



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre : /M2/2023

Mémoire

présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **Images et Vie Artificielle**

Rééclairage d'images de portrait en utilisant une technique d'apprentissage basée sur un Hourglass network

Par :
RIGUET RYM

Soutenu le 02 juillet 2023, devant le jury composé de :

BOUGUETITICHE Amina	MCB	Président
BABAHENINI Mohamed Chaouki	Professeur	Rapporteur
BOUCETTA Mebarek	MAA	Examineur

Remerciements

Tout d'abord, je voudrais remercier **ALLAH S.W.T** pour ses bénédictions, qui m'ont donné force, patience et courage. Sans sa miséricorde, je n'aurais pas pu finir de développer ce projet. Je tiens à exprimer ma sincère gratitude à toutes les personnes qui ont contribué à la réalisation de ce mémoire. Leur soutien et leurs encouragements ont été d'une valeur inestimable tout au long de ce parcours.

Je voudrais tout d'abord remercier mon directeur de mémoire **Pr. BABAHENINI Mouhamed Chaouki**, pour son expertise, ses conseils éclairés et sa disponibilité. Sa patience et ses encouragements constants m'ont permis de surmonter les obstacles et de progresser dans mes recherches.

Je souhaite également exprimer ma reconnaissance envers les membres de mon jury **Dr. BOUGUETITICHE Amina** et **Dr. BOUCETTA Mbarek** pour leur temps, leurs suggestions précieuses et leurs commentaires constructifs lors de l'évaluation de mon mémoire.

Je tiens à remercier les doctorantes **Roumaissa** et **Lina** pour leur enseignement inspirant, leurs conseils avisés et leur dévouement à mon égard. Leur expertise dans le domaine a été une source d'inspiration pour moi et a contribué à façonner la qualité de mon travail.

Un grand merci à **ma famille** et mes amies proches **Zohra** et **Tourkia** pour leur amour, leur soutien indéfectible et leur compréhension pendant cette période intense de ma vie. Leur présence et leur encouragement constant m'ont aidé à rester motivé et à persévérer.

Je n'oublie pas de remercier mes collègues et camarades de classe **IVA** pour leur soutien mutuel, leurs discussions stimulantes et leur esprit d'équipe. Leurs idées et leurs réflexions ont été précieuses pour le développement de mes idées et de mes arguments.

Résumé

Notre travail introduit une approche de rééclairage de portrait basée sur le calcul physique de l'éclairage et permettant de générer un ensemble de données à grande échelle et de haute qualité appelé DPR, qui comprend diverses images de portrait capturées dans des environnements naturels. Nous formons alors un réseau neuronal convolutionnel (CNN) profond à l'aide de cet ensemble de données, ce qui lui permet de produire des images de portraits rallumés en prenant une image source et une configuration d'éclairage cible en entrée. La procédure d'apprentissage incorpore des techniques de régularisation pour améliorer les résultats générés et éliminer les artefacts causés par les méthodes de rééclairage basées sur la physique. En outre, nous utilisons de type Hourglass pour améliorer encore la qualité de l'image du portrait éclairé. Notre réseau entraîné peut rééclairer avec succès des images de portraits.

Pour évaluer les performances de la méthode, nous utilisons l'ensemble de données DPR proposé, ainsi que l'ensemble de données de portraits Flickr et l'ensemble de données Multi-PIE.

Mots-clés : Photo portrait, Apprentissage profond, Hourglass Network, Rééclairage, dataset DPR.

Abstract

Our work introduces a physical illumination computation-based portrait relighting approach to generate a large-scale, high-quality dataset called DPR, which includes various portrait images captured in natural environments. We then train a deep convolutional neural network (CNN) using this dataset, enabling it to produce re-lit portrait images by taking a source image and a target lighting configuration as inputs. The learning procedure incorporates regularisation techniques to improve the generated results and eliminate artifacts caused by physics-based relighting methods. In addition, we use Hourglass to further enhance the image quality of the illuminated portrait. Our trained network can successfully relight portrait images.

To evaluate the method's performance, we use the proposed DPR dataset, the Flickr portrait dataset, and the Multi-PIE dataset.

Keywords : Portrait photo, Deep learning, Hourglass Network, Relighting, DPR dataset

Table des matières

1	Éclairage des scènes virtuelles et Apprentissage profond	17
1.1	Introduction	17
1.2	Partie A : Éclairage en informatique graphique	18
1.2.1	Modèles d'éclairage	19
1.2.1.1	Le modèle diffus idéal ou modèle Lambertien	19
1.2.1.2	Modèle spéculaire : Modèle de Phong	20
1.2.2	Éclairage global	22
1.2.3	Éclairage direct	23
1.2.4	Éclairage indirecte	23
1.2.5	Problème de résolution de l'équation de transport de lumière	24
1.2.6	Les techniques d'éclairage global	25
1.2.6.1	Lancer de rayons (Ray tracing)	25
1.2.6.2	Ray casting	28
1.2.6.3	Radiosité	29
1.2.6.4	Path tracing	33
1.2.6.5	Monte Carlo	34
1.2.6.6	Photon mapping	34
1.3	Partie B : Apprentissage profond	35
1.3.1	Motivations	35
1.3.2	Apprentissage automatique	36
1.3.3	Les réseaux de neurones artificielles	38
1.3.4	Types d'apprentissage	39

1.3.4.1	Apprentissage supervisé	39
1.3.4.2	Apprentissage non supervisé	42
1.3.4.3	Apprentissage par renforcement	43
1.4	Apprentissage profond	43
1.4.1	Réseaux Feedforward	44
1.4.1.1	Les réseaux biologiques	44
1.4.1.2	Les couches	45
1.4.2	Architectures de réseaux neuronaux	46
1.4.2.1	Réseaux neuronaux convolutifs(CNN)	46
1.5	Conclusion	50
2	Rééclairage basé sur l'apprentissage profond à partir d'une seule photo de portrait	51
2.1	Introduction et motivations	51
2.2	Le rendu neuronal : Présentation et définitions	52
2.3	Comparaison entre la technique qui utilise le réseau Hourglass et les autres techniques proposées pour le ré éclairage d'images de portrait	54
2.4	Classification des techniques de ré-éclairage des photos de portrait	55
2.4.1	Techniques de ré éclairage basées sur les images de références(techniques d'optimisation)	55
2.4.1.1	Images de quotient (ratio images) :	55
2.4.1.2	Le ratio de la carte d'environnement de radiance :	56
2.4.1.3	Technique de transfert de réflectance	57
2.4.2	Techniques de ré éclairage basées sur l'apprentissage profond	58
2.4.2.1	Réseaux neuronaux (GAN, CNN, etc.)	58
2.4.2.2	Apprentissage de l'éclairage à partir de données d'entraînement	59
2.4.3	Techniques de rendu inverse des images de portrait	59
2.4.3.1	Modèle 3D morphable (3DMM)	60
2.5	Estimation normale	63

2.5.1	Raffinement normale basé sur 3D Dense Face Alignment (3DDFA)	63
2.5.2	Raffinement normal basé sur ARAP (As-Rigid-As-Possible) . . .	63
2.6	les Réseaux antagonistes génératifs	65
2.6.1	Architecture générale des GANs	66
2.6.1.1	Générateur	67
2.6.1.2	Discriminateur	68
2.6.2	Entraînement de GANs	69
2.6.3	Applications des GANs	70
2.6.3.1	Générer de nouvelles photos	70
2.6.3.2	Coloriser des images en noir et blanc	71
2.6.3.3	Créer des œuvres d'art	71
2.6.3.4	Traduction d'images	72
2.7	Conclusion	72
3	Conception d'une approche de rééclairage de portrait à base de réseau Hourglass	73
	Hourglass	73
3.1	Introduction	73
3.2	Architecture de réseau Hourglass : Généralités	74
3.2.1	Les blocs de construction	75
3.2.1.1	Couches convolutives	75
3.2.1.2	Couches sous-échantillonnage (downsampling layers) . .	76
3.2.1.3	Couches sur-échantillonnage (upsampling layers)	76
3.2.1.4	Sauts de connexion	77
3.3	Proposition d'une application utilisant le réseau Hourglass pour le ré-éclairage des images portrait basé sur l'apprentissage profond	78
3.3.1	Génération de base de données (Deep Portrait Relighting Dataset "DPR")	78
3.3.2	Architecture de réseau Hourglass	80
3.3.3	Entraînement de réseau Hourglass	83

3.4	LSGANs(Réseaux adversariaux génératifs à moindres carrées)	87
3.5	Conclusion	89
4	Mise en œuvre et résultats	90
4.1	Introduction	90
4.2	Préparation de base de données (DPR)	90
4.3	Aperçu de la mise en œuvre	93
4.4	Entraînement de Hourglass	94
4.5	Dispositif expérimental	95
4.5.1	Configuration du matérielle	95
4.5.2	Configuration du logicielle	95
4.5.2.1	PyTorch	96
4.5.2.2	TensorFlow	96
4.5.2.3	Keras	97
4.5.2.4	OpenCV	97
4.5.2.5	Shtools	97
4.6	Mesures d'évaluation	98
4.7	Résultats	100
4.7.1	Résultats sur des images frontales	100
4.7.2	Résultats sur des images non frontales	102
4.7.3	Résultats sur des images difficiles	102
4.7.4	Résultats de données DPR de haute résolution	104
4.7.5	Résultats de données de portraits de Flickr	104
4.7.6	Mauvais Résultats	105
4.8	Conclusion	106
5	Conclusion générale	107

Table des figures

1.1	Le principe de modèle lambert[14]	19
1.2	les trois parties de l'équation : ambiante, diffuse et spéculaire	20
1.3	L'interaction de la lumière avec la surface	21
1.4	Exemple d'application le modèle Phong sur une sphère vert avec un éclairage blanc	22
1.5	Le principe de l'éclairage directe et indirecte[5]	24
1.6	La BRDF	24
1.7	Principe du lancer de rayon	26
1.8	Exemple de rendu avec la méthode de lancer de rayon	28
1.9	Principe de ray casting[1]	28
1.10	Interprétation géométrique.	30
1.11	Schéma explicatif de principe général de la radiosité	31
1.12	Exemple de la division d'une scène	31
1.13	Gathering vs. Shooting	32
1.14	Principe de path tracing	33
1.15	Principe de Monte Carlo	34
1.16	Deux passes de l'algorithme de photon mapping : Pass1 : distribution des photons. Pass 2 : rendu par estimation de luminance.	35
1.17	(a) : Structure d'un réseau de neurone biologique ,(b) le cerveau humain[9]	38
1.18	Exemple d'apprentissage supervisé	40
1.19	Schéma explicatif des étapes de l'apprentissage supervisé	41

1.20	(a) Architecture de réseau neuronal simple , (b) Architecture simple du réseau neuronal profond (DNN)[6]	45
1.21	Exemple des filtres[22]	48
1.22	Exemple des opérations de couche de regroupement[27]	48
1.23	flatterie	49
1.24	Couches complètement connectées	49
2.1	Tableau comparatif des techniques proposées	54
2.2	L'image du milieu est l'entrée. La séquence montre les résultats synthétisés de la rotation 180° de l'environnement d'éclairage. [25]	57
2.3	Exemple de Technique de transfert de réflectance [18]	58
2.4	Exemple d'applique 3DMM sur des images 2D [4]	61
2.5	Le processus de3DMM [16]	62
2.6	Le processus de raffinement normal basé sur ARAP.[28]	64
2.7	Comparaison entre 3DDFA et ARAP. [28]	64
2.8	Architecture générale des GANs[2]	66
2.9	Exemple d'architecture du modèle de réseau adversarial génératif [26]	70
2.10	Visages générés sur Cette personne n'existe pas [26]	70
2.11	modèle de DeOldify [26]	71
2.12	Œuvres d'art générées grâce aux GAN [26]	71
2.13	Traduction d'images avec Cycle GAN [26]	72
3.1	Illustration d'un module "Hourglass" unique [15]	74
3.2	Schéma représente Les blocs de construction de l'encodeur et décodeur de réseau Hourglass[24]	75
3.3	Schéma explicatif de principe de couches sous-échantillonnage (downsampling layers)[3]	76
3.4	Schéma explicatif de principe de couches sur-échantillonnage (upsampling layers). [3]	77
3.5	Exemples de la base CelebA-HQ [26]	79

3.6	Détection de points de repère sur un visage	80
3.7	La structure de réseau Hourglass	80
3.8	Schéma explicatif du contenu de réseau Hourglass	81
3.9	(a) montre l'image d'entrée, (b), (d) et (f) sont des images générées en utilisant L_I , $L_I + L_{GAN}$ et $L_I + L_{GAN} + L_{GAN}$. images générées en utilisant L_I , $L_I + L_{GAN}$ et $L_I + L_{GAN} + L_f$ respectivement. Lf respectivement ; (c), (e) et (g) sont la région du rectangle rouge de (b), (d) et (f) respectivement. Notez le bord au milieu du bruit généré par L_I	85
3.10	(a) sortie du réseau Hourglass, (b) région rectangulaire de (a), (c) sortie du réseau Hourglass. région rectangulaire de (a), (c) sortie du réseau Hourglass avec (d) région rectangulaire de (c). Nous augmentons l'intensité. Augmentation de l'intensité des pixels de (b) et (d) pour une meilleure visualisation.	86
3.11	Schéma explicatif d'un patch GAN	88
4.1	Exemples d'application 3DFFA	92
4.2	Détails sur chaque bloc de réseau Hourglass.[28]	94
4.3	Résultats comparatives d'évaluation	99
4.4	Résultats sur des images frontales	101
4.5	Résultats sur des images non frontales	102
4.6	Résultats sur des images difficiles.	103
4.7	Résultats sur l'ensemble de données DPR.	104
4.8	Résultats sur l'ensemble de données de portraits de Flickr.	105
4.9	Mauvais résultats.	106

Liste des tableaux

4.1	Étude d'ablation sur l'ensemble de données Multi-PIE	99
4.2	Ensemble de données multi-PIE d'évaluation	100

Introduction générale

L'informatique graphique est un domaine de l'informatique qui se concentre sur la création, la manipulation et la représentation graphique d'objets virtuels et d'images. La modélisation et le rendu sont les deux processus fondamentaux dans le processus de génération de mondes virtuels dans l'informatique graphique. Dans l'informatique graphique et la réalité virtuelle, ces processus sont essentiels pour créer des photos virtuelles réalistes et immersives.

L'éclairage d'une scène virtuelle est la simulation du comportement des sources de lumière dans un environnement virtuel. Cela implique de prendre en compte des aspects tels que la position, l'intensité et la couleur des sources de lumière, ainsi que la façon dont la lumière interagit avec les matériaux et les surfaces. Afin de générer des effets d'éclairage réalistes et de représenter efficacement l'interaction entre la lumière et les surfaces, de nombreux modèles d'illumination existent dans la littérature, ces modèles résolvent une équation de rendu de Kajiya avec des méthodes complexes et souvent très gourmandes en ressources informatiques. Depuis 2015 et l'avènement des techniques basées sur le deep learning, de nombreuses théories intégrant ce dernier au calcul de l'éclairage en vue le jour et ont permis d'améliorer considérablement d'une part la qualité du rendu de la scène et d'autre part le temps de calcul.

On s'intéresse dans ce mémoire au problème particulier de rééclairage d'un portrait qui représente un processus captivant qui permet aux photographes et aux artistes de manipuler les conditions d'éclairage d'une image, transformant ainsi l'ambiance, l'atmosphère et l'esthétique générale.

Traditionnellement, l'obtention de tels effets nécessitait des réglages d'éclairage avan-

cés pendant la séance photo ou des techniques de post-traitement approfondies. Cependant, grâce aux récentes avancées en matière d'apprentissage profond et de vision par ordinateur, une nouvelle approche a vu le jour : l'utilisation de réseaux neuronaux convolutionnels en sablier (ou Hourglass Networks) pour le rééclairage (relighting) des portraits.

Dans notre mémoire de fin d'études nous traitant le ré-éclairage d'un portrait à l'aide d'un réseau CNN en sablier, l'objectif principal étant de simuler différentes conditions d'éclairage et d'ajuster l'éclairage du visage du sujet, en améliorant l'attrait visuel et en transmettant une tonalité émotionnelle souhaitée. En entraînant le réseau sur un vaste ensemble de données de portraits capturés dans différents scénarios d'éclairage, il apprend à comprendre la relation complexe entre les caractéristiques du visage et les modèles d'éclairage.

Pour arriver à nos fins, nous avons organisé notre mémoire comme suit :

Le chapitre 1 est composée de deux parties, la première est consacrée à la description du modèle d'illumination locale de Phong, puis à la définition de la formulation mathématique de l'équation de rendu (composantes directe et indirecte), ainsi qu'à un aperçu sur les algorithmes de résolution de cette équation. Quant à la deuxième partie, elle est dédié à l'apprentissage profond et à la présentation des définitions, des différents types d'apprentissage et des réseaux de neurones convolutionnels (CNN).

Dans le chapitre 2, nous avons présenté une classification des techniques existantes dans la littérature sur le ré éclairage des images de portrait, nous avons mis l'accent essentiellement sur les méthodes basées sur l'apprentissage et plus spécifiquement nous avons insisté sur les réseaux GAN (Generative Adversarial Networks) qui était la base de notre investigation de départ.

Le chapitre 3 est dédié à l'exposé général de l'application proposé, pour cela nous avons défini les objectifs à atteindre, l'outil principal utilisé qui consiste en un réseau Hourglass, et l'architecture générale de l'application permettant de résoudre le problème de la modification des conditions d'éclairage d'un portrait sans utiliser de nombreuses images ou des configurations compliquées.

