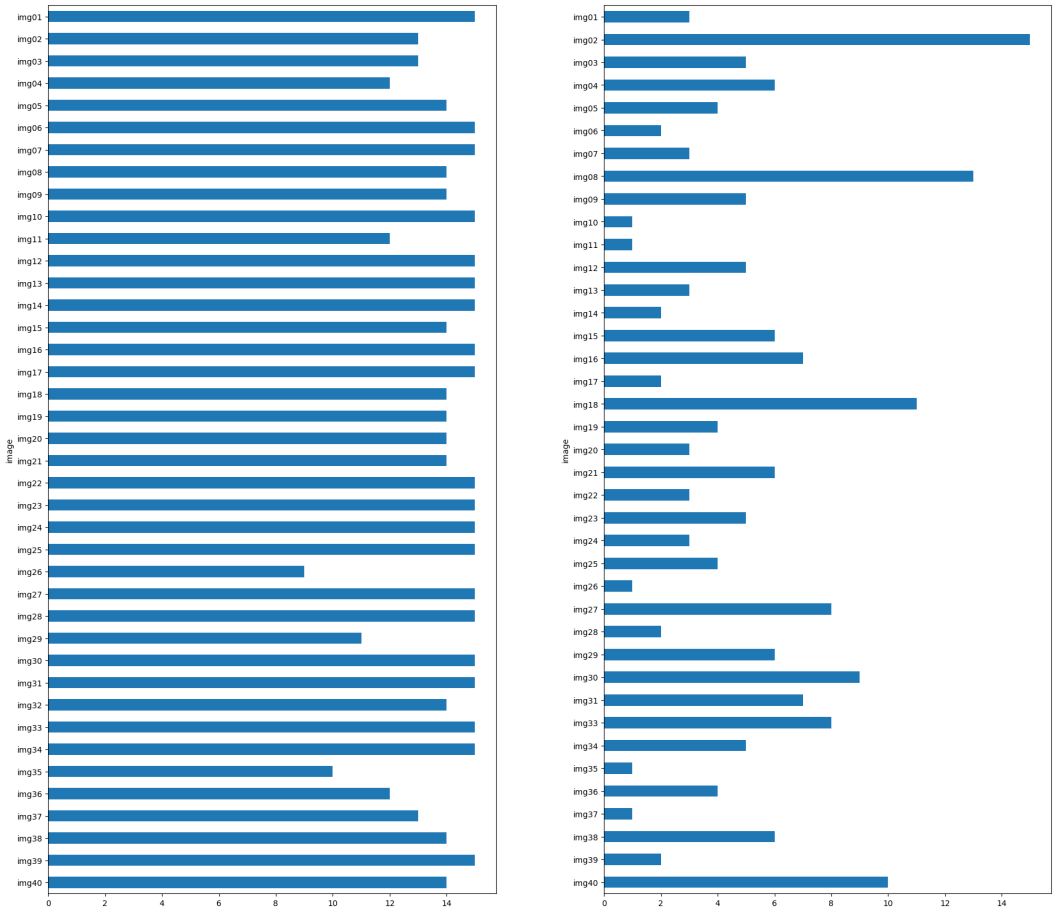
Les résultats de l'expérience sont présentés dans le tableau 5.1 et la figure 5.14. Les participants ont constamment préféré les lignes suggérées par la méthode proposée lorsque qu'elle est confrontée à une distribution aléatoire. D'autre part, lorsque le choix est entre la méthode proposée et la proposition d'un expert, la majorité des sujets ont tendance à choisir la proposition de l'expert, mais la méthode proposée est néanmoins choisie dans environ un tiers des cas. Ces deux effets sont statistiquement significatifs à un seuil d'erreur de 1%. (χ^2 les valeurs sont, respectivement $447,2$ et $77,8$).

Une analyse par image montre que pour toutes les images considérées, le modèle est préféré à une distribution aléatoire. Lorsque le choix est entre les lignes d'un expert et celle du modèle, une grande variation est observée. Bien que la majorité des images choisies sont celles des experts, le modèle est préféré pour certaines images, en particulier les images 2, 8, 18 et 40. Les images 10, 11, 26, 35 et 37 restent problématiques pour le modèle.

Les choix des sujets sont clairement en faveur du modèle par rapport à une distribution aléatoire du même nombre de lignes de force. Par rapport à l'expert le plus représentatif, les choix du modèle sont conservés par les observateurs dans environ un tiers des cas. Cela suggère que les choix de l'algorithme sont de bonne qualité, bien qu'ils ne correspondent pas à la précision des experts. Dans certaines images, le modèle semble même donner une « opinion » que les experts n'avaient pas prise en compte.

5.5.3 Comparaison du modèle avec la vérité terrain

Dans la section 5.3.2, nous avons établi la cohérence des marquages subjectifs entre quatre experts. Par conséquent, nous avons utilisé les résultats des quatre experts comme référence pour valider la cohérence de la sortie de l'algorithme avec les résultats manuels. Nous avons introduit les résultats de l'algorithme en tant que cinquième ensemble de marqueurs experts et avons utilisé notre mesure d'ensemble de lignes pour calculer les distances entre ces cinq ensembles. Ces valeurs de distance constituaient une matrice symétrique 5×5, avec des zéros le long de la diagonale. En observant cette matrice, nous avons noté que lorsque la distance entre deux ensembles était inférieure à $0,2$, comme le montre la figure 5.15 (a), il n'y avait pas de disparités visuelles significatives



(a) (b)

FIGURE 5.14 – Résultats de l'étude subjective. (a) Choix du modèle par rapport à la distribution aléatoire. (b) Choix du modèle par rapport à l'expert le plus représentatif.

entre elles. Lorsque la distance se situait entre $0,2$ et $0,3$, comme l'illustre la figure 5.15 (b), leurs dispositions étaient similaires, bien qu'avec de légères différences de position ou quantitatives. Par conséquent, des distances inférieures à $0,3$ impliquent que les deux ensembles pouvaient être considérés comme cohérents. En revanche, les distances comprises entre $0,3$ et $0,4$, comme le montre la figure 5.15 (d), ont indiqué des différences dans la disposition globale entre l'expert B et l'expert C. Bien que les deux ensembles de lignes variaient dans leur structure globale, certaines lignes avaient des positions similaires. Si la distance entre deux ensembles dépassait $0,4$, comme le montre la figure 5.15 (c), les résultats entre l'algorithme et les experts présentaient des différences significatives à la fois dans la disposition globale et dans les positions des lignes individuelles. Des exemples des quatre cas sont illustrés à la figure 5.16. Par