Dans le chapitre 4 , nous décrivons les outils (matériels et logiciels) ainsi que les plateformes que nous avons utilisées, puis le diagramme de classe et les techniques que nous avons intégrées, enfin nous présentons les résultats, ainsi qu'une analyse quantitative et qualitative sur les images produites par notre application.

Enfin, nous clôturons ce mémoire par une conclusion et des perspectives.

Chapitre 1

Illumination des scènes virtuelles et Apprentissage profond

1.1 Introduction

On rappelle que l'objectif de la synthèse d'images consiste à créer des images aussi exactes et réalistes que possible. Cela se fera en deux étapes : la modélisation géométrique et la visualisation.

La modélisation géométrique consiste à décrire les modèles des objets qui constituent l'environnement. Ces objets peuvent être représentés d'une manière simple, par des points, lignes ou triangles. Ils peuvent aussi se représenter mathématiquement par des surfaces implicites (plan, sphère, cylindre...) plus difficiles à manipuler mais paramétrables.

La visualisation ou le rendu concerne le calcul de la lumière et l'affichage du résultat, il nécessite donc d'une part le respect des caractéristiques physiques des matériaux et de l'environnement et des phénomènes de transport de la lumière pour une simulation correcte de la distribution de lumière, et d'autre part doit tenir compte des effets statiques comme la forme, la couleur, les textures et la position des objets et aussi des effets dynamiques qui ont une influence directe sur les effets visuel comme le changement de la position des objets ou de la caméra. [8]

Le Deep Learning, ou apprentissage profond, est un sous-ensemble du Machine Learning, ou apprentissage automatique, basé sur des réseaux neuronaux artificiels. Le processus d'apprentissage est qualifié de profond vu que la structure des réseaux neuronaux artificiels se compose de plusieurs couches d'entrée, de sortie et masquées. Chaque couche contient des unités qui transforment les données d'entrée en informations que la couche suivante peut utiliser une tâche prédictive spécifique. Grâce à cette structure, une machine est capable d'apprendre au travers de son propre traitement de données [11].

L'objectif de ce chapitre est de donner les définitions nécessaires de l'illumination en informatique graphique et de l'apprentissage profond pour la compréhension de la suite du mémoire, mais surtout de démontrer la complexité des calculs de l'illumination et la nécessité d'intégrer l'apprentissage pour contourner le problème.

1.2 Partie A : Illumination en informatique graphique

L'éclairage dans les graphiques informatiques se réfère à la façon dont la lumière interagit avec les objets 3D virtuels pour créer des images réalistes ou stylisées. L'objectif de l'éclairage est de simuler les propriétés de la lumière du monde réel et de rendre les objets virtuels visuellement convaincants. Un problème typique avec l'éclairage dans les graphiques informatiques est l'équilibre entre le réalisme et l'efficacité. Les approches d'éclairage les plus précises peuvent être assez coûteuses en termes de puissance informatique, ce qui entraîne des temps de rendu extrêmement longs, en particulier pour les scénarios complexes. L'illumination est un domaine vaste et complexe de l'informatique graphique qui continue d'évoluer avec de nouvelles méthodes et algorithmes pour rendre les images virtuelles de plus en plus réalistes.

1.2.1 Modèles d'éclairage

1.2.1.1 Le modèle diffus idéal ou modèle Lambertien

Le modèle d'éclairage de Lambert implique que la lumière sur une surface mat reflète uniformément dans toutes les directions. Il exclut les réflexions spéculaires, les réfractions et d'autres phénomènes optiques plus complexes. En conséquence, ce modèle est une approximation simplifiée de la réalité, mais il est couramment employé en raison de sa simplicité et de son efficacité.[11][19]

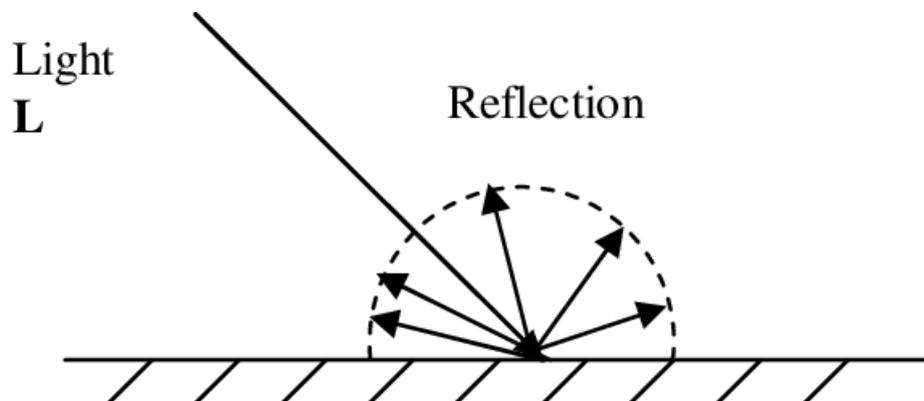


FIGURE 1.1 – Le principe de modèle lambert[14]

La loi de Lambert, communément connue sous le nom de loi de la réflexion diffuse, est utilisée pour calculer l'illumination de Lambert. L'intensité de la lumière reflétée par une surface est exactement proportionnelle à l'angle entre la direction de la lampe incidente et la surface normale, selon cette règle. Plus précisément, l'intensité de la lumière réfléchie est proportionnelle au produit du cosinus de l'angle et à l'intensité de la luminosité. La loi de Lambert peut être exprimée mathématiquement comme suit :

$$I_d = K_d i_{m,d} \cos \theta$$

où $\cos \theta = \frac{\vec{N} \cdot \vec{L}}{\|\vec{N}\| \|\vec{L}\|} = \vec{N} \cdot \vec{L}$, \vec{N} et \vec{L} sont des vecteurs unitaires.

L'illumination de Lambert basée sur la réflexion diffuse de la lumière sur les surfaces mates. Il est une base importante dans les techniques de rendu et peut être combinée avec d'autres modèles d'illumination plus avancés pour obtenir des résultats plus réalistes et complexes.

1.2.1.2 Modèle spéculaire : Modèle de Phong

L'ombrage de Phong a été introduit par Bui Tuong Phong en 1973 .cette technique est encore fréquemment utilisé aujourd'hui.[11][19]

Le modèle d'ombrage Phong prend en compte trois composantes de l'éclairage : éclairage ambiant, diffus et spéculaire. Chaque composant contribue à la couleur finale d'un point de surface en fonction des qualités du matériau de surface et des sources de lumière dans la scène.

1. L'éclairage ambiant : ce composant représente l'éclairage continu et homogène de la scène. Il n'est pas affecté par l'emplacement ou l'orientation de la surface. L'éclairage ambiant influence la couleur de base globale de la surface.
2. L'éclairage diffus : est défini comme la lumière qui est répartie uniformément dans toutes les directions par une surface. Il est déterminé par l'angle formé entre la surface normale et la direction de la source de lumière.
3. L'éclairage spéculaire : Il est déterminé par l'angle formé entre la direction de la lumière réfléchie et la direction de l'observateur. Les surfaces lisses et brillantes ont des accents spéculatifs plus intenses.

$$I_p = k_a i_a + \sum_{m \in \text{lights}} \left(k_d (\hat{L}_m \cdot \hat{N}) i_{m,d} + k_s (\hat{R}_m \cdot \hat{V})^\alpha i_{m,s} \right)$$

Cette équation peut sembler complexe,elle est composée de trois parties principales : ambiante, diffuse et spéculaire. L'indice "m" fait référence au nombre de lumières dans notre scène.

Ambient Diffusion Specular

$$I_p = \boxed{k_a i_a} + \sum_{m \in \text{lights}} \left(\boxed{k_d (\hat{L}_m \cdot \hat{N}) i_{m,d}} + \boxed{k_s (\hat{R}_m \cdot \hat{V})^\alpha i_{m,s}} \right)$$

FIGURE 1.2 – les trois parties de l'équation : ambiante, diffuse et spéculaire

Pour le composant ambiante :

- La valeur k_a : est la constante de réflexion ambiante, le rapport de réflexion du terme ambiant présent en tous les points de la scène rendue.

- La valeur i_a : contrôle l'éclairage ambiant et est parfois calculée comme une somme des contributions de toutes les sources lumineuses.

pour le composant diffuse :

- La valeur K_d : est la constante de réflexion diffuse, le rapport de réflexion du terme diffus de la réflectance lambertienne de la lumière entrante.
- La valeur $(L_m \cdot N)$: est la réflexion diffuse qui représente la réflexion lambertienne. 1.3
- La valeur i_d : est l'intensité d'une source de lumière dans la scène, définie par une valeur de couleur dans ce cas.

pour le composant spéculaire :

- La valeur K_s : est la constante de réflexion spéculaire, le rapport de réflexion du terme spéculaire de la lumière entrante.
- Le vecteur R_m : est la direction que prendrait un rayon de lumière parfaitement réfléchi s'il rebondissait sur la surface. 1.3

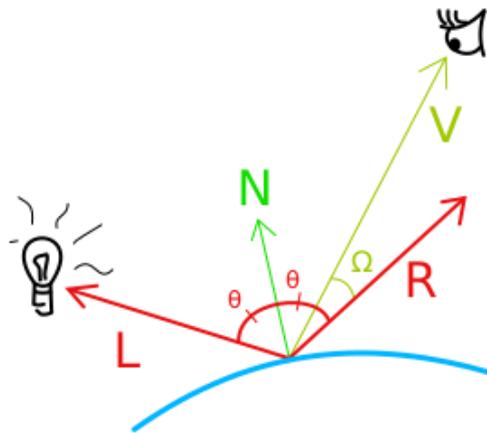


FIGURE 1.3 – L'interaction de la lumière avec la surface

Le modèle de réflexion de Phong calcule la direction du rayon réfléchi R_m à l'aide de la formule suivante :

$$\hat{R}_m = 2 \left(\hat{L}_m \cdot \hat{N} \right) \hat{N} - \hat{L}_m$$

Comme mentionné précédemment, l'indice "m" fait référence au nombre de lumières dans la scène.

Le petit chapeau , au-dessus de chaque lettre signifie que nous devons utiliser la

version normalisée de chaque vecteur.

Le vecteur L_m : fait référence à la direction de la lumière .

Le vecteur N : fait référence à la surface normale .

On ajoute un composant ambiant et spéculaire pour compléter le modèle de réflexion de Phong et avec une seule source de lumière.

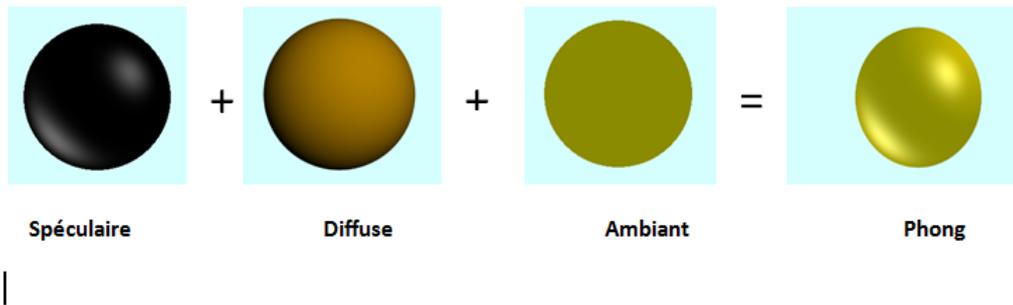


FIGURE 1.4 – Exemple d’application le modèle Phong sur une sphère vert avec un éclairage blanc

Le terme illumination locale est un terme simplifié qui suppose que l’illumination d’un point ou d’un élément de surface dépend de la propriété d’un élément de surface ainsi que de la lumière incidente provenant des sources. Or l’illumination locale est incapable de générer des images vraiment réalistes vu qu’elle tient compte uniquement de la lumière en provenance directement des sources lumineuses. En réalité, les surfaces reçoivent aussi de la lumière indirecte résultant des inter-réflexions avec d’autres surfaces.

1.2.2 Illumination globale

Illumination globale, c’est l’interaction de la lumière entre différents objets d’une scène. En bref, la lumière est réfléchiée par des objets, tout comme cela se produit dans la réalité. La lumière atteint également des zones (comme les coins) que des sources normales de lumière n’éclaireraient pas (dans l’image d’exemple ci-dessus, vous pouvez observer de nombreuses zones noires, hors de portée de toute source de lumière dans la scène rendue sans Illumination globale). Avec l’Illumination globale, quasiment toutes les parties d’un objet peuvent réfléchir la lumière.

1.2.3 Éclairage direct

La lumière provenant d'une source est réfléchié directement sur une surface, puis renvoyée à l'appareil photo dans le cadre d'un éclairage direct voir la figure 1.5 . Il s'agit du paradigme traditionnel pour calculer la lumière en 3D, comme le montre le modèle de Phong 1.2.1.2, La lumière est ensuite divisée en trois composantes : la lumière ambiante (qui représente un éclairage global constant), la lumière diffuse (qui représente la diffusion de la lumière sur la surface) et la lumière spéculaire (qui représente la réflexion de la surface).

1.2.4 Éclairage indirecte

Dans l'éclairage indirecte [17], on considère que chaque surface qui reçoit de la lumière va à son tour en renvoyer. Il n'y a donc plus qu'une seule source de lumière, mais une infinité, correspondant à chaque surface qui renvoie de la lumière voir la figure 1.5. Pour réaliser le rendu d'une scène, il faut donc procéder de la façon suivante :

1. pour chaque surface S_1 visible par la caméra ;
2. pour chaque surface S_2 visible depuis la surface S_1 ;
3. calculer la lumière reçue par la surface S_2 provenant de la source ;
4. calculer la lumière renvoyée par la surface S_2 vers la surface S_1 ;
5. calculer la lumière reçue par la surface S_1 provenant de la surface S_2 ;
6. calculer la lumière totale renvoyée par la surface S_1 vers la caméra ;

Chaque étape doit prendre en compte plusieurs paramètres. Pour (3), il faut par exemple prendre en compte la distance entre la source et la surface S_2 et l'atténuation de la lumière en fonction de la distance. Pour (4), il faut prendre en compte les caractéristiques de la surface et sa capacité à réfléchir la lumière. Pour (5), il faut prendre en compte la distance entre les surfaces S_1 et S_2 et l'angle d'incidence. Pour (6), les capacités de réflexion de la surface S_1 et la distance entre la surface S_1 et la caméra.

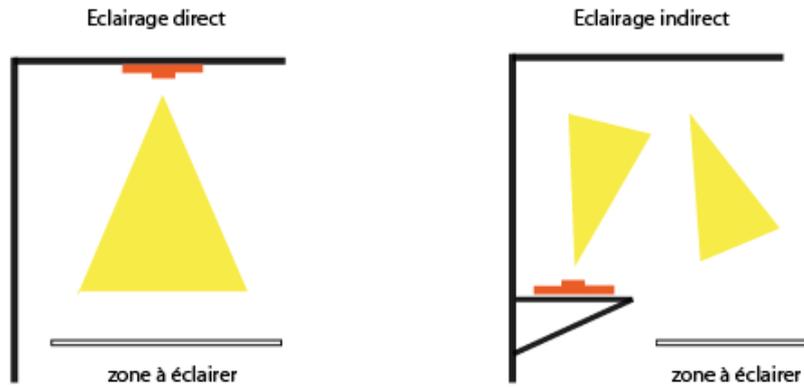


FIGURE 1.5 – Le principe de l'éclairage directe et indirecte[5]

1.2.5 Problème de résolution de l'équation de transport de lumière

L'équation de transport de lumière ou autrement dit l'équation de rendu telle qu'elle est présentée par Kajiya 1986 est formulée comme suit :

$$L_0(x, \omega_0) = L_e(x, \omega_0) + \int_{H(x)} L_i(x, \omega_i) \cdot f_r(x, \omega_i \rightarrow \omega_0) \cdot \cos \theta_i \cdot d\omega_i$$

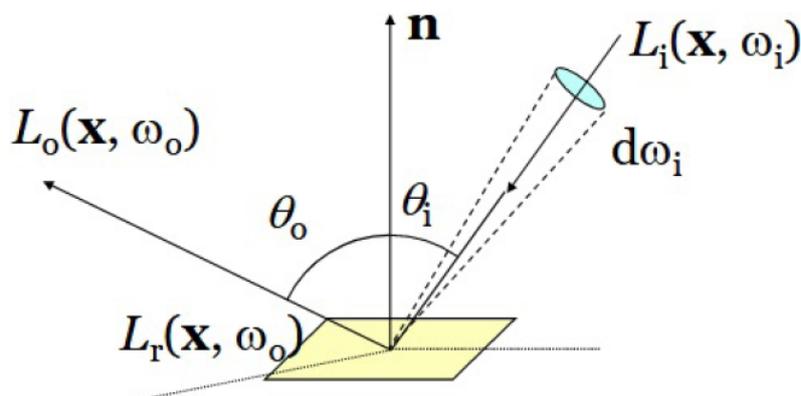


FIGURE 1.6 – La BRDF

Cette équation définit l'éclairage indirect par la luminance en un point 'x' dans la scène comme étant l'intégrale de la luminance incidente $L_i(x, \omega_i)$ provenant de toutes

les directions multipliée par une fonction BRDF $\cdot f_r(x, \omega_i \rightarrow \omega_o)$ et le cosinus de l'angle d'incidence θ_i

Cette équation peut être résolue par une méthode d'éléments finis en remplaçant l'intégrale par une sommation (méthode de radiativité). Elle peut être résolue aussi par des méthodes stochastiques comme la méthode de Monte Carlo. A part ces deux méthodes (physiques) qui calculent l'illumination globale en résolvant l'équation de rendu, Il existe d'autres méthodes non physiques où le de départ n'est pas l'équation de rendu, mais elles focalisent sur le calcul d'éclairage d'une manière approchée et rapide par exemple (méthode de Photo mapping).