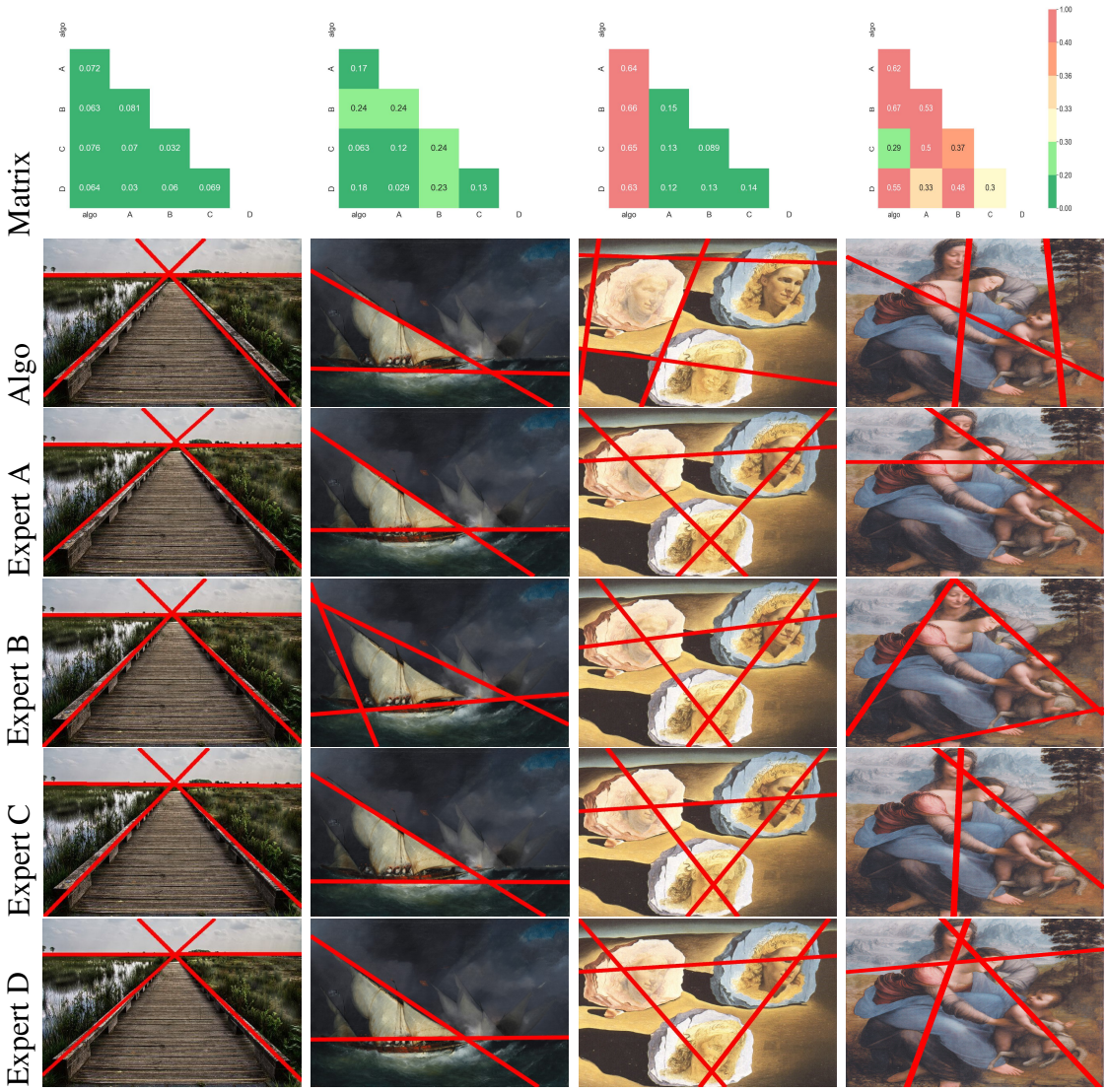


FIGURE 5.15 – Résultats de comparaison entre différents ensembles de lignes. La première ligne montre les distances entre les différents ensembles de lignes de force, les valeurs vertes représentent des distances inférieures à $0,3$ et deux ensembles de lignes avec une distribution spatiale cohérente, et les valeurs non vertes représentent des distances supérieures à $0,3$ et deux ensembles de lignes avec des distributions spatiales incohérentes. (a) indiquent que les résultats de l'algorithme sont cohérents avec les résultats de l'expert. (b) indique que les résultats de l'expert B ont des différences subjectives par rapport aux autres résultats, mais sont cohérents dans l'ensemble. (c) montre un accord entre les quatre experts, mais les résultats de notre algorithme sont incohérents avec les experts. (d) montre qu'il y a de grandes différences subjectives dans cette image, même parmi les experts.

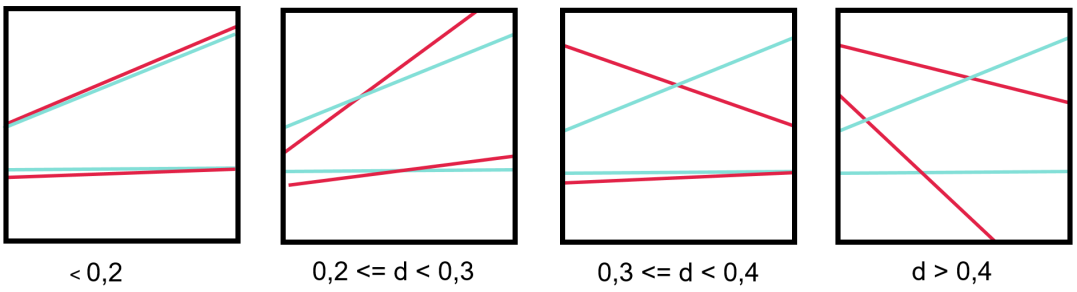


FIGURE 5.16 – Illustration de la distance entre deux ensembles de ligne.

conséquent, des distances supérieures à $0,3$ signifiaient une incohérence entre deux ensembles de lignes. En plus des résultats de la figure 5.15, d'autres résultats quantitatifs sur les différences entre le modèle et les experts sont ajoutés à l'annexe B.1.2.

Par la suite, nous avons établi deux seuils pour ces distances : $\tau_1 = 0,2$ était la distance limite supérieure pour deux ensembles indiquant l'accord global, tandis que $\tau_2 = 0,3$ a servi de distance limite supérieure pour deux ensembles avec des différences acceptables. Pour évaluer la corrélation entre l'algorithme et les quatre experts, nous avons calculé les distances médianes avec et sans les résultats de l'algorithme. Les résultats sont présentés à la figure 5.17, la ligne orange représente la distance médiane entre les quatre experts ; la ligne bleue indique la distance médiane entre les cinq ensembles, y compris l'ensemble de l'algorithme. Comme indiqué dans la figure, les résultats de l'algorithme étaient cohérents avec ceux des experts, à l'exception des images spécifiques qui présentaient des incohérences, même parmi les experts.

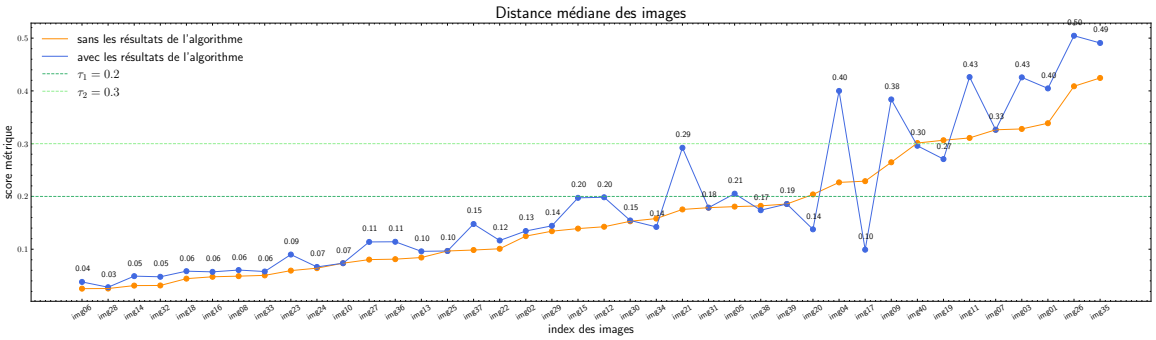
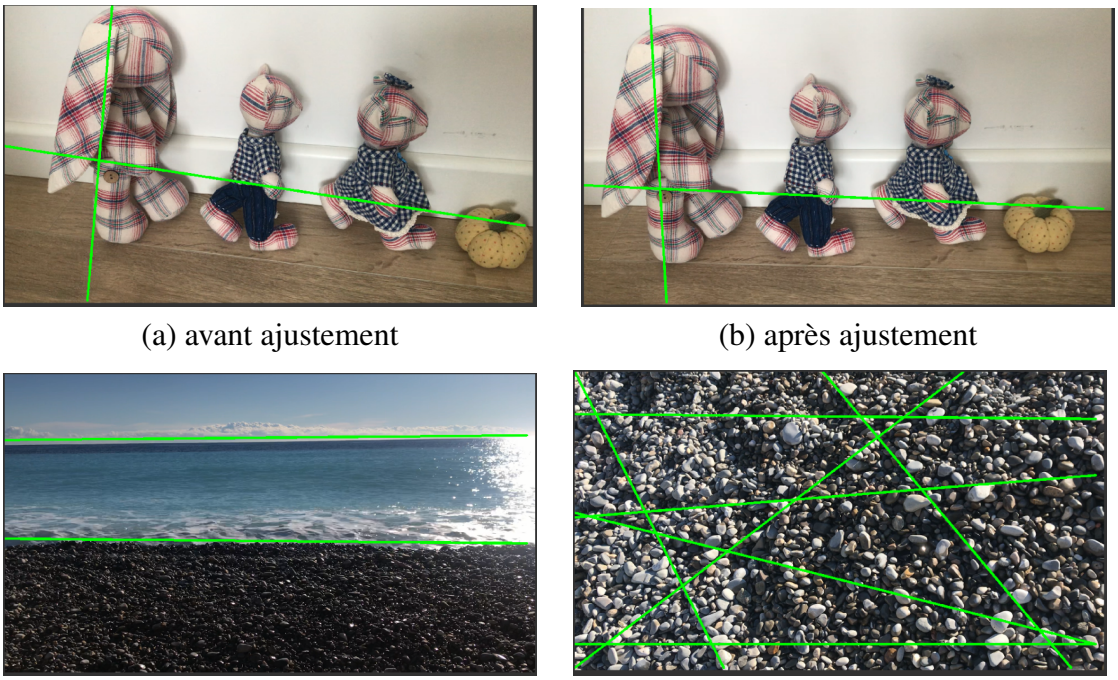


FIGURE 5.17 – Analyse de la distance entre les ensembles de lignes. La ligne orange illustre la distance médiane des quatre experts. La ligne bleue illustre la distance médiane des cinq experts, où le résultat du modèle est considéré comme le cinquième expert.

Dans notre étude empirique de ce modèle, nous l'avons comparé à une distribution aléatoire et analysé la cohérence de ses résultats avec un étiquetage expert. Les résultats expérimentaux ont démontré l'efficacité du modèle dans la reconstruction des lignes de force dans les images. Bien qu'il puisse présenter un biais dans certaines compositions spécifiques, comme le montre la figure 5.15 (c), le modèle s'est avéré applicable dans la majorité des scénarios de reconstruction des lignes de force, comme le montre la figure 5.14.

5.6 Application



(a) avant ajustement (b) après ajustement

(c) composé clairement (d) composé confus

FIGURE 5.18 – Applications des lignes de force. (b) est le résultat optimisé de (a) avec le guide des lignes de force affiché avec la couleur verte. (c) Plage de mer clairement composée, (d) Plage de galets sans composition claire.

La reconstruction des lignes de force aide à comprendre une composition d'image, faisant de notre modèle proposé un outil précieux pour les photographes. Ce modèle pourrait être intégré à un appareil photo, permettant à la position des lignes de force de s'adapter au contenu capturé, offrant ainsi aux photographes un guide de composition. Comme le montre la figure 5.18, (a) représente la composition initiale, tandis que la (b)

présente la composition ajustée en fonction de la position des lignes de force.

De plus, la position et le nombre de lignes de force contribuent également à capturer des images bien composées. Une abondance de lignes de force désorganisées dans une prise de vue indique souvent une composition d'image peu remarquable, comme le montre la figure 5.18 (d). En revanche, les images clairement composées avec des thèmes distincts ont tendance à présenter des lignes de force concises et bien organisées, illustrées à la figure 5.18 (c). L'algorithme de reconstruction des lignes de force fournit une référence objective pour évaluer la composition de l'image.

5.7 Conclusions et travaux futurs

Dans ce chapitre, nous avons démontré un consensus parmi les experts dans l'identification des lignes de force de la composition dans les images, soulignant l'utilité potentielle de la reconstruction automatique pour l'analyse de la composition de l'image et l'assistance lors de la capture d'images. Pour détecter les lignes de force probables, nous avons introduit un algorithme automatique non supervisé, qui calcule initialement toutes les lignes de force pondérées potentielles dans une image en fonction du contraste des pixels. Par la suite, il regroupe les lignes de force pondérées, générant finalement les lignes de force finales en identifiant les centres des groupes. Reconnaissant la variabilité subjective des lignes de force, nous avons conçu une mesure pour quantifier la distance entre deux ensembles de lignes, permettant des comparaisons entre des ensembles avec un nombre variable de lignes. En plus de mener plusieurs études subjectives, nous avons effectué des comparaisons objectives pour évaluer l'exactitude et la robustesse de notre algorithme.

Pour les travaux futurs, nous aspirons à aborder d'autres formes de courbes de force, en nous concentrant particulièrement sur les courbes circulaires. Nous visons également à élargir notre ensemble de données de vérité de base de la composition d'images pour ouvrir la voie à des algorithmes supervisés capables de gérer des compositions plus complexes et des préférences subjectives.

Résumé

La composition d'une image est une caractéristique essentielle choisie par l'auteur pour construire une image qui transmet une narration et les émotions qui s'y rattachent. Les lignes de force nous permettent de mieux comprendre ce que l'image raconte. Bien que les lignes de force ne soient pas toujours explicitement visibles dans l'image, elles relient les points clés de l'image et peuvent également servir de limites entre les différentes zones de l'image. Ce chapitre présente une méthode de calcul automatique pour récupérer les lignes de force qui sous-tendent la composition de l'image. Notre méthode comporte deux étapes : premièrement, sur la base de la détection des caractéristiques, des lignes de force potentielles pondérées sont établies ; deuxièmement, ces lignes de force pondérées sont groupées pour générer les lignes de force de l'image. Nous évaluons notre méthode à travers des évaluations subjectives et objectives, et nous proposons une mesure objective pour comparer deux ensembles de lignes de force.

Les travaux présentés dans ce chapitre ont donné lieu aux publications suivantes : Zhang, J. ; Synave, R. ; Delepoulle, S. ; Cozot, R. Reconstructing Image Composition : Computation of Leading Lines. J. Imaging 2024, 10, 5. https ://doi.org/10.3390/jimaging10010005.

Chapitre 6

Conclusion et perspectives

Synthèse

Dans cette thèse, nous discutons de l'ajustement de la luminance des images à haute gamme dynamique (HDR) en prenant en compte les caractéristiques esthétiques de l'image. Aujourd'hui, les contenus HDR se démocratisent et un nombre croissante d'appareils permettent de les afficher. Par conséquent, nous avons besoin de méthodes pour améliorer l'affichage du contenu HDR en l'adaptant aux caractéristiques des images HDR. Notre objectif est de maximiser la qualité esthétiquement perçue de l'affichage en adaptant les images HDR par des analyses esthétiques des images.

Au chapitre 2, nous introduisons le concept d'images HDR ainsi que les recherches associées. Dans le chapitre 3, nous présentons les méthodes concernant l'ajustement automatique des images HDR. Au début du chapitre, nous présentons une base de données HDR riche en scènes avec des paramètres d'édition. Cette base de données a été créée par un photographe professionnel et nous fournit une base riche pour l'apprentissage supervisé. Ensuite, nous entraînons un réseau de neurones pour mettre en œuvre la prédiction automatique de la courbe d'ajustement. Ce réseau de neurones se compose d'une seule couche entièrement connectée qui établit une relation de correspondance entre l'histogramme de l'image HDR et les cinq points de référence de la courbe d'ajustement. Dans la section 3.3, nous proposons un modèle de réseau de neurones convolutif, qui est nommé ExposureCNN, pour prédire l'ajustement d'exposition des images HDR. ExposureCNN comprend quatre modules de convolution et deux couches entièrement connectées. En extrayant les caractéristiques latentes des images HDR, il intègre également la valeur médiane de luminance de l'image originale pour guider les résultats de prédiction. Les deux méthodes proposées dans ce chapitre visent à améliorer l'affichage des images HDR sur les appareils HDR, et les deux méthodes prédisent les paramètres d'ajustement sur la base des caractéristiques pertinentes du HDR.

Au chapitre 4, nous présentons les recherches liées à l'analyse esthétique des images. Dans le chapitre 5 de la thèse, nous présentons une recherche sur les lignes de force de la composition. Nous décrivons tout d'abord le concept de lignes de force. Dans une image, les lignes de force peuvent guider la perception visuelle globale, nous aidant à comprendre le contenu et le design de la composition de l'image. L'identification des lignes de force est relativement subjective, nous avons d'abord évalué la faisabilité d'une reconstruction automatique des lignes de force. Après avoir vérifié la cohérence des résultats donnés par les experts, nous avons conçu un algorithme de regroupement pour compléter la reconstruction des lignes de force. Notre approche comporte deux étapes : premièrement, sur la base de la détection des caractéristiques, des lignes de force pondérées potentielles sont créés ; deuxièmement, ces lignes de force pondérées sont regroupées pour générer des lignes de force finales de l'image. Nous proposons également une métrique pour comparer les deux ensembles de lignes. Des expériences de comparaison subjectives et objectives confirment la validité de la méthode proposée. Cette méthode de reconstruction de la ligne de force est applicable aux images LDR et HDR, et peut servir d'outil de guidage pour l'édition d'images ou comme outil d'analyse esthétique de la composition.

Perspectives

Actuellement, lorsque nous ajustons le contenu HDR, c'est généralement sur la base d'une configuration spécifique de dispositifs d'affichage afin de maximiser l'expérience visuelle sur les capacités d'affichage existantes. Grâce au développement continu de la technologie de fabrication des dispositifs d'affichage HDR, la capacité à afficher la gamme dynamique des images s'améliore constamment. Cependant, comment adapter le contenu visuel à différentes configurations de dispositifs d'affichage et maximiser la conservation de la perception esthétique du contenu au cours du processus d'adaptation est une question qui continuera d'être explorée.

La méthode de la ligne de force que nous proposons fonctionne actuellement principalement pour l'analyse d'images. À l'avenir, nous espérons l'étendre à l'analyse vidéo pour analyser la fluidité des scènes dynamiques. La position et la direction des lignes de force peuvent refléter la focalisation et la tendance du contenu de l'image. Par conséquent, des lignes de force qui changent de manière régulière peuvent garantir la fluidité de la vidéo.

En outre, lors de l'ajustement d'images HDR, il n'est généralement pas possible d'afficher la gamme dynamique complète de la scène. Par conséquent, nous pouvons combiner la ligne de force pour sélectionner la zone saillante de l'image, en nous concen-

trant sur la préservation des variations de la gamme dynamique de ces zones lors de l'ajustement, afin d'améliorer l'expérience visuelle.

Dans la poursuite de nos recherches dans ce domaine, nous espérons que, tout en ajustant le contenu visuel, l'ordinateur « apprenne » à comprendre l'esthétique des images et à mesurer efficacement la perception esthétique. La compréhension de l'esthétique d'une image ne se limite pas à la compréhension d'un élément naturel particulier, mais doit également prendre en compte les interconnexions entre les différents éléments du contenu et l'image dans son ensemble. Comme le propose la théorie psychologique de la Gestalt, le tout est plus grand que la somme de ses parties. Notre perception d'une image ne provient pas seulement des informations sensorielles de l'image telles que la lumière, la couleur, la composition, etc., mais aussi de notre impression et de notre connaissance de la scène contenue dans l'image, qui s'ajoutent à notre perception de l'image. Peut-être pourrions-nous envisager d'intégrer davantage de résultats expérimentaux issus de la psychologie et des neurosciences, ajoutant différentes dimensions de données à l'analyse. En plus d'analyser les caractéristiques intrinsèques des images, nous examinons l'impact de ces caractéristiques sur les résultats de perception dans différents groupes d'utilisateurs.

Bibliographie

Abbasov, I. B. (2021). Perception of Images. Modern Trends. https://doi.org/10.1002/9781119751991.ch1

Abeln, J., Fresz, L., Amirshahi, S. A., McManus, I. C., Koch, M., Kreysa, H. & Redies, C. (2016). Preference for Well-Balanced Saliency in Details Cropped from Photographs. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00704

Afifi, M., Derpanis, K. G., Ommer, B. & Brown, M. S. (2021, mars 30). *Learning Multi-Scale Photo Exposure Correction*. arXiv : 2003.11596 [cs, eess]. http://arxiv.org/abs/2003.11596

Arnheim, R. (2004, novembre 8). *Art and Visual Perception, Second Edition : A Psychology of the Creative Eye*. University of California Press.