1.2.6 Les techniques d'illumination globale

1.2.6.1 Lancer de rayons (Ray tracing)

La méthode du ray-tracing [17] commence par l'émission d'un rayon à travers chaque pixel de l'image à partir de l'œil virtuel de l'observateur (ou de la caméra). Ce rayon se déplace ensuite dans la scène virtuelle, interagissant avec les éléments qu'il rencontre.

Lorsque le rayon touche un objet, il recueille des données telles que la couleur, la texture, la réflexion, la transmission et les caractéristiques optiques de l'objet. Ces informations sont utilisées pour déterminer l'aspect de l'objet à cet endroit précis.

Cette approche de traçage inverse de l'œil/caméra à la source de lumière a été choisie car elle est considérablement plus efficace que la traçabilité de tous les rayons lumineux émis par des sources lumineuses dans de nombreuses directions.

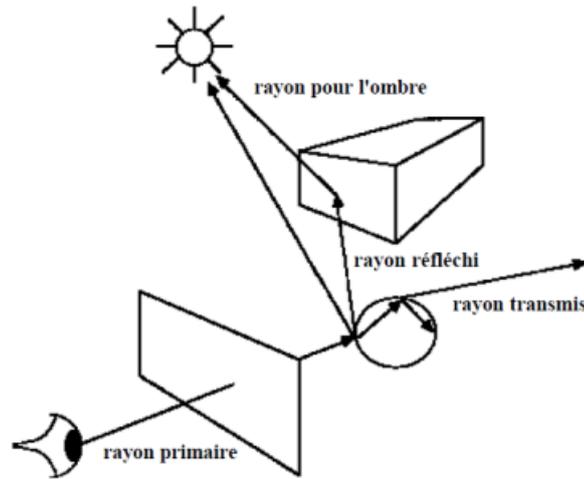


FIGURE 1.7 – Principe du lancer de rayon

L'algorithme de Lancer de rayons est principalement cherche le point de la scène rencontré par un rayon et calcule le flux lumineux le long de ce rayon. La fonction s'appelle elle-même récursivement dans le cas où le point rencontré est sur une surface réfléchissante.

Ce processus récursif s'arrête dans l'un des cas suivants :

1. soit le rayon quitte la scène
2. soit le rayon percute une surface ni spéculaire ni transparente
3. soit la contribution du rayon devient négligeable
4. ou encore une profondeur maximale de la récursivité est atteinte.

L'algorithme du lancer de rayon est très puissant et permet de traiter pratiquement tous les aspects du réalisme. Malheureusement, les calculs d'intersections sont très coûteux en temps, car ils sont étroitement liés au nombre de rayons lancés, et seules des machines puissantes peuvent se permettre de tels calculs dans un temps raisonnable. Il faut donc trouver des méthodes d'optimisation. Pour cela, il existe plusieurs stratégies envisageables ; augmenter la vitesse de traitement de ces intersections, trouver des moyens pour tester plus rapidement s'il y a possibilité d'intersection, diminuer le nombre de rayons lancés, ou encore utiliser des rayons plus complexes que des droites (faisceaux ou cônes, par exemple).

Algorithm 1 Algorithme global de lancer de rayon

Calcul des intersections avec tous les objets de la scène.

if intersection **then**

 on stocke la distance t (analogue à un temps de parcours) entre le point de départ et le point d'intersection.

end if

if aucun objet n'est rencontré **then**

 Renvoi de l'éclairement du fond.

else

 L'intersection la plus proche est retenue, celle dont la valeur de t est la plus petite.

 Le flux est initialisé à zéro.

while Pour chaque source ponctuelle **do**

 Envoi d'un rayon vers la source.

 Les intersections de ce rayon avec les objets de la scène sont recherchées.

if aucune intersection n'est trouvée **then**,

 la contribution de cette source au flux lumineux est ajoutée.

end if

if la surface est réfléchissante **then**

 Calcul du rayon réfléchi par la loi de Descartes de la réflexion spéculaire.

 La fonction lancer rayon est appelée (appel récursif)

 Le résultat est ajouté au flux, avec coefficient de réflexion.

 Renvoi du flux.

end if

end while

end if

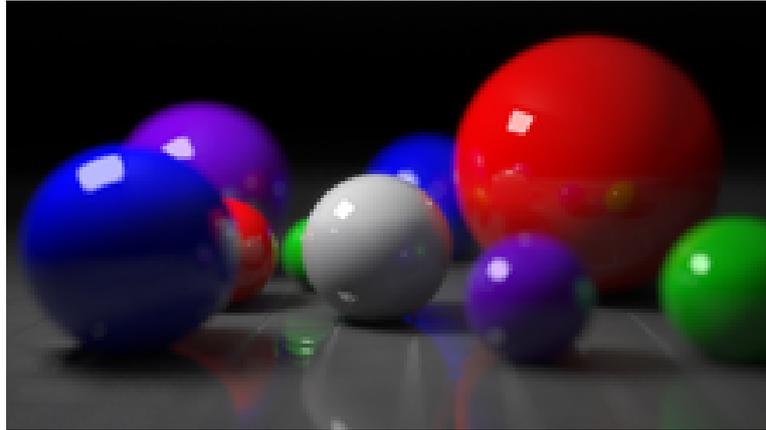


FIGURE 1.8 – Exemple de rendu avec la méthode de lancer de rayon

Dans cette figure On remarque que les reflets et la lumière sont réalistes.

1.2.6.2 Ray casting

Le ray casting [17] est la procédure d'un algorithme de traçage des rayons (ray tracing) dans lequel un ou plusieurs rayons de la caméra (localisation de l'œil) sont tirés à travers chaque pixel dans un plan d'image et ensuite testés pour déterminer si les rayons traversent des primitifs (triangles) dans la scène. Lorsqu'un rayon passe à travers un pixel et sort dans la scène 3D, la distance le long du rayon de l'origine (caméra ou point d'œil) au primitif est calculée, et les données de couleur du primitif contribuent à la couleur finale du pixel. Le faisceau peut également éclater et affecter d'autres objets, capturant des informations de couleur et d'éclairage.

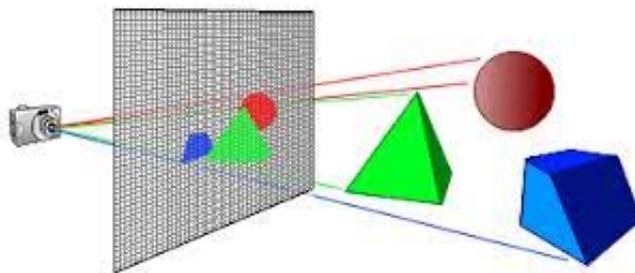


FIGURE 1.9 – Principe de ray casting[1]

1.2.6.3 Radiosité

La radiosité [21] est une méthode d'estimation de l'éclairage global qui est basée sur la physique du transfert de rayonnement lumineux entre les surfaces élémentaires dans une image tridimensionnelle. Il est basé sur des considérations d'équilibre énergétique, de sorte que l'inter-réflexion entre les surfaces diffuses est soigneusement prise en compte. L'Université de Cornell a été la première à utiliser la radiation (Cohen et Greenberg 1985). Malgré le fait qu'il soit assez coûteux en termes de calcul, elle a cependant l'avantage d'être indépendante du point de vue (position de l'observateur), et donc le calcul illumination peut se faire en amont (comme phase de pré traitement), ensuite une navigation dans des environnements virtuels (avec probablement un casque de réalité virtuelle) peut avoir lieu en changeant la position de l'observateur (sans refaire les calculs d'éclairage). L'expression des équations de radiosité est faite sur la base de l'hypothèse que les phénomènes d'émission et de réflexion sont dus à des réflecteurs parfaits, donc la direction des rayons est perdue après toute réflexion. On commence par définir les termes suivants :

1. Radiosité (B) : c'est la quantité de base qu'on cherche calculer pour chaque surface ; c'est une énergie par unité de surface et de temps.
2. Réflectivité R : c'est la fraction de lumière réfléchie sur une surface (elle est normalisée entre 0 et 1). On note que l'absorption vaut (1- R).
3. Facteur de forme (F) : c'est la fraction de la lumière quittant une surface et arrivant à une autre (elle est normalisée entre 0 et 1).
4. Émission (E) : c'est l'énergie émise par la surface (comme pour le cas d'une source lumineuse). Elle est exprimée par unité de surface et de temps.

Pour calculer la radiosité d'un élément de surface dA_i (voir la figure 1.10), l'intensité de cet élément va dépendre de toute la lumière qu'il émet directement plus la lumière qui est réfléchie. Cette lumière réfléchie dépend de la lumière quittant toute autre surface dans l'environnement. Une fraction de la lumière quittant toute autre surface peut arriver à l'élément de surface en question et peut être réfléchie à son tour dans l'environnement.

La fraction dépend du facteur de forme entre les surfaces et la réflectivité de l'élément de surface.

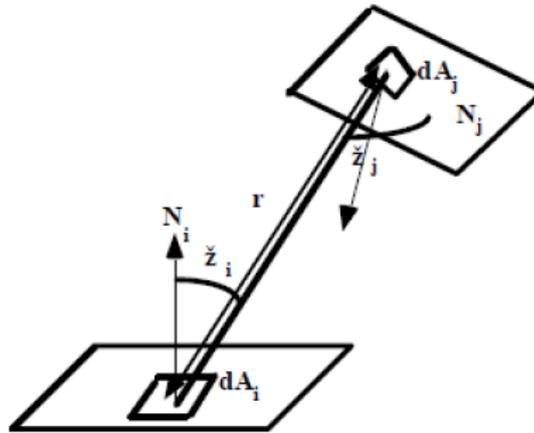


FIGURE 1.10 – Interprétation géométrique.

La radiosit   emise par une facette i B_i est   gale    l'  nergie auto-  mise $B - ei$ plus la somme de toutes les radiosit   re  ues des autres facettes j pond  r  e par un facteur de r  -  mission d  pendant du mat  riau i . L'  nergie re  ue par la facette i de la facette j est   gale au produit de la radiosit   emise par j multipli  e par un facteur de forme F_{ji} , d  pendant de l'orientation relative de i et de j , de leur distance et de la pr  sence d'autres objets entre les deux facettes (ou patches) 1.10. Formellement, la radiosit   est exprim  e comme suit :

$$B_i = B_{ei} + \rho_i \cdot \sum_{j=1}^N B_j$$

B_i :la radiosit   de l'  l  ment de surface i .

B_j :est la radiosit   du j   me   l  ment

F_{ij} : le facteur de forme c'est la fraction de la lumi  re quittant l'  l  ment (ou facette) i et arrivant    l'  l  ment j .

Apr  s la r  solution d'  quation de Kajiya, la m  thode de radiosit   propose de suivre les   tapes suivantes :

On traduit cette algorithme    un sch  ma explicatif .

Algorithm 2 Algorithme global de radiosit 

- Mod liser la sc ne et sa g om trie.
 - Diviser la sc ne en facettes (patches) (voir la figure 1.12).
 - Calculer tous les facteurs de forme.
 - R soudre la matrice de radiosit .
 - Ajouter les conditions de visualisations position et orientation de la cam ra, type projection.
 - Projeter l'image 2D   partir de la sc ne 3D.
-

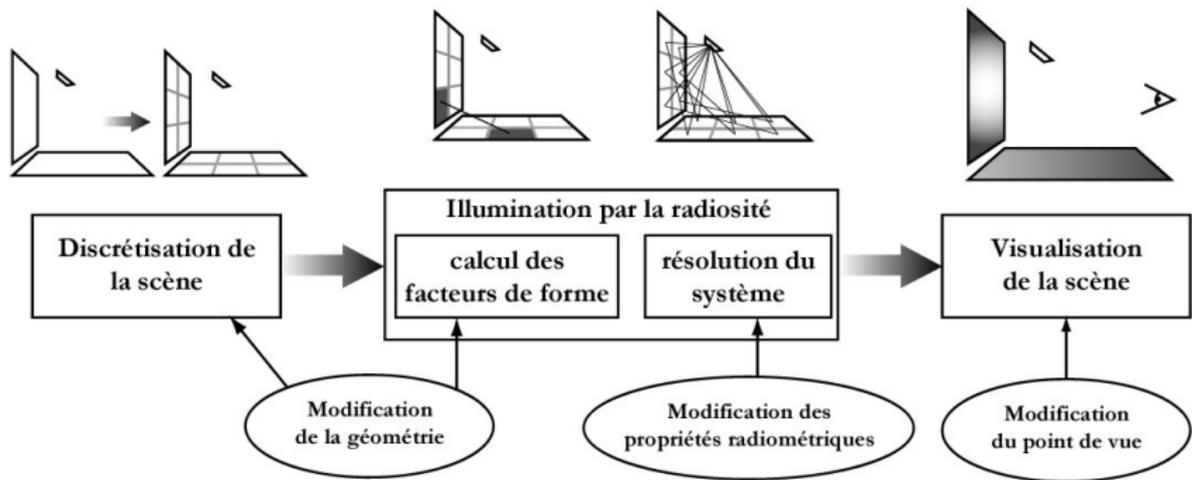


FIGURE 1.11 – Sch ma explicatif de principe g n ral de la radiosit 



FIGURE 1.12 – Exemple de la division d'une sc ne

Résolution informatique du problème de radiosit 

Consid rons que la sc ne 3D est d compos e en plusieurs surfaces rectangulaires (des patches). On peut donc associer   chacun de ces patches une  quation du type de celle d crite pr c demment. En cons quence, le calcul de la radiosit  B_i d'un patch i peut  tre obtenu en r solvant un syst me lin aire suivant de n  quations   n inconnues, qui repr sente un certain " quilibre des transferts  nerg tiques" .

  partir de l' quation de la radiosit  , on peut  crire :

$$\begin{pmatrix} 1 - \rho_1 F_{11} & -\rho_1 F_{1n} & \dots & -\rho_1 F_{1n} \\ -\rho_2 F_{21} & 1 - \rho_2 F_{22} & \dots & -\rho_2 F_{2n} \\ \dots & \dots & \dots & \dots \\ -\rho_n F_{n1} & -\rho_n F_{n2} & \dots & 1 - \rho_n F_{nn} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \dots \\ B_n \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ \dots \\ E_n \end{pmatrix}$$

Pour r soudre le syst me lin aire On utilise des m thodes d'approximations it ratives, comme montr  sur le sch ma ci dessous.

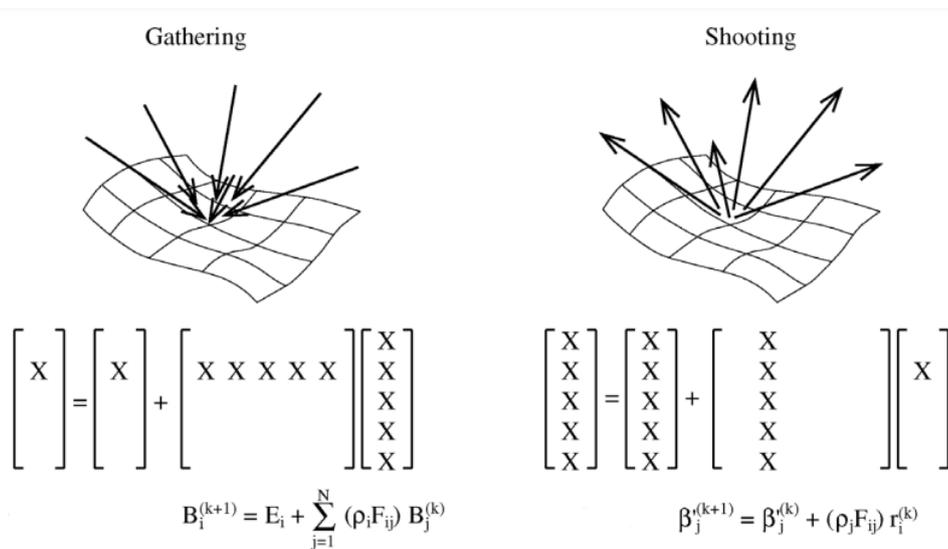


FIGURE 1.13 – Gathering vs. Shooting

A chaque it ration, on s'approche de la solution exacte. On arr te le calcul :

1. Seuil de temps : apr s une certaine p riode de temps
2. Seuil d' nergie : lorsque la majorit  de l' nergie est r partie entre les facettes

Bien entendu, plus la sc ne contient de patches, plus cette r solution est longue et

plus l'espace mémoire occupé est important. Il existe heureusement des méthodes qui permettent de converger vers la solution de ce système sans pour autant le résoudre.

Les calculs de radiosit , contrairement   ceux du lancer de rayons, sont ind pendants de la position de l'observateur (c'est   dire qu'  aucun moment, dans les calculs, la notion d'observateur n'intervient).

1.2.6.4 Path tracing

Le tra age de parcours (path tracing) est une approche graphique qui reproduit la propagation de la lumi re en tirant des rayons   partir d'une cam ra virtuelle et en suivant leur parcours dans l'environnement. Elle prend en compte les interactions entre les mat riaux, l' clairage direct et indirect et l' chantillonnage multiple pour produire des images r alistes avec des effets d'illumination, de r flexion et de diffusion exacts.