Aydin, T., Mantiuk, R., Myszkowski, K. & Seidel, H.-P. (2008). Dynamic Range Independent Image Quality Assessment. *Turk, Greg : Proceedings of ACM SIGGRAPH 2008, ACM, Art.69.1-10 (2008)*, *27*. https://doi.org/10.1145/1360612.1360668

Aydin, T. O., Smolic, A. & Gross, M. (2015). Automated Aesthetic Analysis of Photographic Images. *IEEE Transactions on Visualization and Computer Graphics*, *21*(1), 31-42. https://doi.org/10.1109/TVCG.2014.2325047

Aydın, T. O., Mantiuk, R. & Seidel, H.-P. (2008, février 14). Extending quality metrics to full luminance range images. In B. E. Rogowitz & T. N. Pappas (Éd.). https://doi.org/10.1117/12.765095

Bang, M. (2000). *Picture this : how pictures work*. SeaStar Books.

Banterle, F., Artusi, A., Debattista, K. & Chalmers, A. (2018). *Advanced high dynamic range imaging* (Second edition). Taylor & Francis, CRC Press.

Banterle, F., Artusi, A., Moreo, A. & Carrara, F. (2020). Nor-Vdpnet : A No-Reference High Dynamic Range Quality Metric Trained On Hdr-Vdp 2. *2020 IEEE International Conference on Image Processing (ICIP)*, 126-130. https://doi.org/10.1109/ICIP40778.2020.9191202

Banterle, F., Ledda, P., Debattista, K., Chalmers, A. & Bloj, M. (2007). A Framework for Inverse Tone Mapping. *The Visual Computer*, *23*, 467-478. https://doi.org/10.1007/s00371-007-0124-9

Baumgartner, T., Paassen, B. & Klatt, S. (2023). Extracting spatial knowledge from track and field broadcasts for monocular 3D human pose estimation. *Scientific Reports*, *13*(1), 1-12.

Bianco, S., Cusano, C., Piccoli, F. & Schettini, R. (2020). Personalized Image Enhancement Using Neural Spline Color Transforms. *IEEE Transactions on Image Processing*, *PP*, 1-1. https://doi.org/10.1109/TIP.2020.2989584

Birkhoff, G. D. (2013, octobre 1). Aesthetic Measure. *Aesthetic Measure*. Harvard University Press. https://doi.org/10.4159/harvard.9780674734470

Bist, C., Cozot, R., Madec, G. & Ducloux, X. (2016). Style Aware Tone Expansion for HDR Displays. *Proceedings of the 42nd Graphics Interface Conference*, 57-63.

Bist, C., Cozot, R., Madec, G. & Ducloux, X. (2017). Tone expansion using lighting style aesthetics. *Computers & Graphics*, *62*, 77-86. https://doi.org/10.1016/j.cag.2016.12.006

Brown, B. (2016). *Cinematography : theory and practice : imagemaking for cinematographers and directors* (Third edition). Routledge, Taylor & Francis Group.

Brown, M., Windridge, D. & Guillemaut, J.-Y. (2015). A generalisable framework for saliency-based line segment detection. *Pattern Recognition*, *48*(12), 3993-4011. https://doi.org/10.1016/j.patcog.2015.06.015

Bychkovsky, V., Paris, S., Chan, E. & Durand, F. (2011). Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs.

Cao, F., Musé, P. & Sur, F. (2005). Extracting Meaningful Curves from Images. *Journal of Mathematical Imaging and Vision*, *22*, 159-181. https://doi.org/10.1007/s10851-005-4888-0

Celona, L., Leonardi, M., Napoletano, P. & Rozza, A. (2021, novembre 8). *Composition and Style Attributes Guided Image Aesthetic Assessment*.

Cerdá, X., Párraga, C. A. & Otazu, X. (2018). Which tone-mapping operator is the best ? A comparative study of perceptual quality. *Journal of the Optical Society of America A*, *35*(4), 626. https://doi.org/10.1364/JOSAA.35.000626

Chambe, M., Cozot, R. & Le Meur, O. (2019, novembre). *Behaviour of Recent Aesthetics Assessment Models with Professional Photography*. https://hal.science/hal-02374494

Chandra, A. L. (2022, septembre 27). *McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron*. Medium. https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1

Chang, H., Fried, O., Liu, Y., DiVerdi, S. & Finkelstein, A. (2015). Palette-based photo recoloring. *ACM Transactions on Graphics*, *34*(4), 1-11. https://doi.org/10.1145/2766978

Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C. & Zimmerman, K. (1993). Spatially Nonuniform Scaling Functions for High Contrast Images. *Graphics Interface '*.

Chopra, S., Hadsell, R. & Lecun, Y. (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. *1*, 539-546 vol. 1. https://doi.org/10.1109/CVPR.2005.202

CIE-17-21-003. (s. d.). https://cie.co.at/eilvterm/17-21-003

CIE-e-ILV. (s. d.). https://cie.co.at/e-ilv

Cmglee. (s. d.). *Density of Photoreceptors*. https://commons.wikimedia.org/w/index.php?curid=29924570

Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T. & Xu, Y.-Q. (2006). Color Harmonization. *ACM Transactions on Graphics*, *25*(3), 624-630. https://doi.org/10.1145/1141911.1141933

Csurka, G., Dance, C., Fan, L., Willamowski, J. & Bray, C. (2004). Visual Categorization with Bags of Keypoints. *Work Stat Learn Comput Vision, ECCV*, *Vol. 1*.

Dai, X., Gong, H., Wu, S., Yuan, X. & Ma, Y. (2022). Fully convolutional line parsing. *Neurocomputing*, *506*, 1-11. https://doi.org/https://doi.org/10.1016/j.neucom.2022.07.026

Datta, R., Joshi, D., Li, J. & Wang, J. Z. (2006). Studying Aesthetics in Photographic Images Using a Computational Approach. In A. Leonardis, H. Bischof & A. Pinz (Éd.), *Computer Vision – ECCV 2006* (p. 288-301). Springer Berlin Heidelberg. https://doi.org/10.1007/11744078_23

Datta, R. & Wang, J. Z. (2010). ACQUINE : aesthetic quality inference engine - real-time automatic rating of photo aesthetics. *Proceedings of the International Conference on Multimedia Information Retrieval*, 421-424. https://doi.org/10.1145/1743384.1743457

Debevec, P. E. & Malik, J. (1997). Recovering High Dynamic Range Radiance Maps from Photographs.

Debnath, S., Roy, R. & Changder, S. (2022). A novel approach using deep convolutional neural network to classify the photographs based on leading-line by fine-tuning the pre-trained VGG16 neural network. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13338-5

Deng, Y., Loy, C. C. & Tang, X. (2018). Aesthetic-Driven Image Enhancement by Adversarial Learning, 870-878. https://doi.org/10.1145/3240508.3240531

Desolneux, A., Moisan, L. & Morel, J.-M. (2000). Meaningful alignments. *International Journal of Computer Vision*, *40*(1), 7-23. https://doi.org/10.1023/A:1026593302236

Dhar, S., Ordonez, V. & Berg, T. L. (2011). High level describable attributes for predicting aesthetics and interestingness. *CVPR 2011*, 1657-1664. https://doi.org/10.1109/CVPR.2011.5995467

Douchová, V. (2016). Birkhoff's aesthetic measure. *AUC PHILOSOPHICA ET HISTORICA*, *2015*(1), 39-53. https://doi.org/10.14712/24647055.2016.8

DPChallenge - A Digital Photography Contest. (s. d.). https://www.dpchallenge.com/index.php

Drago, F., Myszkowski, K., Annen, T. & Chiba, N. (2003). Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Computer Graphics Forum*, *22*(3), 419-426. https://doi.org/10.1111/1467-8659.00689

Duda, R. O. & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, *15*(1), 11-15. https://doi.org/10.1145/361237.361242

Durand, F. & Dorsey, J. (2002). Fast Bilateral Filtering for the Display of High-Dynamic-Range Images.

Dykinga, J. W. (2014). *Capture the magic : train your eye, improve your photographic composition* (1st ed). Rocky Nook.

Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R. & Unger, J. (2017). HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics*, *36*. https://doi.org/10.1145/3130800.3130816

Fairchild, M. & Chen, P.-H. (2011). Brightness, Lightness, and Specifying Color in High-Dynamic-Range Scenes and Images. *Proceedings of SPIE - The International Society for Optical Engineering*, *7867*, 23. https://doi.org/10.1117/12.872075

Fairchild, M. & Johnson, G. (2004). The iCAM Framework for Image Appearance, Image Differences, and Image Quality. *J. Electronic Imaging*, *13*, 126-138. https://doi.org/10.1117/1.1635368

Fairchild, M. D. (2005, juillet 8). *Color Appearance Models*. John Wiley & Sons.

Fairchild, M. D. (2007). The HDR Photographic Survey. *15th Color and Imaging Conference, CIC 2007, Albuquerque, New Mexico, USA, November 5-9, 2007*, 233-238. http://www.ingentaconnect.com/content/ist/cic/2007/00002007/00000001/art00044

Fairchild, M. D. & Johnson, G. M. (2002). Meet iCAM : A Next-Generation Color Appearance Model. *Color and Imaging Conference*, *10*(1), 33-38. https://doi.org/10.2352/CIC.2002.10.1.art00008

Fattal, R., Lischinski, D. & Werman, M. (2002). Gradient domain high dynamic range compression. *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, 249-256. https://doi.org/10.1145/566570.566573

Fechner, G. T. (1966). *Elements of psychophysics.* Holt, Rinehart and Winston, Inc.

Fernandes, L. A. & Oliveira, M. M. (2008). Real-time line detection through an improved Hough transform voting scheme. *Pattern Recognition*, *41*(1), 299-314. https://doi.org/10.1016/j.patcog.2007.04.003

fidle. (s. d.). https://fidle.cnrs.fr/w3/

Filmic. (2015). SlideShare. https://www.slideshare.net/hpduiker/filmic-tonemapping-for-realtime-rendering-siggraph-2010-color-course

Fischer, M., Kobs, K. & Hotho, A. (2020, décembre 3). *NICER : Aesthetic Image Enhancement with Humans in the Loop*. arXiv : 2012.01778 [cs]. http://arxiv.org/abs/2012.01778

flickr. (2024, février 26). https://www.flickr.com/

Freeman, M. (2007). *The Photographer's Eye : Composition and Design for Better Digital Photos*. Focal Press.

Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A. & Brendel, H. (2014). *Creating Cinematic Wide Gamut HDR-Video for the Evaluation of Tone Mapping Operators and HDR-Displays*. https://doi.org/10.1117/12.2040003

Funt, B. & Shi, L. (2010, février 4). The effect of exposure on MaxRGB color constancy. In B. E. Rogowitz & T. N. Pappas (Éd.). https://doi.org/10.1117/12.845394

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd). O'Reilly Media, Inc.

Ghadiyaram, D. & Bovik, A. C. (2016). Massive Online Crowdsourced Study of Subjective and Objective Picture Quality. *IEEE Transactions on Image Processing*, *25*(1), 372-387. https://doi.org/10.1109/TIP.2015.2500021

Gharbi, M., Chen, J., Barron, J. T., Hasinoff, S. W. & Durand, F. (2017). Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, *36*(4), 1-12. https://doi.org/10.1145/3072959.3073592

González-Casillas, A., Parra, L., Martin, L., Avila-Contreras, C., Ramirez-Pedraza, R., Vargas, N., del Valle-Padilla, J. L. & Ramos, F. (2018). Towards a model of visual recognition based on neurosciences. *Biologically Inspired Cognitive Architectures*, *25*, 119-129. https://doi.org/10.1016/j.bica.2018.07.018

Gringer, P. R. (2018, mars 30). *Situation Du Visible Dans Le Spectre Électromagnétique*. https://commons.wikimedia.org/wiki/File:EM_spectrumrevised_fr.png

Grompone Von Gioi, R., Jakubowicz, J., Morel, J.-M. & Randall, G. (2012). LSD : a Line Segment Detector. *Image Processing On Line*, *2*, 35-55. https://doi.org/10.5201/ipol.2012.gjmr-lsd

Guo, C. & Jiang, X. (2021). Deep Tone-Mapping Operator Using Image Quality Assessment Inspired Semi-Supervised Learning. *IEEE Access*, *PP*, 1-1. https://doi.org/10.1109/ACCESS.2021.3080331

Guo, X. (2016, juillet 24). *LIME : A Method for Low-light IMage Enhancement*. http://arxiv.org/abs/1605.05034

He, K., Zhang, X., Ren, S. & Sun, J. (2015, décembre 10). *Deep Residual Learning for Image Recognition*. arXiv : 1512.03385 [cs]. https://doi.org/10.48550/arXiv.1512.03385

He, S., Ming, A., Li, Y., Sun, J., Zheng, S. & Ma, H. (2023). Thinking Image Color Aesthetics Assessment : Models, Datasets and Benchmarks. *2023 IEEE/CVF*

International Conference on Computer Vision (ICCV), 21781-21790. https://doi.org/10.1109/ICCV51070.2023.01996

Henry, C. (1885). *Introduction a une esthétique scientifique*. La Revue contemporaine.

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-term Memory. *Neural computation*, *9*, 1735-80. https://doi.org/10.1162/neco.1997.9.8.1735

Hoenig, F. (2005). *Defining Computational Aesthetics*. The Eurographics Association. https://doi.org/10.2312/COMPAESTH/COMPAESTH05/013-018 Accepted : 2013-10-22T07 :40 :19Z

Hosu, V., Goldlucke, B. & Saupe, D. (2019, avril 2). *Effective Aesthetics Prediction with Multi-level Spatially Pooled Features* (1). arXiv : 1904.01382 [cs]. http://arxiv.org/abs/1904.01382

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. (2017, avril 16). *MobileNets : Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv : 1704.04861 [cs]. https://doi.org/10.48550/arXiv.1704.04861

Hristova, H., Le Meur, O., Cozot, R. & Bouatouch, K. (2017). High-Dynamic-Range Image Recovery from Flash and Non-Flash Image Pairs. *The Visual Computer*, *33*, 1-11. https://doi.org/10.1007/s00371-017-1399-0

Hu, Y., He, H., Xu, C., Wang, B. & Lin, S. (2018, février 6). *Exposure : A White-Box Photo Post-Processing Framework*. http://arxiv.org/abs/1709.09602

Huang, K., Wang, Y., Zhou, Z., Ding, T., Gao, S. & Ma, Y. (2018). Learning to Parse Wireframes in Images of Man-Made Environments. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 626-635. https://doi.org/10.1109/CVPR.2018.00072

ITU. (1990). ITU-R Recommendation BT.709, Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange. *Geneva : ITU*. https://www.itu.int/rec/R-REC-BT.709-6-201506-I/en

Jin, D., Lee, J.-T. & Kim, C.-S. (2022). Semantic Line Detection Using Mirror Attention and Comparative Ranking and Matching.

Jin, D., Park, W., Jeong, S.-G. & Kim, C.-S. (2021). Harmonious Semantic Line Detection via Maximal Weight Clique Selection.

Jin, X., Li, X., Lou, H., Fan, C., Deng, Q., Xiao, C., Cui, S. & Singh, A. K. (2022, juillet 5). *Aesthetic Attribute Assessment of Images Numerically on Mixed Multi-attribute Datasets* (1). arXiv : 2207.01806 [cs]. http://arxiv.org/abs/2207.01806

Jin, X., Wu, L., Li, X., Chen, S., Peng, S., Chi, J., Ge, S., Song, C. & Zhao, G. (2017, novembre 20). *Predicting Aesthetic Score Distribution through Cumulative Jensen-Shannon Divergence*. arXiv : 1708.07089 [cs]. http://arxiv.org/abs/1708.07089

Jin, X., Wu, L., Zhao, G., Li, X., Zhang, X., Ge, S., Zou, D., Zhou, B. & Zhou, X. (2019, juillet 29). *Aesthetic Attributes Assessment of Images*. arXiv : 1907.04983 [cs]. http://arxiv.org/abs/1907.04983

Joly, M. (2015, mai 6). *Introduction à l'analyse de l'image - 3e édition*. Armand Colin.

Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.-T., Wang, J., Li, J. & Luo, J. (2011). Aesthetics and Emotions in Images. *IEEE Signal Processing Magazine*, *28*(5), 94-115. https://doi.org/10.1109/MSP.2011.941851

Kang, C., Valenzise, G. & Dufaux, F. (2020). EVA : An Explainable Visual Aesthetics Dataset. *Joint Workshop on Aesthetic and Technical Quality Assessment of Multimedia and Media Analytics for Societal Trends (ATQAM/MAST'20), ACM Multimedia*, 5-13. https://doi.org/10.1145/3423268.3423590

Kao, Y., He, R. & Huang, K. (2017). Deep Aesthetic Quality Assessment with Semantic Information. *IEEE Transactions on Image Processing*, *26*(3), 1482-1495. https://doi.org/10.1109/TIP.2017.2651399

Kingma, D. P. & Ba, J. (2017, janvier 29). *Adam : A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Kipf, T. N. & Welling, M. (2017, février 22). *Semi-Supervised Classification with Graph Convolutional Networks*. arXiv : 1609.02907 [cs, stat]. http://arxiv.org/abs/1609.02907

Kiryati, N., Eldar, Y. & Bruckstein, A. (1991). A probabilistic Hough transform. *Pattern Recognition*, *24*(4), 303-316. https://doi.org/10.1016/0031-3203(91)90073-E

Kong, S., Shen, X., Lin, Z., Mech, R. & Fowlkes, C. (2016, juillet 26). *Photo Aesthetics Ranking Network with Attributes and Content Adaptation*. arXiv : 1606.01621 [cs]. https://doi.org/10.48550/arXiv.1606.01621

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Kuang, J., Johnson, G. & Fairchild, M. (2007). iCAM06 : A Refined Image Appearance Model for HDR Image Rendering. *Journal of Visual Communication and Image Representation*, *18*, 406-414. https://doi.org/10.1016/j.jvcir.2007.06.003

Kuang-Yu Chang, Lu, K.-H. & Chen, C.-S. (2017). Aesthetic Critiques Generation for Photos. *2017 IEEE International Conference on Computer Vision (ICCV)*, 3534-3543. https://doi.org/10.1109/ICCV.2017.380

Larson, G., Rushmeier, H. & Piatko, C. (1997). A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, *3*(4), 291-306. https://doi.org/10.1109/2945.646233

Lauga, P., Koz, A., Valenzise, G. & Dufaux, F. (2013). Segmentation-Based Optimized Tone Mapping for High Dynamic Range Image and Video Coding, 257-260. https://doi.org/10.1109/PCS.2013.6737732

Le Robert Illustré. (2018).