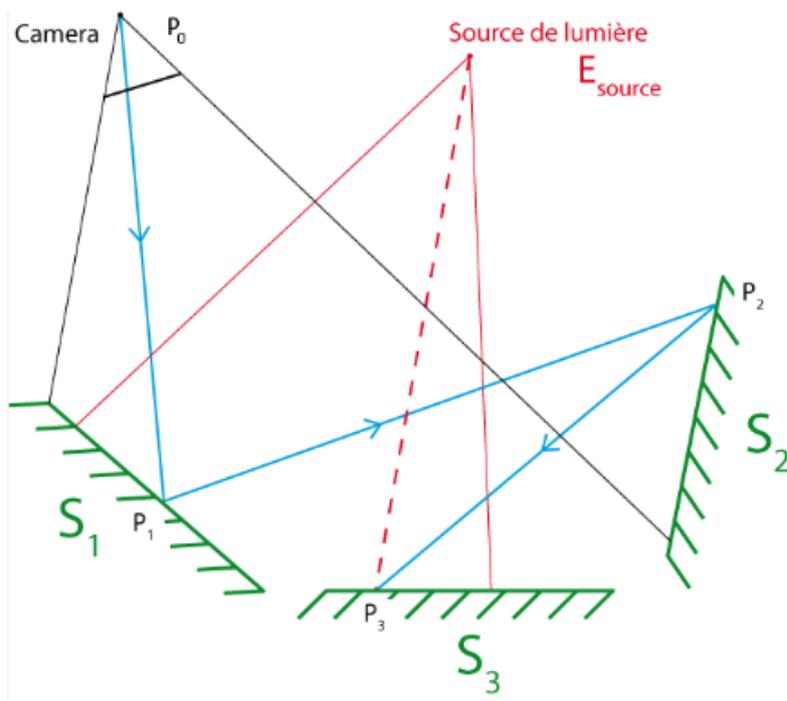


FIGURE 1.14 – Principe de path tracing

1.2.6.5 Monte Carlo

La technique de Monte Carlo est une méthode d'estimation de nombres inconnus ou difficiles à calculer analytiquement. son principe c'est un tirage aléatoire des des échantillons dans la scène (voir la figure 1.15). Des estimations statistiques sont créées à partir de ces échantillons afin de fournir une approximation de la réponse souhaitée. Le nombre d'échantillons utilisés détermine la précision des estimations, et le processus converge vers la valeur correcte lorsque le nombre d'échantillons s'approche de l'infini.

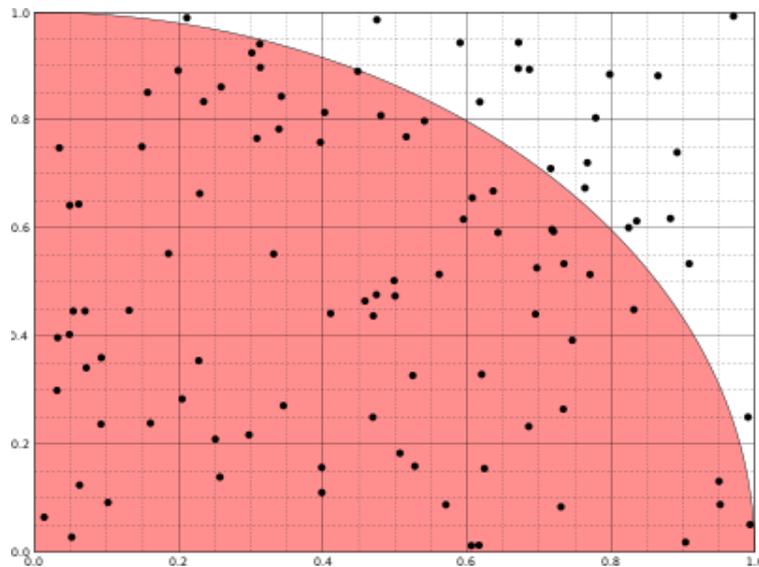


FIGURE 1.15 – Principe de Monte Carlo

1.2.6.6 Photon mapping

Photon mapping (Jensen 1996) est une technique à deux passes qui suit les chemins d'éclairage des sources de lumière à l'œil de l'observateur et vice versa. Cependant, pour des raisons d'efficacité, cette stratégie enregistre et réutilise des paramètres d'éclairage pré-calculés. Les photons sont suivis à partir de sources de lumière dans la scène au premier passage. Ces photons, qui transmettent des informations, sont stockés dans une structure de données connue sous le nom de cartes photoniques. L'information de la carte photonique est utilisée pour produire une image dans le second passage.

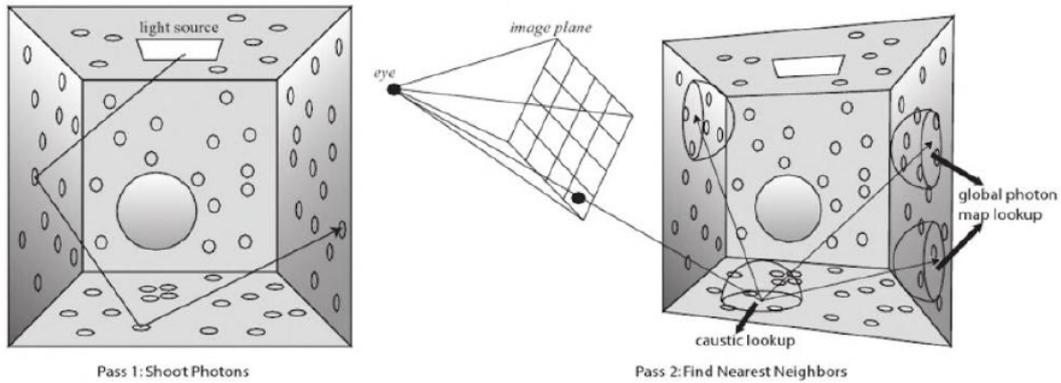


FIGURE 1.16 – Deux passes de l’algorithme de photon mapping : Pass1 : distribution des photons. Pass 2 : rendu par estimation de luminance.

1. Première passe : Distribution des photons

Le traçage de photons est le processus d’émission de photons discrets à partir des sources de lumière à travers la scène. L’objectif principal de cette passe est de remplir les cartes de photons utilisées dans la passe de rendu pour calculer la luminance réfléchie sur les surfaces, et la luminance diffuse dans les milieux participants.

- Émission de photons
- Diffusion de photons
- Stockage de photons

2. Deuxième passe : rendu

Au cours de la passe de rendu, l’algorithme utilise une approximation de l’équation de rendu pour calculer la luminance réfléchie aux endroits de la surface.

1.3 Partie B : Apprentissage profond

1.3.1 Motivations

Aujourd’hui, l’apprentissage profond est devenu l’un des domaines les plus populaires et les plus visibles de l’apprentissage automatique, en raison de son succès dans

une variété d'applications, telles que la vision par ordinateur, le traitement du langage naturel et l'éducation renforcée.

L'apprentissage profond est une branche de l'apprentissage automatique basée sur des réseaux neuronaux artificiels. Il est capable d'apprendre des schémas et des relations complexes au sein des données. Dans l'apprentissage profond, nous n'avons pas besoin de tout programmer explicitement. Il est devenu de plus en plus populaire ces dernières années en raison des progrès en matière de puissance de traitement et de la disponibilité de grands ensembles de données. Parce qu'il est basé sur des réseaux neuronaux artificiels (ANN) également connus sous le nom de réseaux nerveux profonds (DNNs). Ces réseaux neuronaux sont inspirés par la structure et la fonction des neurones biologiques du cerveau humain, et ils sont conçus pour apprendre de grandes quantités de données.

1. L'apprentissage automatique fonctionne uniquement avec des ensembles de données structurées et semi-structurées, tandis que l'apprentissage profond fonctionne à la fois avec des données structurées et non structurées.
2. Les algorithmes d'apprentissage profond peuvent effectuer des opérations complexes efficacement, tandis que les algorithmes d'enseignement automatique ne peuvent pas.
3. algorithmes d'apprentissage automatique utilisent des données d'échantillon étiquetées pour extraire des modèles, tandis que l'apprentissage profond accepte de grands volumes de données en tant qu'entrée et analyse les données d'entrée afin d'extraire des caractéristiques à partir d'un objet.
4. Les performances des algorithmes d'apprentissage automatique diminuent avec l'augmentation du nombre de données, donc pour maintenir les performances du modèle, nous avons besoin d'un apprentis profond.

1.3.2 Apprentissage automatique

Le Apprentissage automatique est une branche de l'intelligence artificielle (IA) qui se concentre sur le développement d'algorithmes et de modèles qui permettent aux ordinateurs d'apprendre et de prendre des prévisions ou des décisions sans programmation

explicite. Il implique l'étude de modèles statistiques et d'algorithmes qui analysent et interprètent automatiquement les schémas dans les données, permettant aux ordinateurs d'améliorer leurs performances sur une tâche spécifique à travers l'expérience ou la formation.

L'objectif principal de l'apprentissage automatique est de créer des algorithmes qui peuvent généraliser leurs connaissances à partir de données pour faire des prédictions précises ou des choix sur de nouvelles données inexplorées.

L'apprentissage automatique comporte généralement deux phases :

la Phase d'apprentissage (l'entraînement) : Dans cette phase, le modèle de machine automatique est entraîné sur un ensemble de données d'apprentissage. Voici les étapes clés de la phase d'apprentissage : dans cette phase L'apprentissage automatique nécessite la préparation et la collecte de données. Cela peut impliquer la collecte de données brutes, le nettoyage en éliminant les valeurs anormales (outils), le traitement des données manquantes, la normalisation ou la modification des caractéristiques pour les rendre appropriés pour l'apprentissage automatique, etc. ensuite , Selon le type de problème et les caractéristiques des données, un algorithme de machine automatique approprié est choisi. Il peut s'agir d'un algorithme de régression linéaire, d'un réseau de neurones..etc

la phase d'évaluation : Cette phase est essentielle pour évaluer la qualité du modèle et déterminer s'il répond aux exigences et aux objectifs fixés, une fois que le modèle a été entraîné ,Une collection distincte de données d'essai est créée qui n'a pas été utilisée pour la formation du modèle ou la validation du modèle. La performance du modèle sur les données fraîches est évaluée de manière objective en utilisant ce ensemble de tests, qui sert également à mesurer la capacité du modèle à généraliser . ensuite Les données de test doivent être pré-traitées en utilisant les mêmes transformations et normalisations que les données de formation, Les prévisions générées à travers la base des données de test sont contrastées avec les valeurs étiquetées des données de test correspondant aux valeurs réelles. Pour comprendre les avantages et les inconvénients du modèle, les résultats de l'évaluation sont examinés. Cela peut impliquer la recherche de groupes ou

de situations dans lesquelles le modèle fonctionne bien ou mal, la détection de biais ou d'incohérences, l'identification d'erreurs typiques, etc...

1.3.3 Les réseaux de neurones artificielles

Les réseaux neuronaux artificiels (RNA) [10], également connus sous le nom de perceptrons multicouches, qui sont des modèles mathématiques qui ont été influencés par la façon dont le cerveau humain fonctionne. Ils sont conçus pour effectuer des tâches d'apprentissage automatique en simulant le comportement des neurones biologiques.

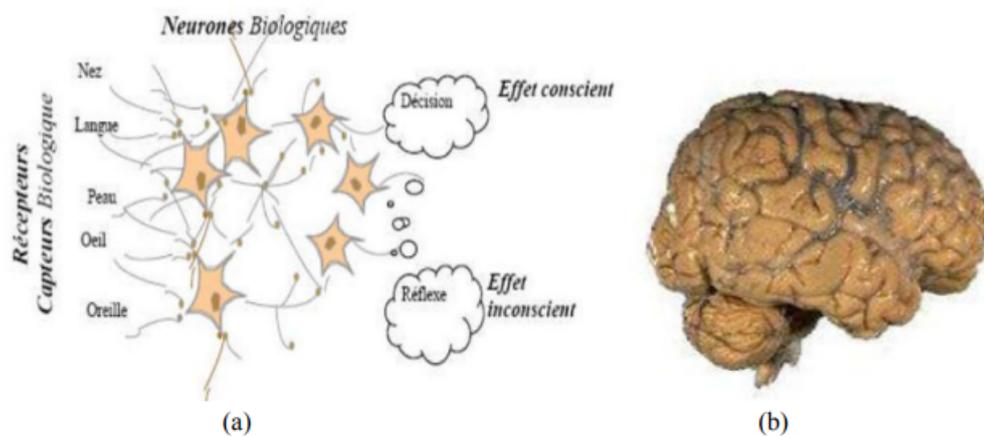


FIGURE 1.17 – (a) : Structure d'un réseau de neurone biologique ,(b) le cerveau humain[9]

L'origine des réseaux de neurones vient de l'essai de modélisation mathématique du cerveau humain les premiers travaux datent de 1943 et sont l'œuvre de W.M. Culloch et W. Pitts. Ils supposent que l'impulsion nerveuse est le résultat d'un calcul simple effectué par chaque neurone et que la pensée née grâce à l'effet collectif d'un réseau de neurone inter connecté .

Un réseau de neurones est un assemblage de constituants élémentaires inter connectés (appelés «neurones» en hommage à leur modèle biologique), qui réalisent chacun un traitement simple mais dont l'ensemble en interaction fait émerger des propriétés globales complexes. Chaque neurone fonctionne indépendamment des autres de telle sorte que l'ensemble forme un système massivement parallèle. L'information est stockée de manière distribuée dans le réseau sous forme de coefficients synaptiques ou de fonctions

d'activation, il n'y a donc pas de zone de mémoire et de zone de calcul, l'une et l'autre sont intimement liés. Un réseau de neurone ne se programme pas, il est entraîné grâce à un mécanisme d'apprentissage. Les tâches particulièrement adaptées au traitement par réseau de neurones sont : l'association, la classification, la discrimination, la prévision ou l'estimation, et la commande de processus complexes.

Les réseaux de neurones artificiels consistent en des modèles plus ou moins inspirés du fonctionnement cérébral de l'être humain en se basant principalement sur le concept de neurone.

1.3.4 Types d'apprentissage

Les algorithmes d'apprentissage peuvent se catégoriser selon le mode d'apprentissage qu'ils emploient. Il existe :

1.3.4.1 Apprentissage supervisé

Un modèle est formé en utilisant la technique d'apprentissage automatique connue sous le nom de supervisé automatique sur une collection de données étiquetées, ou des données pour lesquelles les bonnes réponses sont déjà connues. L'objectif de l'apprentissage supervisé est d'enseigner au modèle comment prédire les étiquettes ou les valeurs cibles en fonction des caractéristiques ou des variables fournies comme input.

Avec l'apprentissage supervisé, la machine peut apprendre à faire une certaine tâche en étudiant des exemples de cette tâche. Par exemple, elle peut apprendre à reconnaître une photo de chien après qu'on lui ait montré des millions de photos de chiens. Ou bien, elle peut apprendre à traduire le français en chinois après avoir vu des millions d'exemples de traduction français-chinois.

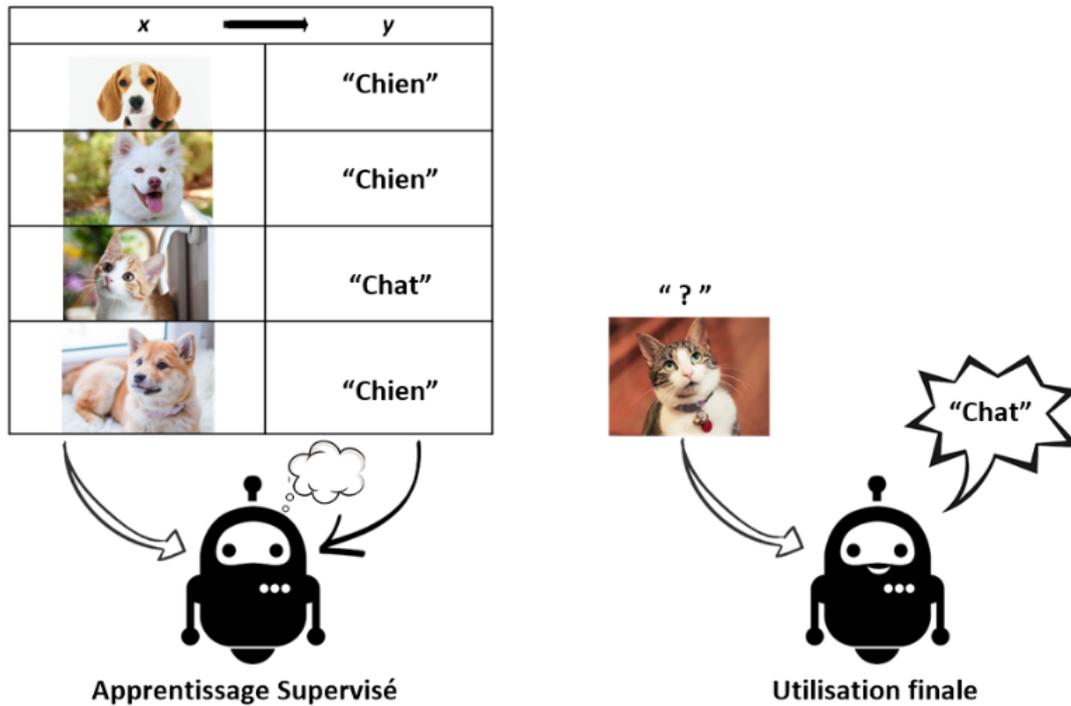


FIGURE 1.18 – Exemple d'apprentissage supervisé

L'apprentissage supervisé fonctionne en 4 étapes :

1. Importer un Dataset (x, y) qui contient nos exemples
2. Développer un Modèle aux paramètres aléatoires
3. Développer une Fonction Coût qui mesure les erreurs entre le modèle et le Dataset.
4. Développer un Algorithme d'apprentissage pour trouver les paramètres du modèle qui minimisent la Fonction Coût

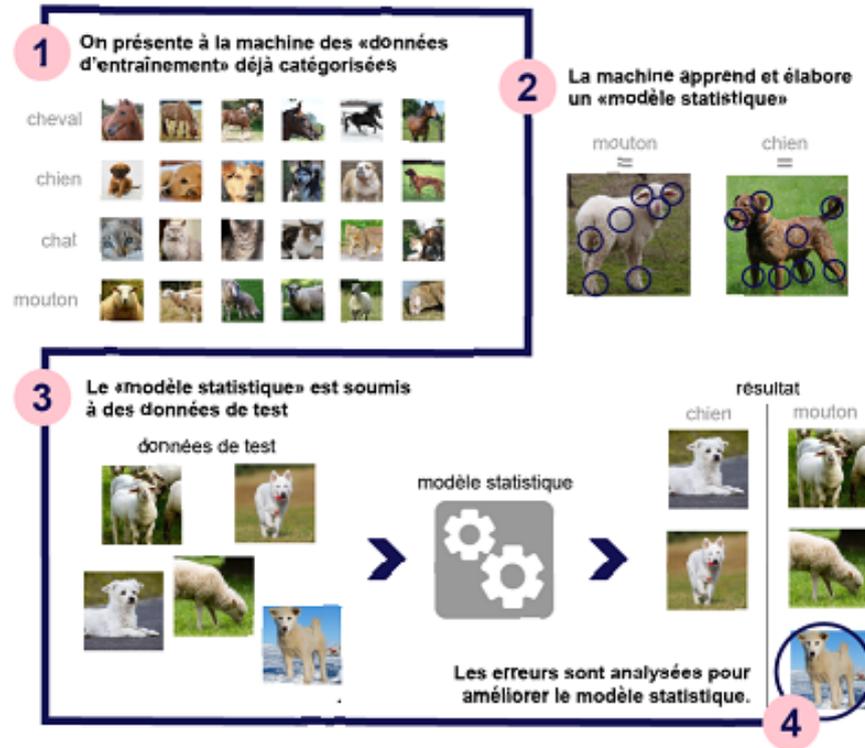


FIGURE 1.19 – Schéma explicatif des étapes de l'apprentissage supervisé

1. le dataset : Apprendre à partir d'exemples

Dans la cas de 1.18, cette apprentissage est dit supervisé car nous supervisons la reconnaissance des données en entrée (e.g les photos) en explicitant leur catégorie . L'algorithme a besoin de nous. On dit que les données sont labellisées.