Ledda, P., Chalmers, A., Troscianko, T. & Seetzen, H. (2005). Evaluation of Tone Mapping Operators Using a High Dynamic Range Display. *ACM Transactions on Graphics*, *24*(3), 640-648. https://doi.org/10.1145/1073204.1073242

Lee, J.-T., Kim, H.-U., Lee, C. & Kim, C.-S. (2017). Semantic Line Detection and Its Applications. *2017 IEEE International Conference on Computer Vision (ICCV)*, 3249-3257. https://doi.org/10.1109/ICCV.2017.350

Levina, E. & Bickel, P. (2001). The Earth Mover's Distance Is the Mallows Distance : Some Insights from Statistics. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, *2*, 251-256 vol.2. https://doi.org/10.1109/ICCV.2001.937632

Li, D., Wu, H., Zhang, J. & Huang, K. (2018, mars 12). *A2-RL : Aesthetics Aware Reinforcement Learning for Image Cropping*. arXiv : 1709.04595 [cs]. http://arxiv.org/abs/1709.04595

Lo, K.-Y., Liu, K.-H. & Chen, C.-S. (2013). Intelligent Photographing Interface with On-Device Aesthetic Quality Assessment. In J.-I. Park & J. Kim (Éd.), *Computer Vision - ACCV 2012 Workshops* (p. 533-544). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37484-5_43

Lu, X., Lin, Z., Jin, H., Yang, J. & Wang, J. Z. (2014). RAPID : Rating Pictorial Aesthetics using Deep Learning. *Proceedings of the 22nd ACM International Conference on Multimedia*, 457-466. https://doi.org/10.1145/2647868.2654927

Lu, X., Lin, Z., Shen, X., Mech, R. & Wang, J. Z. (2015). Deep Multi-patch Aggregation Network for Image Style, Aesthetics, and Quality Estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, 990-998. https://doi.org/10.1109/ICCV.2015.119

Ma, S., Liu, J. & Chen, C. W. (2017, avril 1). *A-Lamp : Adaptive Layout-Aware Multi-Patch Deep Convolutional Neural Network for Photo Aesthetic Assessment* (1). arXiv : 1704.00248 [cs]. http://arxiv.org/abs/1704.00248

Mai, L., Jin, H. & Liu, F. (2016). Composition-Preserving Deep Photo Aesthetics Assessment. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 497-506. https://doi.org/10.1109/CVPR.2016.60

Mantiuk, R., Daly, S. & Kerofsky, L. (2008). Display adaptive tone mapping. *ACM SIGGRAPH 2008 Papers*, 1-10. https://doi.org/10.1145/1399504.1360667

Mantiuk, R. K. & Azimi, M. (2021). PU21 : A novel perceptually uniform encoding for adapting existing quality metrics for HDR. *2021 Picture Coding Symposium (PCS)*, 1-5. https://doi.org/10.1109/PCS50896.2021.9477471

Mantiuk, R. K., Hammou, D. & Hanji, P. (2023, avril 26). *HDR-VDP-3 : A Multi-Metric for Predicting Image Differences, Quality and Contrast Distortions in High Dynamic Range and Regular Content*. arXiv : 2304.13625 [cs, eess]. http://arxiv.org/abs/2304.13625

Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G. & Michallon, L. (2012). Évaluation Automatique de La Qualité Esthétique Des Photographies à l'aide de Descripteurs d'images Génériques. *RFIA 2012 (Reconnaissance Des Formes et Intelligence Artificielle)*, 978-2-9539515-2-3. https://hal.science/hal-00656535

Matković, K. & Purgathofer, W. (1998). Automatic Exposure in Computer Graphics Based on the Minimum Information Loss Principle., 666.

Mertens, T., Kautz, J. & Van Reeth, F. (2007). Exposure Fusion, 382-390. https://doi.org/10.1109/PG.2007.17

Ming-Ming, C., Guo-Xin, Z., Mitra, N. J., Xialei, H. & Shi-Min, H. (2011). CVPR.

Murray, N., Marchesotti, L. & Perronnin, F. (2012). AVA : A large-scale database for aesthetic visual analysis. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2408-2415. https://doi.org/10.1109/CVPR.2012.6247954

Myndex. (2022, avril 4). *CIE 1931 Chromaticity Diagram*. https://commons.wikimedia.org/wiki/File:CIE1931xy_gamut_comparison_of_sRGB_P3_Rec2020.svg

Newell, D. B. & Tiesinga, E. (2019, août). *The international system of units (SI) : 2019 edition* (NIST SP 330-2019). National Institute of Standards and Technology. Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.330-2019

Nieto, D. V., Celona, L. & Fernandez-Labrador, C. (2022, septembre 21). *Understanding Aesthetics with Language : A Photo Critique Dataset for Aesthetic Assessment* (3). arXiv : 2206.08614 [cs]. http://arxiv.org/abs/2206.08614

Nishiyama, M., Okabe, T., Sato, I. & Sato, Y. (2011). Aesthetic quality classification of photographs based on color harmony. *CVPR 2011*, 33-40. https://doi.org/10.1109/CVPR.2011.5995539

Nsampi, N. E., Zhongyun, H. & Qing, W. (2021). Learning Exposure Correction Via Consistency Modeling.

Panetta, K., Kezebou, L., Oludare, V., Agaian, S. & Xia, Z. (2021). TMO-Net : A Parameter-Free Tone Mapping Operator Using Generative Adversarial Network, and Performance Benchmarking on Large Scale HDR Dataset. *IEEE Access*, *9*, 39500-39517. https://doi.org/10.1109/ACCESS.2021.3064295

Pattanaik, S. N., Fairchild, M. D., Ferwerda, J. A. & Greenberg, D. P. (1998). Multiscale Model of Adaptation, Spatial Vision and Color Appearance. *Color and Imaging Conference*, *6*(1), 2-7. https://doi.org/10.2352/CIC.1998.6.1.art00002

Perronnin, F. & Dance, C. (2007). Fisher Kernels on Visual Vocabularies for Image Categorization, 1-8. https://doi.org/10.1109/CVPR.2007.383266

PhotoNet Home. (2024, février 26). Photo.net. https://www.photo.net

Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M. & Battisti, F. (2013). COLOR IMAGE DATABASE TID2013 : PECULIARITIES AND PRELIMINARY RESULTS.

Princen, J., Illingworth, J. & Kittler, J. (1989). A Hierarchical Approach to Line Extraction. *Proceedings CVPR '89 : IEEE Computer Society Conference on Computer*

Vision and Pattern Recognition, 92-97. https://doi.org/10.1109/CVPR.1989.37833

Rana, A., Singh, P., Valenzise, G., Dufaux, F., Komodakis, N. & Smolic, A. (2020). Deep Tone Mapping Operator for High Dynamic Range Images. *IEEE Transactions on Image Processing*, *29*, 1285-1298. https://doi.org/10.1109/TIP.2019.2936649

Rana, A. A., Valenzise, G. & Dufaux, F. (2017). Learning-Based Adaptive Tone Mapping for Keypoint Detection. *IEEE International Conference on Multimedia & Expo (ICME'2017)*. https://doi.org/10.1109/icme.2017.8019394

Reinhard, E. & Devlin, K. (2005). Dynamic Range Reduction Inspired by Photoreceptor Physiology. *IEEE Transactions on Visualization and Computer Graphics*, *11*(01), 13-24. https://doi.org/10.1109/TVCG.2005.9

Reinhard, E. (Éd.). (2010). *High dynamic range imaging : acquisition, display, and image-based lighting* (2nd ed). Morgan Kaufmann/Elsevier.

Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A. & Damberg, G. (2012). Calibrated Image Appearance Reproduction. *ACM Transactions on Graphics (TOG)*, *31*, 201. https://doi.org/10.1145/2366145.2366220

Reinhard, E., Stark, M., Shirley, P. & Ferwerda, J. (2002). Photographic Tone Reproduction for Digital Images.

Rempel, A., Trentacoste, M., Seetzen, H., Young, D., Heidrich, W., Whitehead, L. & Ward, G. (2007). Ldr2Hdr : On-the-fly Reverse Tone Mapping of Legacy Video and Photographs.

RGBCMF. (s. d.). https://en.wikipedia.org/wiki/File:CIE1931_RGBCMF.svg

Robertson, M., Borman, S. & Stevenson, R. (2000). Dynamic Range Improvement Through Multiple Exposures. *1999 International Conference on Image Processing*.

Santos, M. S., Tsang, R. & Khademi Kalantari, N. (2020). Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. *ACM Transactions on Graphics*, *39*(4). https://doi.org/10.1145/3386569.3392403

Schulz, S., Grimm, M. & Grigat, R.-R. (2007). Using Brightness Histogram to Perform Optimum Auto Exposure. *WSEAS Trans. System and Control*, *2*.

Shamoi, P., Inoue, A. & Kawanaka, H. (2022). Color Aesthetics and Context-Dependency, 1-7. https://doi.org/10.1109/SCISISIS55246.2022.10001872

Shannon, C. E. (1984). A Mathematical Theory of Communication.