2. Le modèle : Développer un modèle à partir du Dataset

En Machine automatique, on développe un modèle à partir de ce Dataset. A la différence du modèle illustré plus haut, un modèle de Machine automatique ne repose pas sur une démonstrations mathématique ou une équation physique. A la place, il est construit à partir de données, comme un modèle statistique.

Si par exemple le Dataset donne le nuage de point suivant, alors la machine devra trouver le modèle qui rentre le mieux dans ce nuage de point.

Dans les faits, c'est à nous de choisir le type de modèle (c'est-à-dire la fonction mathématique) et c'est à la machine de trouver les coefficients(a,b,c) de cette fonction qui donnent les meilleurs résultats. Par convention on appelle ces coefficients les paramètres du modèle.

Avec le Supervisée automatique on peut développer des modèles pour résoudre 2 types de problèmes :

- Les problèmes de Régression
- Les problèmes de Classification

3. La Fonction coût : mesure de la performance

Autre chose à noter qu'un modèle nous retourne des erreurs par rapport à notre Dataset. On appelle Fonction Coût l'ensemble de ces erreurs.

Avoir un bon modèle, c'est avoir un modèle qui nous donne de petites erreurs, donc une petite Fonction Coût.

4. L'Algorithme d'apprentissage : Apprendre, c'est minimiser la Fonction Coût.

Ainsi l'objectif central en Apprentissage supervisé, c'est de trouver les paramètres du modèle qui minimisent la Fonction Coût. Pour cela, on utilise un algorithme d'apprentissage comme ;l'algorithme de Gradient Descent.

1.3.4.2 Apprentissage non supervisé

Un modèle est formé sur une collection de données non étiquetées, c'est-à-dire des données pour lesquelles les réponses ou les étiquettes ne sont pas connues à l'avance, dans un processus connu sous le nom d'apprentissage non supervisé. Contrairement à l'apprentissage supervisé, le modèle doit indépendamment révéler toutes les structures ou modèles présents dans les données.

Dans l'apprentissage non supervisé, les données sont généralement regroupées en fonction de similarités ou de caractéristiques communes. Parmi Les objectifs principaux de l'apprentissage non supervisé est le regroupement (clustering) il s'agit de regrouper des données similaires. Sans aucune connaissance préalable de ces groupements, le regroupement vous permet de trouver des groupes ou des classes inhérents dans les données.

1.3.4.3 Apprentissage par renforcement

L'apprentissage du renforcement est le processus de choix du bon cours d'action pour maximiser la récompense dans une certaine circonstance. Il est utilisé par une variété de logiciels et de robots pour déterminer le cours optimal d'action à poursuivre dans une circonstance donnée. Contrairement à l'apprentissage supervisé, où les données de formation comprennent la clé de réponse, ce qui permet au modèle d'être entraîné avec la réponse correcte, le renforcement apprenant s'appuie sur l'agent renforçant pour déterminer le meilleur cours d'action pour compléter la tâche à portée de main. Sans un ensemble de données de formation, il va inévitablement acquérir des connaissances à partir de l'expérience.

L'apprentissage du renforcement est utile pour aider nos robots à apprendre à remplir leurs devoirs sans avoir à coder leurs procédures pour de nombreuses applications robotiques (think manufacturing lines). Cela peut être une solution plus abordable, plus sûre et encore moins susceptible d'échouer. Afin d'économiser de l'argent, nous pouvons également fournir des incitations pour que nos appareils consomment moins d'énergie. Plus que cela, nous pouvons commencer l'ensemble du processus dans une simulation pour éviter de gaspiller de l'argent au cas où nous endommagerions notre machine.

1.4 Apprentissage profond

La méthode de deep learning [12], un domaine récent de la recherche sur l'apprentissage automatique, permet à l'ordinateur de apprendre à exécuter des fonctions qui viennent naturellement au cerveau, telles que l'identification d'image. Les techniques d'apprentissage profond (DL) ont eu une influence significative sur les applications pour la vision informatique et l'analyse de l'image, y compris la classification des images, la segmentation, l'achèvement des images et d'autres. Les réseaux neuronaux artificiels, une sorte d'apprentissage automatique qui tire son inspiration du fonctionnement du cerveau humain, sont l'objet d'une formation profonde.

1.4.1 Réseaux Feedforward

La capacité des réseaux neuronaux artificiels, également connus sous le nom d'ANN, à apprendre des cartes complexes à partir d'exemples, ou ce qui est connu sous le titre d'approximation universelle, a suscité beaucoup d'attention ces dernières années. Ils sont connus sous le nom de neurones puisque la neurosciences sert de source d'inspiration. Bien que les données d'entraînement ne spécifient pas ce que chaque couche individuelle devrait faire.

l'algorithme d'apprentissage doit déterminer comment utiliser ces couches pour obtenir la sortie prévue. Bien que les processus biologiques que les scientifiques ont pu étudier dans le cerveau dans les années 1950 ont servi d'inspiration pour les neurones artificiels, les neurones artificiel et biologique diffèrent de plus en plus que simplement les matériaux de leurs conteneurs. Donc, nous devons discuter du bloc de construction fondamental d'un réseau neuronal.

1.4.1.1 Les réseaux biologiques

La composante fondamentale du cerveau et du système nerveux central est le neurone. Des cellules spécialisées appelées neurones transmettent des informations électriques et chimiques. Les seules cellules autres que les neurones qui composent le cerveau sont les cellules gliales, qui donnent aux neurones un soutien et une forme. Les 86 milliards ou plus de neurones du système nerveux collaborent pour communiquer avec le reste du corps. Ils sont responsables de tout, y compris la douleur et la faim, ainsi que de la conscience et de la cognition.

1. dendrite :sont des structures qui ressemblent à des branches qui s'éloignent du corps cellulaire. Ils sont responsables de recevoir des informations des neurones voisins et de leur permettre d'atteindre le corps cellulaire.
2. le corps cellulaire : qui contient un noyau, lisse et rugueuse endoplasmique le réticul, l'appareil de Golgi, les mitochondries et d'autres composants cellulaires.
3. axon :À son niveau le plus basique, un axon est une structure semblable à un tube qui transporte un impulsion électrique du corps cellulaire (ou des dendrites

d'une autre cellule) aux terminaux axoniques à l'autre bout du neurone, qui peut ensuite transmettre l'impulsion à un autre neurone.

4. le terminal axonique :Le lien chimique entre les terminaux axoniques d'un neurone et les dendrites du neurone suivant est connu sous le nom de synapse. Au lieu d'une structure, c'est un espace où des interactions chimiques complexes peuvent avoir lieu.

1.4.1.2 Les couches

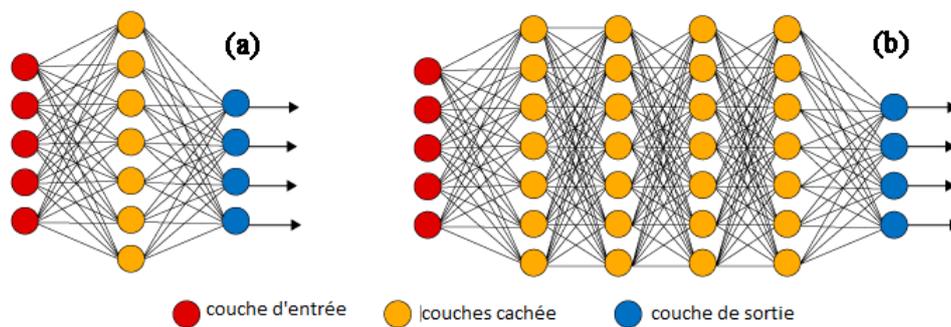


FIGURE 1.20 – (a) Architecture de réseau neuronal simple , (b) Architecture simple du réseau neuronal profond (DNN)[6]

La couche d'entrée : c'est La première couche d'un réseau neural . Il est composé de neurones d'entrée et envoie des données brutes aux couches cachées pour un traitement ultérieur.

La couche cachée : entre les couches d'entrée et de sortie, une couche ou une collection de couches. Beaucoup de ces calculs sont faits dans la (les) couche cachée de l'apprentissage profond, qui est un problème d'optimisation qui cherche la meilleure solution à un problème extrêmement compliqué. Le nombre de couches cachées à utiliser dépend de la complexité des données ; pour les données moins complexes, il est conseillé d'utiliser moins de couche cachées ; pour des données plus complexes des couches plus cachées devraient être utilisées. L'utilisation d'un design de base dans un problème complexe entraîne une sous-configuration, tandis que le fait de le faire dans une situation simple pourrait conduire à un surconfigurateur. Les couches cachées du réseau sont généralement évaluées par vecteur. La largeur du modèle dépend de la dimensionnalité de ces couches cachées.

La couche de sortie : Dans un réseau neuronal, la couche de sortie est la dernière couche qui génère les sorties du programme. Pour les tâches de classification, la taille de la couche de sortie est égale au nombre de classes.

1.4.2 Architectures de réseaux neuronaux

Il existe de nombreuses architectures de réseaux neuronaux différents, telle que les Réseaux neuronaux convolutifs(CNN).

1.4.2.1 Réseaux neuronaux convolutifs(CNN)

Les réseaux neuronaux convolutifs (CNN) sont une famille de réseaux neuro-artificiels qui sont le plus souvent utilisés dans l'apprentissage profond pour interpréter les données visuelles. Dans au moins une de ses couches, les CNN remplacent le processus mathématique de convolution pour la multiplication de matrice générique. Ils sont employés dans le traitement de l'image et la reconnaissance et ont été créés particulièrement pour traiter les données de pixels .

Les réseaux neuronaux convolutifs (CNN) utilisent la structure hiérarchique des données. Afin d'extraire des informations utiles des données. Les CNN divisent une image ou une entrée en caractéristiques plus petites et plus simples, qui sont représentées par des filtres, plutôt que d'essayer de traiter l'image complète ou l'entrée tout à la fois. Pour extraire les données pertinentes, ces filtres sont utilisés sur différents emplacements d'entrée. Ces éléments sont fusionnés et construits en modèles plus complexes au fur et à mesure que le réseau avance à travers les couches, ce qui lui permet d'acquérir des représentations de plus en plus abstraites de l'entrée. Tout en réduisant la possibilité de surinstallation, cette technique hiérarchique permet aux CNN d'apprendre efficacement des schémas compliqués dans les données. Les CNN sont donc à l'extrémité inférieure du spectre de connectivité et de complexité.

Une couche d'entrée, des couches cachées et une couche de sortie constituent un réseau neuronal convolutif^{1.20}. Les couches cachées d'un réseau neuronal convolutif contiennent une ou plusieurs couches convolutives. Ceci contient généralement une

couche qui fait un produit point de la matrice d'entrée de la couche avec le noyau de convolution. Le mécanisme d'activation de ce produit, qui est souvent le produit interne de Frobenius, est souvent ReLU. La procédure de convolution développe une carte des caractéristiques alors que le noyau de convolutions se déplace au-dessus de la matrice d'entrée pour la couche, ajoutant à l'entrée du niveau suivant. Ensuite, il y a des couches supplémentaires telles que la normalisation, le regroupement et les couches entièrement liées.

Il existe 4 couches primaires ou étapes dans la conception d'une CNN :

1. Couche d'entrée : est l'endroit où les représentations tensorielles multidimensionnelles des données d'entrée, souvent des images, sont transmises. Vous pouvez penser à chaque pixel de l'image comme une valeur dans la matrice d'entrée.

Couches convolutionnaires : Les composants de base d'une CNN sont des couches de convolution. Chaque couche de convolution est composée d'un certain nombre de filtres (ou de noyaux) qui convolutionnent les données entrantes (les pixels). Ces filtres recherchent certains modèles visuels tels que des bords, des textures ou des formes 1.21. pour chaque filtre sera produit Une carte d'activation, également appelée carte des caractéristiques c'est la multiplication des pixels d'entrée et un filtre 1.22 .

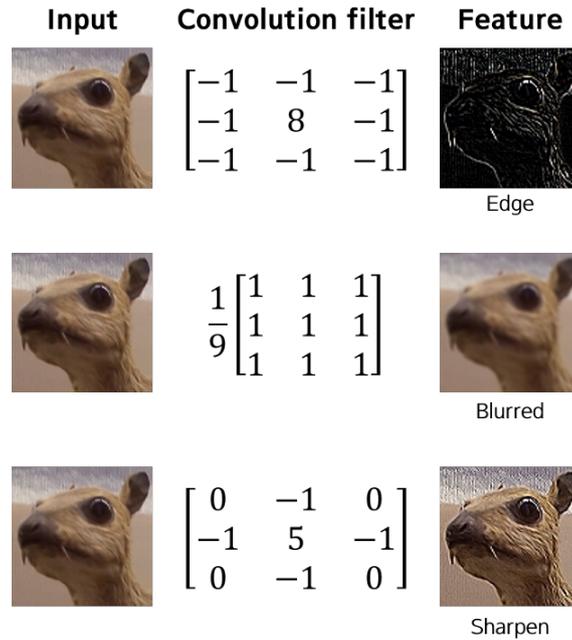


FIGURE 1.21 – Exemple des filtres[22]

2. Couche de Couverture (Subsampling or pooling layer) :

Les dimensions spatiales des cartes des caractéristiques (largeur et hauteur) sont réduites, mais les données les plus cruciales sont préservées par des couches de regroupement. Les deux types de regroupement les plus couramment utilisés sont le maximum et le moyen. Alors que le regroupement moyen utilise la valeur moyenne de chaque groupe local de neurones dans la carte des caractéristiques, le regroupement maximal emploie la valeur maximale de chacun comme dans la figure ci-dessous.

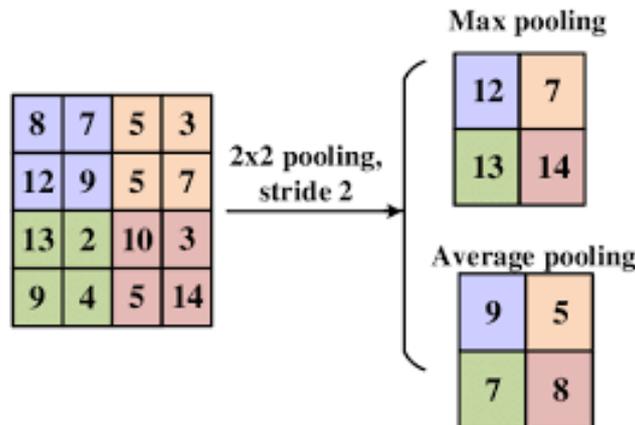


FIGURE 1.22 – Exemple des opérations de couche de regroupement[27]

3. Couche d'aplatissement (flatten layer) :Après la génération des cartes des caractéristiques regroupées (pooled feature map) à travers la couches de Couverture (Subsampling or pooling layer)on les aplater en une vecteur comme dans l'image ci-dessous.