She, D., Lai, Y.-K., Yi, G. & Xu, K. (2021). Hierarchical Layout-Aware Graph Convolutional Network for Unified Aesthetics Assessment. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8471-8480. https://doi.org/10.1109/CVPR46437.2021.00837

Sheng, K., Dong, W., Ma, C., Mei, X., Huang, F. & Hu, B.-G. (2018). Attention-Based Multi-Patch Aggregation for Image Aesthetic Assessment, 879-886. https://doi.org/10.1145/3240508.3240554

Smith, T. & Guild, J. (1931). The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*, *33*(3), 73-134. https://doi.org/10.1088/1475-4878/33/3/301

Stevens, S. S. (1970). Neural Events and the Psychophysical Law. *Science*, *170*(3962), 1043-1050. https://doi.org/10.1126/science.170.3962.1043

Stockham, T. (1972). Image processing in the context of a visual model. *Proceedings of the IEEE*, *60*(7), 828-842. https://doi.org/10.1109/PROC.1972.8782

Suárez, I., Buenaposada, J. M. & Baumela, L. (2021). ELSED : Enhanced Line SEgment Drawing.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2015, décembre 11). *Rethinking the Inception Architecture for Computer Vision*. arXiv : 1512.00567 [cs]. https://doi.org/10.48550/arXiv.1512.00567

Talebi, H. & Milanfar, P. (2018). NIMA : Neural Image Assessment. *IEEE Transactions on Image Processing*, *27*(8), 3998-4011. https://doi.org/10.1109/TIP.2018.2831899

Tan, M. & Le, Q. V. (2020, septembre 11). *EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks*. arXiv : 1905.11946 [cs, stat]. https://doi.org/10.48550/arXiv.1905.11946

Teplyakov, L., Erlygin, L. & Shvets, E. (2022). LSDNet : Trainable Modification of LSD Algorithm for Real-Time Line Segment Detection. *IEEE Access*, *10*, 45256-45265. https://doi.org/10.1109/ACCESS.2022.3169177

Tong, H., Li, M., He, J. & Zhang, C. (2004). Classification of Digital Photos Taken by Photographers or Home Users. *3331*, 198-205. https://doi.org/10.1007/978-3-540-30541-5_25

Tumblin, J. & Rushmeier, H. (1993). Tone Reproduction for Realistic Images. *Computer Graphics and Applications, IEEE*, *13*, 42-48. https://doi.org/10.1109/38.252554

Tumblin, J. & Turk, G. (1999). LCIS : a boundary hierarchy for detail-preserving contrast reduction. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '99*, 83-90. https://doi.org/10.1145/311535.311544

Urbano, C., Magalhães, L., Moura, J. P., Bessa, M., Fernandes-Marcos, A. & Chalmers, A. (2010). Tone Mapping Operators on Small Screen Devices : An Evaluation Study. *Comput. Graph. Forum*, *29*, 2469-2478. https://doi.org/10.1111/j.1467-8659.2010.01758.x

Valenzise, G., Kang, C. & Dufaux, F. (2022). Advances and Challenges in Computational Image Aesthetics. *Human Perception of Visual Information : Psychological and Computational Perspectives* (p. 133-181). Springer. https://doi.org/10.1007/978-3-030-81465-6_6

Vesa Certified DisplayHDR. (s. d.). VESA Certified DisplayHDR. https://displayhdr.org/

Wandell, ©. B. A. & Use, S. U. |. T. of. (1995). *Foundations of Vision » Chapter 9 : Color*. https://foundationsofvision.stanford.edu/chapter-9-color/

Wang, W. & Shen, J. (2017, octobre 22). *Deep Cropping via Attention Box Prediction and Aesthetics Assessment*. arXiv : 1710.08014 [cs]. http://arxiv.org/abs/1710.08014

Wang, Y., Liu, Y. & Xu, K. (2019). An Improved Geometric Approach for Palette-based Image Decomposition and Recoloring. *Computer Graphics Forum*, *38*(7), 11-22. https://doi.org/https://doi.org/10.1111/cgf.13812

Ward, G. (1994). A Contrast-Based Scalefactor for Luminance Display. In Paul Heckbert, Editor, Graphics Gems IV, Chapter VII.2, Pages 415–421. Aca- Demic Press, Boston, MA, USA, 1994. *Graphics Gems IV* (p. 415-421).

Wei Luo, Xiaogang Wang & Tang, X. (2011). Content-based photo quality assessment. *2011 International Conference on Computer Vision*, 2206-2213. https://doi.org/10.1109/ICCV.2011.6126498

Wisslar, V. (2012). *Illuminated Pixels*. Course Technology PTR.

Wright, W. D. (1929). A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, *30*(4), 141-164. https://doi.org/10.1088/1475-4878/30/4/301

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. & Bengio, Y. (2016, avril 19). *Show, Attend and Tell : Neural Image Caption Generation with Visual Attention*. arXiv : 1502.03044 [cs]. http://arxiv.org/abs/1502.03044

XYZ Color Matching Functions.svg - Wikipedia. (2009, mars 15). https://commons.wikimedia.org/wiki/File:CIE_1931_XYZ_Color_Matching_Functions.svg

Yan Ke, Xiaoou Tang & Feng Jing. (2006). The Design of High-Level Features for Photo Quality Assessment. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*, *1*, 419-426. https://doi.org/10.1109/CVPR.2006.303

Yang, C., Zhang, L., Lu, H., Ruan, X. & Yang, M.-H. (2013). Saliency Detection via Graph-Based Manifold Ranking. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3166-3173. https://doi.org/10.1109/CVPR.2013.407

Yee, Y. H. & Pattanaik, S. (2003). Segmentation and adaptive assimilation for detail-preserving display of high-dynamic range images. *The Visual Computer*, *19*(7-8), 457-466. https://doi.org/10.1007/s00371-003-0211-5

Yeganeh, H. & Wang, Z. (2013). Objective Quality Assessment of Tone-Mapped Images. *IEEE Transactions on Image Processing*, *22*(2), 657-667. https://doi.org/10.1109/TIP.2012.2221725

Yi, R., Tian, H., Gu, Z., Lai, Y.-K. & Rosin, P. L. (2023, mars 27). *Towards Artistic Image Aesthetics Assessment : A Large-scale Dataset and a New Method* (1). arXiv : 2303.15166 [cs]. http://arxiv.org/abs/2303.15166

Yoshida, A., Blanz, V., Myszkowski, K. & Seidel, H.-P. (2005, mars 18). Perceptual evaluation of tone mapping operators with real-world scenes. In B. E. Rogowitz, T. N. Pappas & S. J. Daly (Éd.). https://doi.org/10.1117/12.587782

Yuan, L. & Sun, J. (2012). Automatic Exposure Correction of Consumer Photographs. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato & C. Schmid (Éd.), *Computer Vision – ECCV 2012* (p. 771-785). Springer. https://doi.org/10.1007/978-3-642-33765-9_55

Yun, Y. K. & Lin, W. (2022). SelfReformer : Self-Refined Network with Transformer for Salient Object Detection.

Zhang, B., Niu, L. & Zhang, L. (2021, octobre 17). *Image Composition Assessment with Saliency-augmented Multi-pattern Pooling* (2). arXiv : 2104.03133 [cs]. http://arxiv.org/abs/2104.03133

Zhang, F.-L., Wang, M. & Hu, S.-M. (2013). Aesthetic Image Enhancement by Dependence-Aware Object Recomposition. *IEEE Transactions on Multimedia*, *15*(7), 1480-1490. https://doi.org/10.1109/TMM.2013.2268051

Zhang, H., Luo, Y., Qin, F., He, Y. & Liu, X. (2021). ELSD : Efficient Line Segment Detector and Descriptor.

Zhang, J., Yang, J., Fu, F. & Ma, J. (2024). Structural asymmetric convolution for wireframe parsing. *Engineering Applications of Artificial Intelligence*, *128*, 107410. https://doi.org/https://doi.org/10.1016/j.engappai.2023.107410

Zhang, Q., Nie, Y. & Zheng, W.-S. (2019, octobre 30). *Dual Illumination Estimation for Robust Exposure Correction*. http://arxiv.org/abs/1910.13688

Zhang, Y., Zhang, J. & Guo, X. (2019, mai 4). *Kindling the Darkness : A Practical Low-light Image Enhancer*. http://arxiv.org/abs/1905.04161

Zhao, K., Han, Q., Zhang, C.-B., Xu, J. & Cheng, M.-M. (2021). Deep Hough Transform for Semantic Line Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. https://doi.org/10.1109/TPAMI.2021.3077129

Annexes

Annexe A

Traitement d'image HDR

A.1 Données d'analyses

Source	URL
HDRIhaven	https ://hdrihaven.com/hdris/
Funt	https ://www2.cs.sfu.ca/ colour/data/funt_hdr/#DATA
Fairchild	http ://rit-mcsl.org/fairchild//HDRPS/HDRthumbs.html
HDRMAPS	http ://hdrmaps.com/freebies
Ward	http ://www.anyhere.com/gward/hdrenc/pages/originals.html
Freeskies	https ://joost3d.com/hdris/
Noemotion	http ://noemotionhdrs.net/hdrother.html
HDRI hub	https ://www.hdri-hub.com/hdrishop/freesamples
Openfootage	https ://www.openfootage.net/hdri-360-thumersbach-austria-winter-morning-at-the-lake/
Viz people	https ://www.viz-people.com/portfolio/free-hdri-maps/
HDRSID	https ://qualinet.github.io/databases/image/high_dynamic_range_specific_image_dataset_hdrsid/