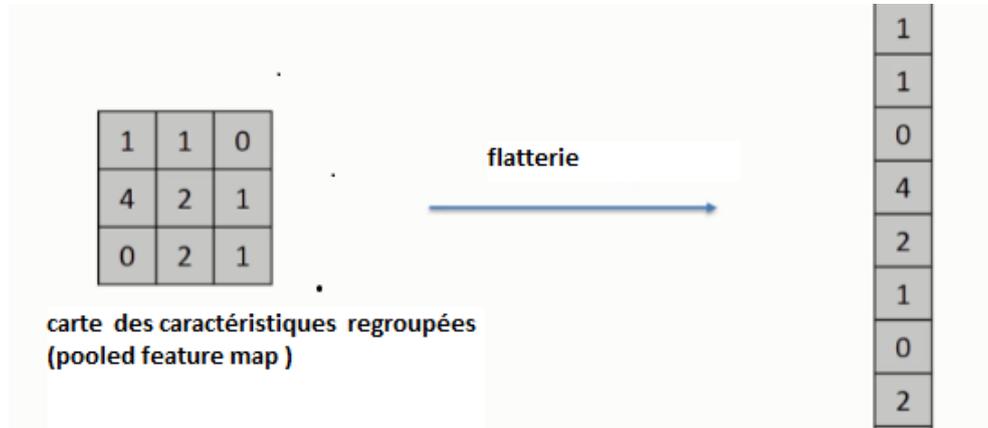


FIGURE 1.23 – flatterie

La raison pour laquelle nous faisons cela est que nous allons devoir insérer ces données dans un réseau neuronal artificiel plus tard (Fully connected layers).

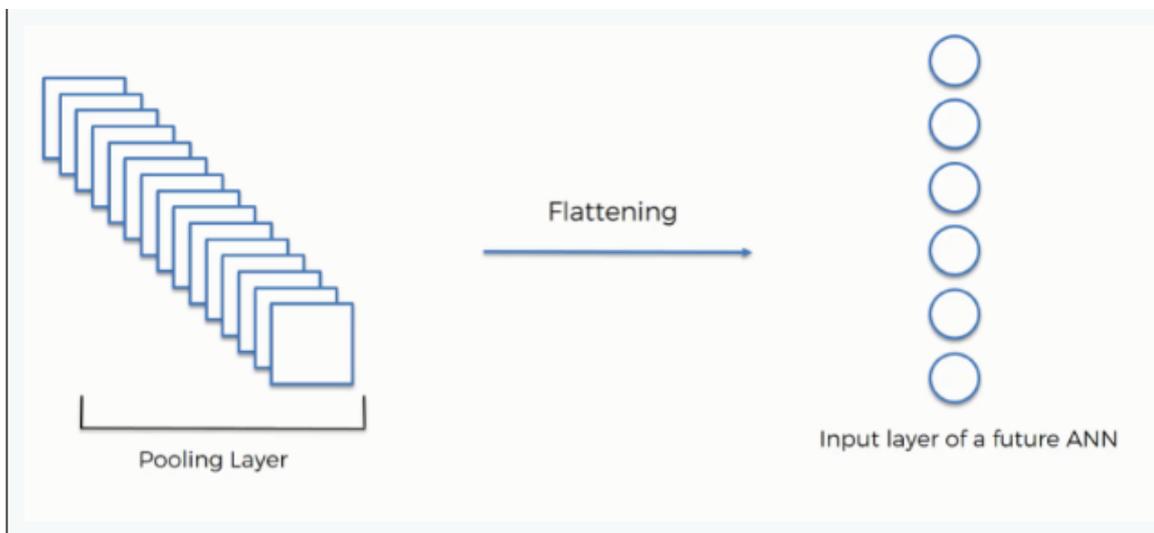


FIGURE 1.24 – Couches complètement connectées

4. Des lignes de couches complètement connectées(Fully connected layers)dans cette couche Chaque neurone de la couche précédente est liée aux neurones de la prochaine couche via cette couche (Perceptron multicouches). l'entrée du ANN c'est la sortie de précédente couche qui est un vecteur de caractéristique 1.23 pour

classifier les images.

1.5 Conclusion

Nous avons mis en évidence, dans ce chapitre, les notions fondamentales à la compréhension de la complexité de l'illumination, qui constituent le premier début à la génération des images photo-réalistes. Pour ce faire, nous avons étudié les deux grandes classes de l'illumination des scènes 3D à savoir, l'illumination locale et l'illumination globale. L'objectif était de montrer la complexité des calculs physiques qui génèrent un coût de calcul énorme, lorsqu'on désire obtenir un rendu précis et efficace.

Nous avons ensuite exposé les notions de l'apprentissage et focaliser surtout sur l'apprentissage profond et les CNN car nous avons vu que la majorité des techniques de rendu et des études de recherches récentes se dirigeaient vers le rendu neuronal, que nous allons expliquer dans le deuxième chapitre de ce mémoire. newpage

Chapitre 2

Rééclairage basé sur l'apprentissage profond à partir d'une seule photo de portrait

2.1 Introduction et motivations

Après avoir présenté dans le chapitre précédent, les deux domaines sur lesquels se basent notre problématique, nous allons dans ce chapitre justifier les choix qui nous ont poussés à introduire l'apprentissage profond dans le domaine de l'illumination des scènes virtuelles.

Ainsi nous allons nous intéresser à une branche de la vision informatique et du traitement de l'image appelée « Deep Single-Image Portrait Relighting » [28], celle-ci tente de créer de nouveaux portraits d'une personne dans divers scénarios d'éclairage à partir d'un seul portrait et d'un éclairage cible. L'importance de ce domaine découle du fait que la photographie traditionnelle implique de modifier les configurations d'éclairage pour capturer de nouvelles photos, ce qui peut être coûteux, chronophage et pas toujours pratique.

Les réseaux de neurones convolutifs (CNN) et les réseaux antagonistes génératifs (GAN), deux approches d'apprentissage en profondeur, ont joué un rôle crucial dans le

développement d'algorithmes de rééclairage de portrait à partir d'une seule image en profondeur. Ces méthodes ont été utilisées pour créer de nouvelles images à la fois réalistes et cohérentes avec l'image originale, ainsi que pour apprendre des correspondances complexes entre des photographies prises dans différentes conditions d'éclairage.

2.2 Le rendu neuronal : Présentation et définitions

Le rendu neuronal [23] est un concept classique issu du développement de l'infographie. Un pipeline de rendu neuronal apprend à représenter des scènes à partir d'images du monde réel. Son entrée est soit un ensemble non ordonné d'images représentant différentes vues de la même scène, soit un ensemble structuré d'images ou de vidéos fournissant des angles spécifiques et connus de la scène.

Le rendu neuronal simule la physique d'une caméra capturant la scène. Une caractéristique importante du rendu neuronal 3D est de séparer la représentation de la scène en 3D du processus de capture de la caméra. Il est ainsi plus facile de reconstruire de nouvelles images à partir de la scène avec un niveau élevé de cohérence 3D.

Le rendu neuronal tridimensionnel implique généralement un modèle de génération d'images tel que la projection de points, la rasterisation ou l'intégration de volumes pour découpler les processus physiques tels que la projection des représentations 3D d'une scène. Le modèle fait appel à la physique, en tirant parti de conditions telles que l'éclairage et la relation de la scène avec la caméra. Une équation de rendu formule le transport de la lumière.

L'infographie utilise souvent différentes approximations de l'équation de rendu en fonction du type de représentation de la scène (par exemple, traçage de chemin, tramage ou consolidation de volume). Le système de rendu neuronal 3D exploite la méthode de rendu préférée. Il est important de faire la distinction entre la méthode de rendu et la représentation de la scène pour permettre un entraînement basé sur des images réelles.

Dans cette étude, nous exploitons cette caractéristique en entraînant un réseau générateur pour reproduire ces correspondances tout en imposant une perte contradictoire basée uniquement sur des images réelles. Pour ce faire, nous commençons par créer un

vaste ensemble de données d'entraînement à l'aide d'une technique de ré éclairage basée sur la physique. La perte contradictoire garantit que les apprentissages se basent sur une variété d'images authentiques et sont exempts d'erreurs de l'approche de ré éclairage basée sur la physique, tandis que la perte de reconstruction supervisée permet au réseau d'apprendre à réaliser le ré éclairage.

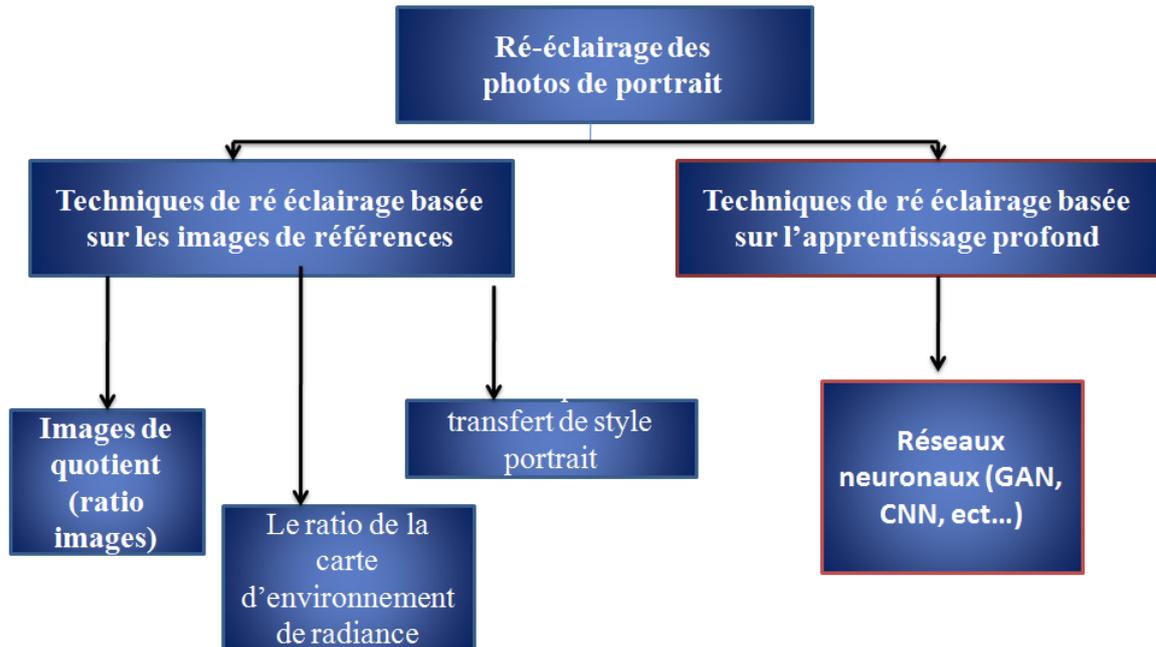
2.3 Comparaison entre la technique qui utilise le réseau Hourglass et les autres techniques proposées pour le rééclairage d'images de portrait

Approche	Principe	Avantages	Inconvénients
Li et al	Fournir une solution de forme fermée pour générer des images stylisées photoréalistes	<ul style="list-style-type: none"> - Fournit une méthode rapide et efficace pour générer des images stylisées photoréalistes. - Maintient un contrôle fin sur les paramètres de stylisation - Conserve une qualité de perception et un réalisme élevés dans les images stylisées. 	<ul style="list-style-type: none"> - Limitation de la gamme de styles et d'effets artistiques disponibles - Perte de réalisme dans les portraits fortement stylisés - Sensibilité à la qualité de l'image d'entrée et aux résolutions
Shih et al	Transférer le style artistique d'une image de référence sur un portrait de tête cible.	<ul style="list-style-type: none"> - Permet la transformation artistique et la personnalisation des portraits. - Préservation des traits essentiels du visage et de l'identité. - Améliore l'attrait visuel et la profondeur des portraits. 	<ul style="list-style-type: none"> - L'inadéquation entre le style et le contenu peut conduire à des résultats visuellement incohérents. - Perte de réalisme dans les portraits fortement stylisés - Manque de contrôle précis sur le processus de stylisation
Shu et al	Transfert des conditions d'éclairage d'une image de référence sur un portrait cible	<ul style="list-style-type: none"> - Transfert réaliste des conditions d'éclairage - Préserve les traits du visage et l'identité - Permet un contrôle précis du processus de transfert de l'éclairage - Flexibilité dans la sélection des images de référence pour les effets d'éclairage souhaités 	<ul style="list-style-type: none"> - Nécessite une estimation fiable et précise des informations relatives à la géométrie et à la réflectance. - Peut être sensible aux variations des expressions faciales ou des poses. - Les performances peuvent se dégrader en cas de conditions d'éclairage complexes ou de portraits difficiles. - Processus à forte intensité de calcul
sfsNet	Apprendre la forme, la réflectance et l'éclairement des visages en conditions réelles	<ul style="list-style-type: none"> - Estimation précise de la forme, de la réflectance et de l'éclairement - Fonctionne bien avec les visages dans différentes conditions d'éclairage et d'environnement - Peut gérer les poses non frontales, les occlusions et les variations de maquillage. - Offre une approche basée sur les données pour le ré-éclairage des visages. 	<ul style="list-style-type: none"> - Peut nécessiter une supervision supplémentaire ou des images de référence pour guider le processus de ré-éclairage. - Les performances peuvent être affectées par des variations extrêmes des conditions d'éclairage ou des ombres importantes. - La complexité de calcul peut être élevée, en particulier avec des ensembles de données à grande échelle. - Limité dans la capture des détails fins et des textures des visages rallumés
La méthode proposée	Améliorer et modifier les conditions d'éclairage d'une seule image portrait	<ul style="list-style-type: none"> - Permet un éclairage réaliste et personnalisable des images de portrait - Permet d'ajuster avec souplesse les conditions d'éclairage et les effets artistiques - Préserve les traits du visage et l'identité - Peut gérer les poses non frontales, les occlusions et les variations de maquillage. 	<ul style="list-style-type: none"> - Sensible à la qualité et à la résolution de l'image d'entrée - La complexité informatique peut être élevée, en particulier pour les images à grande échelle ou à haute résolution. - Dépend de la disponibilité de diverses données d'entraînement et de modèles d'apprentissage profond robustes

FIGURE 2.1 – Tableau comparatif des techniques proposées

2.4 Classification des techniques de ré-éclairage des photos de portrait

Les techniques de ré-éclairage des photos de portrait peuvent être classées en plusieurs catégories en fonction de la méthode utilisée pour modifier l'éclairage de la photo portrait.



2.4.1 Techniques de ré-éclairage basées sur les images de références (techniques d'optimisation)

2.4.1.1 Images de quotient (ratio images) :

Cette technique consiste à calculer l'image quotient [20] en soustrayant pixel par pixel les valeurs des images d'origine et d'une image de référence. L'image quotient représente les différences d'éclairage entre les deux images. On utilise l'image quotient pour ré-éclairer l'image d'origine en ajoutant ou en supprimant les variations d'éclairage capturées par rapport à l'image de référence. Cela permet de modifier l'apparence de l'image d'origine en simulant un nouvel éclairage, tout en préservant les détails et les caractéristiques de l'image d'origine.

Le processus de création d'une image quotient implique une comparaison pixel par pixel entre les valeurs de luminosité des pixels de l'image d'origine et de l'image de référence correspondante. La différence d'éclairage entre ces deux images est calculée en soustrayant les valeurs de luminosité des pixels correspondants.

2.4.1.2 Le ratio de la carte d'environnement de radiance :

Une autre technique pour le ré éclairage des images portrait basée sur Le ratio de la carte d'environnement de radiance s'effectue généralement selon les étapes suivantes :

1. **Acquisition de la carte d'environnement de radiance :** La carte d'environnement de radiance est une représentation des informations d'éclairage de l'environnement capturées à partir d'un panorama sphérique ou hémisphérique. Elle contient les valeurs de radiance (intensité et couleur) pour chaque direction de l'environnement
2. **Calcul du rapport :** Le rapport de la carte d'environnement de radiance est calculé en prenant le rapport entre les valeurs de radiance de la carte d'environnement cible et de la carte d'environnement de référence. Ce rapport représente les différences relatives d'éclairage entre les deux environnements.
3. **Appliquer le rapport à l'image :** Le rapport calculé est ensuite appliqué à l'image originale pour modifier son éclairage. Pour ce faire, on multiplie les valeurs des pixels de l'image originale par les valeurs correspondantes des pixels de la carte de rapport. Le résultat est une nouvelle image qui présente les caractéristiques d'éclairage de l'environnement cible.

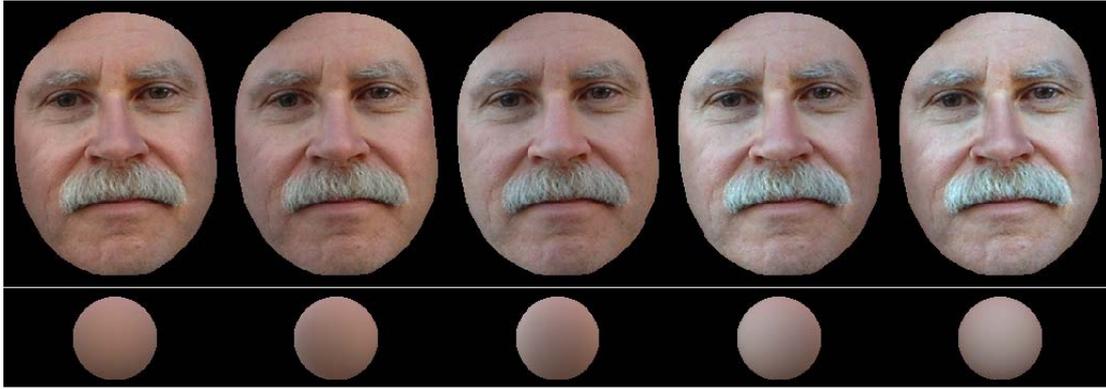


FIGURE 2.2 – L'image du milieu est l'entrée. La séquence montre les résultats synthétisés de la rotation 180° de l'environnement d'éclairage. [25]

Cette méthode présente l'avantage de ne pas nécessiter la séparation de l'éclairage de la réflectance, et elle est simple à mettre en œuvre et fonctionne à une vitesse interactive.