Annexe B

Reconstruction de la composition d'une image : calcul des lignes de force

B.1 Reconstruction de la composition d'une image : Calcul des lignes de force

B.1.1 Méthode : Calcul des lignes de force de la composition

La distance entre l'ensemble de lignes $\mathcal{F}$ à l'ensemble de lignes $\mathcal{G}$ est calculé comme suit :

$$d_{LS}(\mathcal{F},\mathcal{G}) = \frac{1}{N_{\mathcal{F}}} \sum_{i=1}^{N_{\mathcal{F}}} d_{LS}(f_i,\mathcal{G}) \tag{B.1}$$

La distance entre l'ensemble de lignes $\mathcal{G}$ à l'ensemble de lignes $\mathcal{F}$ est calculé comme suit :

$$d_{LS}(\mathcal{G},\mathcal{F}) = \frac{1}{N_{\mathcal{G}}} \sum_{i=1}^{N_{\mathcal{G}}} d_{LS}(g_i,\mathcal{F}) \tag{B.2}$$

Où N_F est le nombre de lignes dans l'ensemble $\mathcal{F}$, et N_G est le nombre de lignes dans l'ensemble $\mathcal{G}$. Ensuite, nous utilisons la moyenne de deux distances comme distance finale entre deux ensembles, qui est calculée comme suit :

$$\mathcal{D}_{LS}(\mathcal{F},\mathcal{G}) = \frac{1}{2}\left(d_{LS}(\mathcal{F},\mathcal{G}) + d_{LS}(\mathcal{G},\mathcal{F})\right) \tag{B.3}$$

B.1.2 Résultats et discussion

Une compilation complète de tous les résultats comparatifs est accessible via l'URL suivante : https://projets.jrcandev.netlib.re/leadinglines/.

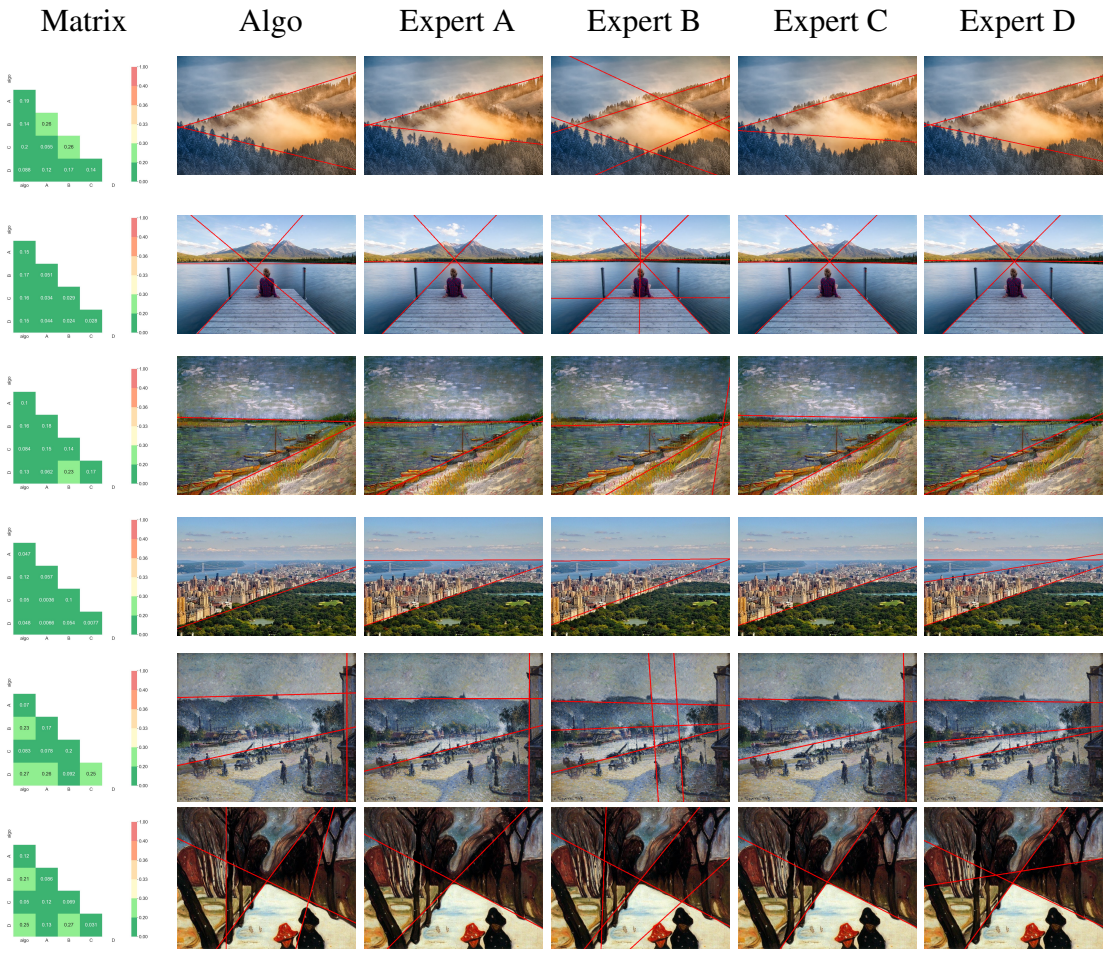


FIGURE B.1 – Plus de résultats et de matrices de comparaison. Ces résultats semblent soutenir l'hypothèse selon laquelle la production du modèle s'aligne étroitement sur celle des experts. Au-delà de la dimension quantitative, il est à noter que les propositions du modèle, même lorsqu'elles divergent du consensus moyen, ne semblent pas anormales et coïncident parfois avec les jugements de l'un des experts. Pour étudier cette proposition, d'autres examens impliquant un éventail plus large d'experts seraient nécessaires.

PRISE EN COMPTE DE L'ESTHÉTIQUE DANS LA GESTION DES GAMMES DE LUMINANCE DES IMAGES

Résumé

L'analyse des caractéristiques esthétiques d'images numériques permettent d'améliorer la qualité esthétique du contenu visuel. En analysant les caractéristiques esthétiques qui influencent la perception visuelle à travers les données de l'image, les ordinateurs peuvent effectuer des tâches telles que l'édition assistée d'image, l'amélioration de la qualité esthétique et le filtrage de la meilleure image. Cette thèse intègre l'analyse des caractéristiques esthétique des images avec l'imagerie à haute gamme dynamique (HDR). Nous prenons en compte les propriétés du HDR et les caractéristiques esthétiques lors du traitement des images HDR. L'objectif est de maximiser la préservation des caractéristiques esthétiques originales des images lors de l'ajustement des effets d'affichage HDR, afin d'atteindre l'expérience visuelle plus agréable. Dans cette thèse, nous proposons deux approches d'auto-ajustement des images HDR et une méthode de reconstruction des lignes de force de la composition.

Concernant l'auto-ajustement des images HDR, nous développons un modèle basé sur un réseau de neurones pour prédire la courbe d'ajustement des images HDR, et un modèle utilisant un réseau de neurones convolutifs pour estimer la valeur d'ajustement de l'exposition, en analysant les caractéristiques potentielles des images HDR. Ces deux méthodes consistent à améliorer automatiquement la perception de la qualité esthétique des images HDR sur des dispositifs d'affichage HDR. Elles le font en entraînant des réseaux de neurones à apprendre des paramètres d'édition d'experts à partir de jeux de données HDR. Afin d'analyser l'esthétique de la composition d'une image, nous proposons de reconstruire les lignes de force. Tout comme la couleur, les lumières ou le grain de l'image, les lignes de force font partie des caractéristiques esthétique qui doivent être analysées. La méthode proposée identifie les lignes de force implicites dans l'image par un algorithme de regroupement des lignes. Nous avons initialement mené une analyse de cohérence entre experts pour démontrer la faisabilité de notre méthode. Par ailleurs, nous proposons une métrique pour comparer les deux ensembles de lignes de force.

Mots clés : esthétique d'image, imagerie HDR, perception visuelle, apprentissage automatique

---

Abstract

Aesthetic analysis of digital images enhances the visual content's aesthetic quality. By analyzing aesthetic features that influence visual perception through image data, computers can perform tasks like assisted image editing, aesthetic quality enhancement, and filtering for the best image. This thesis merges aesthetic image analysis with high dynamic range (HDR) imaging. We consider both the properties of HDR and the aesthetic characteristics of images during HDR image processing. The aim is to maximize the preservation of the original aesthetic features of images when adjusting HDR image display results, thereby achieving a pleasant visual experience. In this thesis, we propose a composition leading lines reconstruction method and two HDR image auto-adjustment methods.

Regarding automatic adjustment of HDR images, we are developing a model based on a neural network to predict the adjustment curve of HDR images, and a model using a convolutional neural network to estimate the exposure adjustment value, by analyzing potential features of HDR images. Both methods automatically enhance the aesthetic quality perception of HDR images on HDR display devices by training neural networks to learn expert editing parameters from an HDR database. In order to analyze the aesthetics of image composition, we propose to reconstruct the leading lines of the image. Just like color, lighting, or the grain of the image, the leading lines are among the aesthetic features that need to be analyzed. The proposed method identifies implicit leading lines in the image through a line regrouping algorithm. We initially carried out an inter-expert consistency analysis to demonstrate the feasibility of our method. In addition, we propose a metric for comparing the two sets of leading lines.

Keywords: image aesthetics, HDR image, visual perception, machine learning

---