2.4.1.3 Technique de transfert de réflectance

Pieter Peers et al [18] proposent une technique révolutionnaire de relâchement des performances du visage pour les artistes humains en post-production. Cette méthode repose uniquement sur un ensemble de données d'expressions faciales dépendantes de la vue avec des expressions neutres acquises pour une personne statique à l'aide d'un dispositif Light Stage. Cependant, pendant la performance réelle, un acteur potentiellement différent est pris sous l'éclairage connu, mais statique. Le champ de réflexion de l'acteur de base de données est transféré sur la performance dynamique pendant la post-production, ce qui permet de relâcher l'ensemble de la séquence sur la base d'images. Parce qu'il ne nécessite pas d'équipement supplémentaire pendant le tournage primaire, cette solution rend le relâchement post-production plus pratique et peut simplement être inclus dans un flux de travail de production standard. De plus, le système est adapté à l'édition d'éclairage post-production en temps réel.



FIGURE 2.3 – Exemple de Technique de transfert de réflectance [18]

dans cette figure De gauche à droite : un sujet de référence (source) rallumé avec l'éclairage souhaité. Une image de l'acteur de performance (cible) capturée sous un éclairage uniforme. L'image quotient de l'acteur de référence s'est rallumée à l'aide de l'éclairage cible et de l'éclairage de performance uniforme (déformé dans la pose de l'acteur de performance). L'acteur de performance avec l'ombrage souhaité transféré de l'acteur de performance rallumé à l'aide de l'image quotient.

2.4.2 Techniques de ré éclairage basées sur l'apprentissage profond

Les techniques basées sur l'apprentissage profond pour le ré-éclairage des photos de portrait exploitent les capacités des réseaux neuronaux, tels que les réseaux génératifs adverses (GAN) et les réseaux de neurones à convolution (CNN), pour apprendre et ajuster l'éclairage des images.

2.4.2.1 Réseaux neuronaux (GAN, CNN, etc.)

Les réseaux neuronaux, notamment les GAN et les CNN, sont utilisés pour apprendre les relations complexes entre les images d'origine et leurs éclairages correspondants. Les GAN sont particulièrement utiles pour générer de nouvelles images d'apparence réaliste en apprenant à partir d'un ensemble de données d'entraînement.

GAN : Les réseaux génératifs adverses sont utilisés pour générer de nouvelles images ré-éclairées en apprenant à partir d'un ensemble d'images d'origine et de leurs correspondances avec différents éclairages. Le générateur du GAN apprend à transformer l'image

d'origine pour qu'elle corresponde à l'éclairage souhaité, tandis que le discriminateur évalue la qualité des images générées.

CNN : Les réseaux de neurones à convolution sont utilisés pour extraire des caractéristiques spécifiques de l'image et de l'éclairage, ce qui permet d'apprendre des modèles d'ajustement de l'éclairage. Les CNN peuvent être utilisés dans des architectures plus complexes pour apprendre à ré-éclairer les images de portrait de manière réaliste.

2.4.2.2 Apprentissage de l'éclairage à partir de données d'entraînement

Cette approche implique d'utiliser un ensemble de données d'entraînement contenant des paires d'images de portrait et de leurs éclairages correspondants. Les algorithmes d'apprentissage automatique, tels que les réseaux neuronaux, sont utilisés pour apprendre à partir de ces données et générer des modèles d'ajustement de l'éclairage.

L'apprentissage supervisé : Dans cette approche, les données d'entraînement sont annotées avec des informations d'éclairage spécifiques. Les réseaux neuronaux sont entraînés à partir de ces données pour prédire l'éclairage optimal pour une image donnée, en se basant sur les caractéristiques de l'image.

L'apprentissage non supervisé : Dans cette approche, les données d'entraînement ne sont pas annotées avec des informations d'éclairage spécifiques. Les réseaux neuronaux apprennent à partir de ces données pour découvrir et apprendre les modèles d'éclairage intrinsèques, ce qui leur permet d'ajuster l'éclairage des images de portrait de manière non supervisée.

2.4.3 Techniques de rendu inverse des images de portrait

IBR (image based rendering) est une technique permettant de créer des modèles 3D photo réalistes à partir des images 2D, et ces modèles sont utilisés pour restituer de nouvelles images à partir de différents points de vue. Comme il influence la qualité des images produites, l'éclairage de la scène est un élément crucial de l'IBR. L'utilisation des données d'éclairage de l'image d'origine pour éclairer le modèle 3D est une méthode d'éclairage de scène.

L'approche de rendu basée sur la physique est une autre méthode d'éclairage de scène dans IBR, qui reproduisent le comportement de la lumière dans une scène sur la base de principes physiques. Par rapport à la méthode basée sur l'image, le rendu basé sur la physique peut fournir des effets d'éclairage plus précis et réalistes.

La technique suggérée utilise une bibliothèque de scans faciaux 3D et un modèle statistique de visages 3D. Il est possible de créer de nouveaux modèles de visages 3D compatibles avec les statistiques rapportées en utilisant le modèle statistique pour refléter les différences de forme et de texture des visages qui sont visibles dans la base de données.

L'approche suggérée extrait d'abord une collection de points de repère de l'image qui correspondent à des traits faciaux importants (tels que les coins des yeux, la pointe du nez, etc.) afin de créer un modèle 3D à partir d'une image 2D. La position et le contour du visage sont ensuite déduits des repères 2D en utilisant une variété d'algorithmes. La forme prédite est ensuite modifiée pour refléter la texture du visage telle que montrée sur la photo. En projetant l'image de texture 2D sur la surface du modèle 3D, une méthode connue sous le nom de cartographie de la texture est utilisée pour y parvenir.

Le rendu inverse d'images de portraits à partir du modèle 3D morphable (3DMM) implique le processus d'estimation des paramètres du 3DMM qui représentent le mieux une image d'entrée donnée

2.4.3.1 Modèle 3D morphable (3DMM)

3DMM [4] une nouvelle technique de modélisation de visages 3D texturés est introduite. Il est un modèle statistique qui représente les variations de forme et de texture des visages humains en 3D. 3DMM fournit un paramétrage compact de la géométrie et de l'apparence du visage, permettant la synthèse, l'analyse et la manipulation de visages 3D réalistes.



FIGURE 2.4 – Exemple d'applique 3DMM sur des images 2D [4]

Les visages 3D peuvent être générés automatiquement à partir d'une ou plusieurs photographies, soit être modélisés directement via une vue intuitive. Les utilisateurs sont assistés dans deux problèmes clés de la modélisation de visages assistée par ordinateur. Premièrement, de nouvelles images de visages ou de nouveaux modèles de visages en 3D peuvent être enregistrés automatiquement en calculant une correspondance dense et univoque avec un modèle de visage interne. Deuxièmement, l'approche régule le caractère naturel des visages modélisés en évitant les visages à l'apparence d'apparence "improbable". À partir d'un ensemble de modèles de visages en 3D, nous dérivons un modèle de visage morphable (3DMM) en transformant l'apparence des visages.

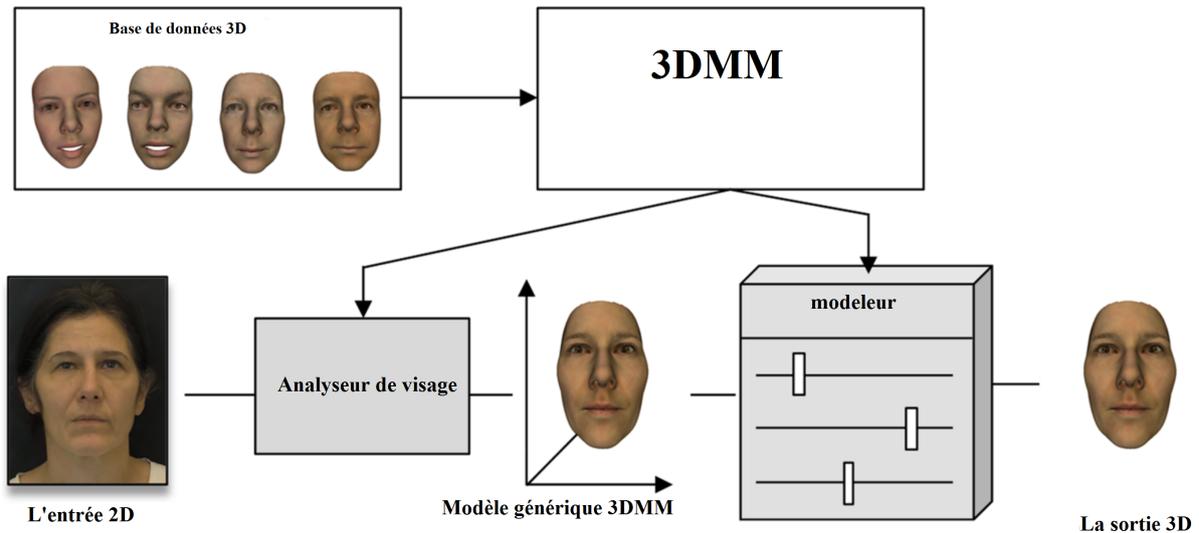


FIGURE 2.5 – Le processus de3DMM [16]

Le cadre général de la reconstruction du visage en 3D d'un individu est illustré à la figure 2.5. Pour commencer, un modèle de visage morphable en 3D est généré à partir d'un ensemble de scans de visage en 3D. Ensuite, à partir de l'image 2D d'entrée, le modèle appris extrait les caractéristiques et estime les paramètres correspondants du modèle 3D morphable.

De nombreuses méthodes de rendu inverse pour les images de portrait ont été proposées .

1. **Les méthodes basées sur l'optimisation** : Ces méthodes décomposent une image de portrait en fonction de la réflectance, de la géométrie et de l'éclairage. Une image de portrait éclairée peut ensuite être rendue en modifiant l'éclairage et en conservant la géométrie et l'éclairage ,ces méthodes prennent beaucoup de temps
2. **Les méthodes basées sur l'apprentissage profond** :dans ce cas en changeant l'éclairage et en gardant la géométrie et la réflectance fixes.

2.5 Estimation normale

Il existe plusieurs méthodes pour estimer les normales à partir d'une image de portrait, c'est-à-dire aider pour reconstruire un visage 3D à partir de telle que :

2.5.1 Raffinement normale basé sur 3D Dense Face Alignment (3DDFA)

3DDFA est une technique très appréciée pour la reconstruction 3D du visage, pour obtenir la carte de profondeur, le modèle 3D et les points de repère du visage à partir d'une seule image de visage, cette technique produit des paramètres de forme 3DMM qui peuvent être utilisés pour créer des images normales de portrait à n'importe quelle résolution, même si 3DDFA prend en compte l'expression du visage lors de l'ajustement du 3DMM. Mais les normales prédites ne peuvent toujours pas être précisément associées à l'image du portrait parce que le 3DMM ne peut représenter qu'un ensemble fini de changements dans la géométrie des faces car il est basé sur un ensemble fini de faces pour résoudre ce problème nous suggérons d'utiliser une technique de raffinement normal basée sur ARAP pour aligner les normales estimées avec l'image de portrait afin d'éviter les artefacts dans les photos de relief.

Pour éviter les artefacts dans les images de relief, ils proposent d'aligner les estimations normales avec l'image de portrait en utilisant une normale basée sur ARAP algorithm de raffinement.

2.5.2 Raffinement normal basé sur ARAP (As-Rigid-As-Possible)

Zhou et al présentent Une technique d'amélioration du portrait de relief à partir d'une seule photo. La méthode de déformation As-Rigid-As-Possible (ARAP) est utilisée dans leur méthode pour modifier les normales de surface du visage et fournir une illumination plus réaliste. On crée un maillage par l'utilisation des paramètres 3DMM prédit par 3DDFA, ensuite on projette la carte de réflectance générique du modèle 3DMM sur ce

maillage pour obtenir la "réflexion" l'image du portrait.

Après cela, ils ont utilisé un algorithme pour calculer avec précision la position des marques faciales pour trouver 68 points de visage. La « réflexion » c'est ça. Afin de construire un filet triangulaire sur la « réflexion », ces 68 visages découverts points de référence et les 198 points uniformément échantillonnés autour des bords de l'image sont joints en tant que « points d'ancrage » et se basent sur la triangulation de Delaunay. De même, un filet triangulaire est fabriqué pour le portrait.

La fonction de distorsion estimée par ARAP (warp function) est ensuite appliquée aux normales de face estimées par 3DDFA pour générer des normales raffinées comme illustré à la figure 2.3 .

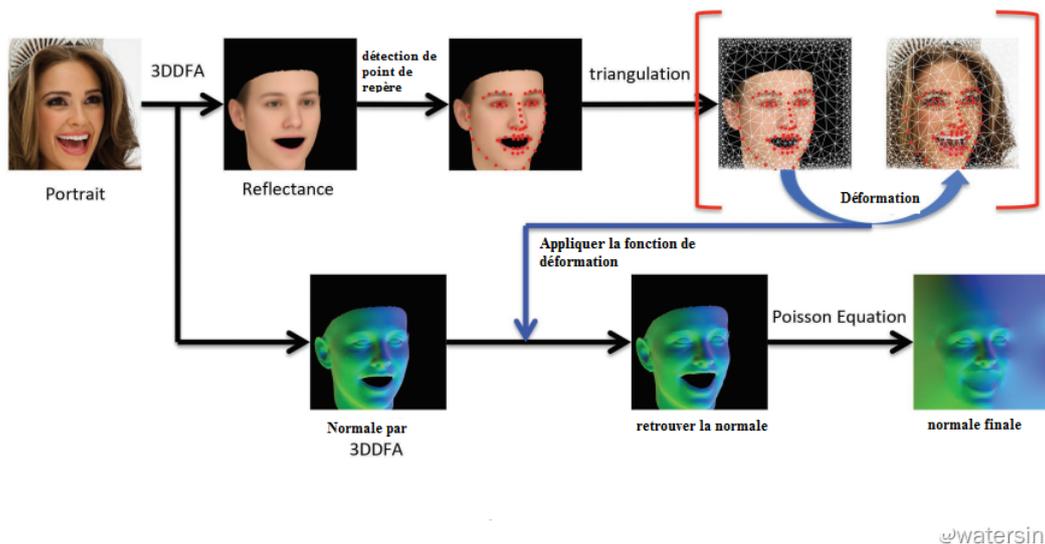


FIGURE 2.6 – Le processus de raffinement normal basé sur ARAP.[28]

Nous montrons le visage original recouvert de normal estimée par 3DDFA dans la figure suivante.

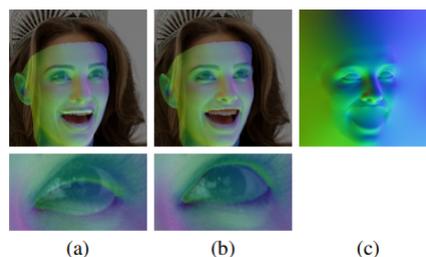


FIGURE 2.7 – Comparaison entre 3DDFA et ARAP. [28]

(a) et le normale raffinée proposée dans (b). La deuxième ligne des points a) et b) indique la région de l'œil droit. c) affiche la carte normale finale.

Pour démontrer l'efficacité du raffinement normal proposé, la méthode est remplacée par des estimations de normales 3DDFA, et le standard raffiné (ARAP) avec l'image originale, et montrer a) et b) respectivement. Il est clair que la qualité de l'alignement de la normale de l'image de portrait Les yeux et la bouche s'améliorent considérablement par la méthode (ARAP) proposée.

Nous voyons que le désalignement des zones de l'oreille et du cou ne peut pas être corrigé en utilisant notre approche de raffinement normal ARAP suggérée. C'est parce que, au mieux de nos connaissances, il n'y a pas de technique d'identification historique pour les oreilles et le cou, 3DMM est incapable de représenter avec précision la déformation de l'oreille et du cou. Nous éliminons les zones de l'oreille et du cou des normes révisées pour empêcher les artefacts potentiels dans les photos de relit. on résolvons une équation de Poisson pour compléter les normales manquantes pour l'oreille, le cou, la bouche et la région de fond afin d'obtenir une image normale complète. Les normales sont représentées après que la région manquante ait été remplie.

L'intelligence artificielle évolue rapidement, permettant aux robots de devenir de plus en plus sophistiqués. Les machines découvrent de nouvelles façons de nous aider à effectuer des tâches plus complexes parce qu'elles ont le potentiel d'évoluer de manière indépendante en fonction des données qui leur sont présentées.

En intelligence artificielle, le système Generative Adversarial Networks (GAN) [13] est une approche très prometteuse qui a déjà produit des résultats étonnants dans une variété de domaines. Les GAN sont principalement utilisés pour générer des images.

2.6 les Réseaux antagonistes génératifs

Un réseau adversaire génératif (GAN) [13] est une approche d'apprentissage non supervisée puissante et largement utilisée dans la vision informatique.

Cette forme de méthode d'apprentissage automatique qui se compose de deux réseaux neuronaux distincts : le générateur et le discriminateur, les deux réseaux s'entraînent

l'un contre l'autre pour générer de fausses images et générer des données réalistes qui sont difficiles à distinguer des données réelles.

2.6.1 Architecture générale des GANs

L'objectif de cette architecture est de permettre l'éclairage des portraits à partir d'une seule photographie en recueillant les complexités de la texture du portrait et en créant des photos réalistes qui semblent avoir été prises dans des situations de lumière diverses.

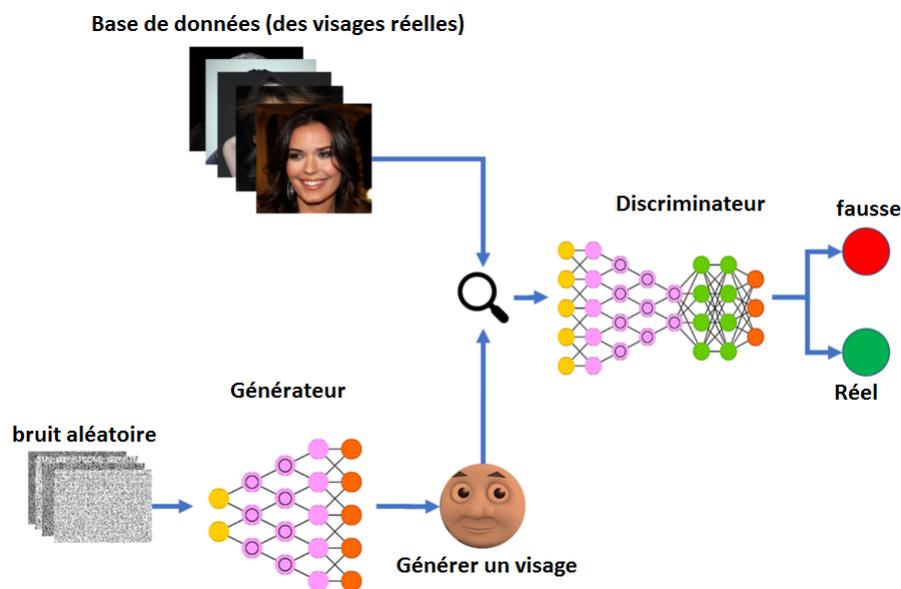


FIGURE 2.8 – Architecture générale des GANs[2]

Dans l'architecture générale des GAN, le générateur et le discriminateur sont conduits de manière antagoniste.

Le générateur cherche à produire des photos de portrait réalistes et intéressantes, tandis que le discriminateur s'efforce d'identifier les images générées à partir de celles authentiques.

Une fonction de perte qui comprend à la fois l'erreur de génération du générateur et l'erreur de discrimination du discriminateur est utilisée pour propulser le générateur ainsi que le discriminateur en même temps.

2.6.1.1 Générateur

Dans la génération de portraits d'image unique profonde à l'aide de GAN, le générateur joue un rôle crucial dans la création d'images de portrait réalistes et détaillées à partir de bruits aléatoires ou de vecteurs latents.

Architecture générale

Le générateur est construit comme un réseau neuronal profond, avec des couches convolutionnelles et déconvolutionnelles (également connues sous le nom de convolutions de transposition) c'est-à-dire un réseau de type codeur décodeur. Le vecteur latent sert d'entrée au générateur et est généralement un vecteur de bruit aléatoire de haute dimension. Il augmente les dimensions spatiales des données en utilisant des procédures d'échantillonnage (telles que les convolutions de transposition ou l'interpolation du voisin le plus proche) pour les rapprocher de la taille de sortie souhaitée pour l'image de portrait.

Pour extraire des caractéristiques essentielles de la représentation échantillonnée, les couches convolutionnelles de le générateur aident à capturer les nuances et les motifs nécessaires pour créer des caractéristiques et des structures faciales réalistes.

Après chaque couche (jusqu'à la couche de sortie), des fonctions d'activation telles que ReLU ou LeakyReLU sont appliquées pour ajouter de la non-linéarité et permettre au réseau d'apprendre des cartographies compliquées entre le vecteur latent et l'image de portrait. mais dans la couche de sortie du réseau du générateur aucune fonction d'activation est utilisée. Au lieu de cela, il produit l'image créée comme la sortie finale du réseau du générateur. Les valeurs de pixel dans l'image résultante produite par la couche de sortie peuvent varier en fonction de la normalisation des données $[0, 1]$ ou $[-1, 1]$ utilisée pendant la formation. Selon les procédures de normalisation employées dans la phase de formation.

L'image est encore raffinée par le générateur en utilisant des couches convolutionnelles ou déconvolutionnelles supplémentaires. Cela permet au réseau d'augmenter la qualité des caractéristiques produites, d'améliorer les détails du visage et de générer des textures plus réalistes (voir (b) dans la figure 2.7).

2.6.1.2 Discriminateur

Le réseau de discriminateurs dans la génération de portraits d'images simples profondes à l'aide de GAN joue un rôle critique dans la distinction entre les images de portrait réelles et générées. Sa fonction principale est d'évaluer le réalisme et la qualité des images générées par le générateur

Architecture générale :

Le réseau de discriminateurs est souvent construit comme un réseau neuronal profond avec des couches convolutives. L'architecture spécifique peut différer en fonction de la complexité de la tâche et de l'ensemble de données utilisé.

Le réseau de discriminateur reçoit soit des photos de portrait réelles à partir de l'ensemble de données de formation, soit des images créées à partir du réseau du générateur. Ces photos sont souvent scalées et pré-traitées pour répondre aux critères d'entrée du discriminateur .

Le réseau de discriminateurs analyse l'image d'entrée en utilisant de nombreuses couches d'algorithmes convolutifs. Chaque couche extrait et apprend des aspects hiérarchiques de l'image d'entrée afin de capturer ses propriétés visuelles, Après chaque couche convolutionnelle, des fonctions d'activation telles que ReLU (Rectified Linear Unit) ou LeakyReLU (Leaky Rectified linear unit) sont fréquemment appliquées pour introduire la non-linéarité et permettre au réseau discriminateur d'apprendre des cartographies complexes entre les images d'entrée et leur authenticité.

Pour minimiser les dimensions spatiales de l'image d'entrée, le discriminateur peut effectuer des techniques d'échantillonnage. (such as pooling or strided convolutions). Cela permet au réseau de collecter des caractéristiques globales et de haut niveau.

La couche de sortie du réseau discriminateur est souvent composée d'un seul neurone ou d'une couche dense avec une fonction d'activation sigmoïde. Cette couche génère un résultat de classification binaire (0 ou 1), 1 si l'image d'entrée est réelle ,sinon 0 c'est une image fabriquée(fake).

2.6.2 Entraînement de GANs

L'entraînement du générateur et du discriminateur sont répétées de manière itérative pour améliorer les performances du réseau. L'entraînement de GAN se fait en plusieurs étapes itératives.

1. Collecte de données :

Recueillez un ensemble de données de portrait authentiques qui seront utilisées comme référence pendant la formation. Pré-traitement des images en modifiant, normalisant et effectuant toute autre préparation requise.

2. Initialisation :

définissez les poids du générateur et du discriminateur au hasard.

3. Formation du générateur :

fournir au générateur un bruit latent aléatoire ou un vecteur latent, puis générer des images. Montrez au discriminateur ces images créées et construisez une fonction de perte générale basée sur les prédictions du discriminateur. Le but du générateur est de tromper le discriminateur en générant des portraits indistincts du vrai portrait.

4. Formation de discriminateur :

montrer au discriminateur une collection d'images authentiques du groupe de données ainsi qu'une collection d'images créées par le générateur à partir du bruit latent. Le discriminateur crée une fonction de perte discriminatoire basée sur ces prédictions et les étiquettes d'accompagnement après avoir déterminé si chaque image est réelle ou fabriquée.

5. Reproduction et mise à jour des poids de discriminateur et de générateur :

On Calcule les gradients de la fonction de perte du discriminateur par rapport à ses poids, puis effectuez une mise à jour des poids à l'aide d'un algorithme d'optimisation.

Faites de même pour le générateur en calculant les gradients de la fonction

de perte générale. Répétez ces étapes d'entraînement pour améliorer les performances du GAN.

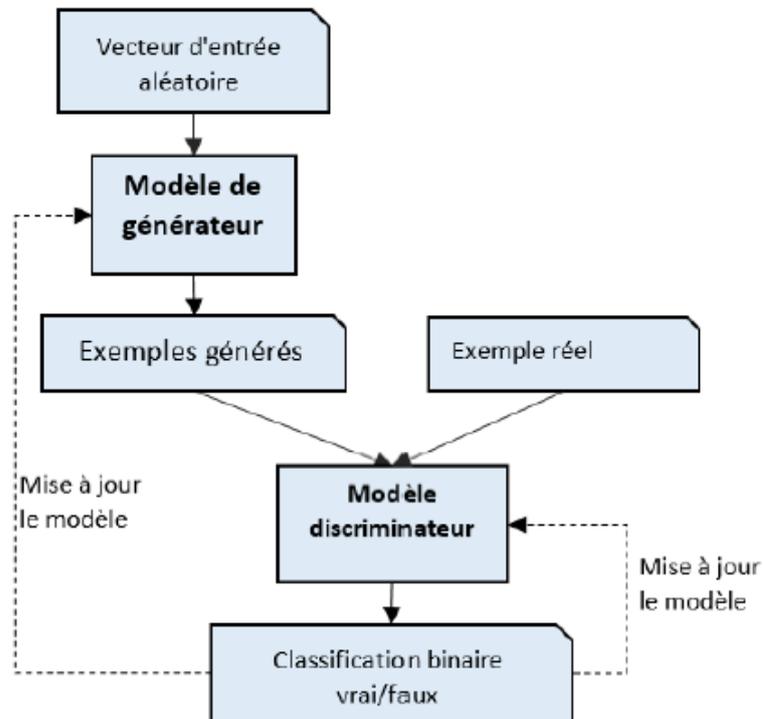


FIGURE 2.9 – Exemple d'architecture du modèle de réseau adversarial génératif [26]

2.6.3 Applications des GANs

Les GANs ont de nombreuses applications dans différents domaines.

2.6.3.1 Générer de nouvelles photos



FIGURE 2.10 – Visages générés sur Cette personne n'existe pas [26]

Une IA a totalement produit ces visages. Les GAN C'est aussi une des applications qui, en collaboration avec les artistes de l'IA, a permis au GAN de se démocratiser.

Par conséquent, l'IA doit avoir une compréhension complète de ce qu'est un visage humain. Pour cela, il faut entraîner le réseau neuronal avec des centaines de photos de visages.

2.6.3.2 Coloriser des images en noir et blanc



FIGURE 2.11 – modèle de DeOldify [26]

les GANs permettent de coloriser des images de façon très réaliste.

Plusieurs modèles, comme DeOldify par exemple, permettent de redonner vie à vos anciennes images.

2.6.3.3 Créer des œuvres d'art



FIGURE 2.12 – Œuvres d'art générées grâce aux GAN [26]

La création d'œuvres d'arts est l'application des GAN qui a fait le plus parler. Les IA se sont longtemps limitées à des tâches répétitives et sans intérêts pour l'humain.

Aujourd'hui elles investissent des domaines qui nous étaient jusque là réservés.

Même si en rentrant dans les détails on voit bien qu'il ne s'agit que de copier les styles des peintres humains, on ne peut que être étonnés des résultats que ces techniques permettent d'obtenir.

2.6.3.4 Traduction d'images



FIGURE 2.13 – Traduction d'images avec Cycle GAN [26]

La traduction d'images est une technologie qui permet d'échanger des objets, des visages ou d'autres choses dans des photos ou des vidéos. C'est d'ailleurs cette technique qui est à l'origine des deep fake. Ce sont des vidéos dans lesquels on voit des personnes dire des choses qu'elles n'ont jamais dites. C'est évidemment un danger sérieux. Et je vous conseil de le prendre en compte lorsque vous voyez des informations sur les réseaux sociaux. Ces systèmes reposent sur les GAN, avec une architecture spéciale appelées CycleGAN.

2.7 Conclusion

Nous avons présenté dans ce chapitre les notions générales sur le rééclairage d'images en générale et particulièrement celles concernant le rééclairage à partir d'une seule photo de portrait basé sur l'apprentissage profond.

Nous avons ensuite introduit le concept de l'image quotient, et présenté les différentes techniques existantes dans la littérature utilisant l'apprentissage pour rééclairer les images portraits. Nous avons enfin présenté plus en détail sur les réseaux GAN dont une variante (Hourglass network) va nous servir pour mettre en œuvre la technique que nous allons présenter dans le chapitre trois. newpage

Chapitre 3

Conception d'une approche de rééclairage de portrait à base de réseau Hourglass

3.1 Introduction

L'utilisation des techniques d'apprentissage profond pour les tâches liées aux portraits est motivée par leur capacité à révolutionner les tâches de vision par ordinateur et à améliorer la qualité et l'efficacité de la création, de l'édition ou de l'amélioration des portraits.

L'apprentissage profond a permis des avancées significatives dans la génération et la manipulation d'images, offrant ainsi de nouvelles perspectives pour les portraits.

Les méthodes traditionnelles présentent des limitations en termes d'effort manuel, de temps nécessaire et de flexibilité, ce qui rend l'utilisation de l'apprentissage profond particulièrement pertinente pour surmonter ces défis.

En exploitant des réseaux de neurones convolutionnels (CNN) et des réseaux génératifs antagonistes (GAN), entre autres architectures d'apprentissage profond, il est possible de créer des portraits réalistes et de haute qualité, d'améliorer les portraits existants, et même de réaliser des tâches spécifiques telles que la progression de l'âge ou

la manipulation des expressions.

Les techniques de portrait à partir d'une seule image offrent des applications potentielles dans les domaines du divertissement, de la publicité, de l'art numérique, de la réalité virtuelle et de la création de contenu personnalisé. Dans ce chapitre, on va présenter le domaine des portraits réalisés grâce à l'apprentissage profond.

L'éclairage d'un portrait consiste à modifier les conditions d'éclairage d'une prise de vue afin de produire de nouveaux effets de lumière. En modifiant la direction, l'intensité et la couleur de la lumière qui éclaire le visage .

3.2 Architecture de réseau Hourglass : Généralités

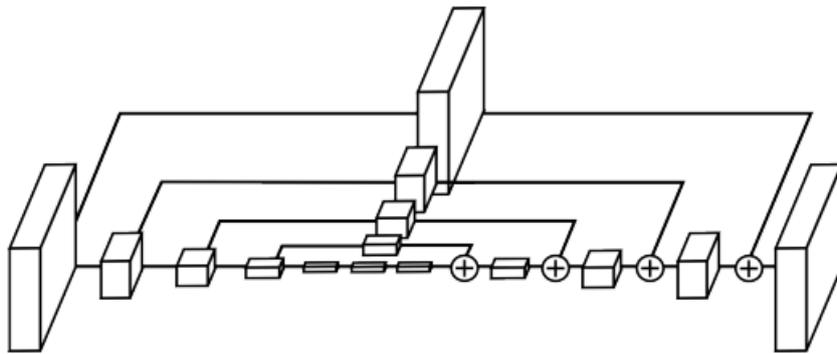


FIGURE 3.1 – Illustration d'un module "Hourglass" unique [15]

Le réseau Hourglass [15] est une architecture d'apprentissage profond (CNN) de type encodeur et décodeur ,il a d'abord été développée pour estimer les poses humaines en 2016 par par Newell et al. Il a une forme distincte qui ressemble à un sablier, d'où son nom. Cette forme caractéristique indique la structure en cascade symétrique du réseau, qui se compose de nombreux blocs de construction. La capacité du réseau Hourglass à rassembler des informations à plusieurs échelles et à combiner des éléments à plusieurs niveaux d'abstraction est largement reconnue.

3.2.1 Les blocs de construction

Les éléments constitutifs du réseau Sablier sont les composants de base de son architecture. Ces blocs sont chargés de traiter et de modifier les données entrantes ou les cartes des caractéristiques afin de permettre une analyse et une compréhension efficaces. Les couches convolutives, les couches de pooling pour le sous-échantillonnage et les couches de sur-échantillonnage pour la restauration de la résolution spatiale sont des éléments de construction courants pour les deux parties : encodeur et décodeur.

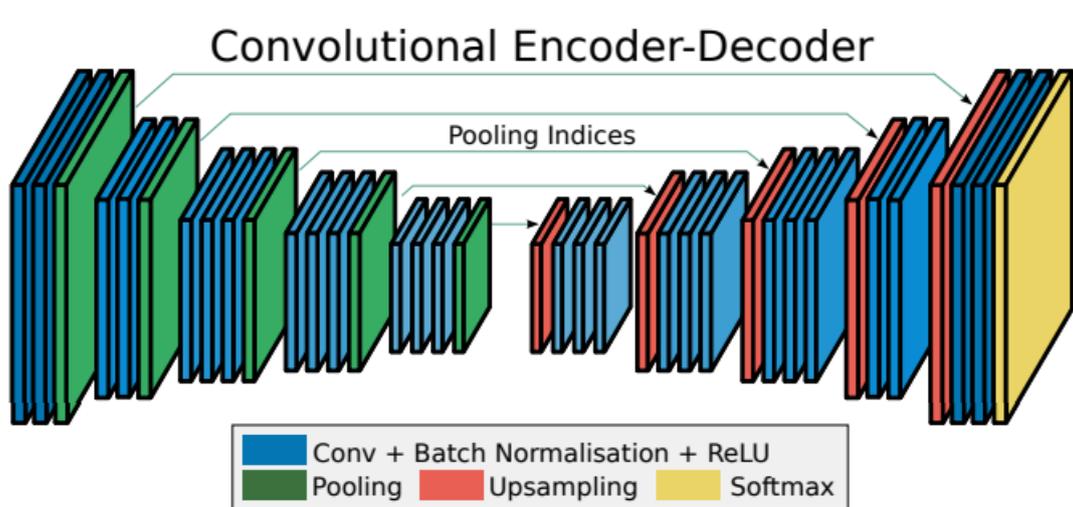


FIGURE 3.2 – Schéma représente Les blocs de construction de l'encodeur et décodeur de réseau Hourglass[24]

3.2.1.1 Couches convolutives

Dans le réseau de Hourglass, les couches convolutives jouent un rôle crucial dans la capture et le traitement des informations pour permettre une analyse précise des données visuelles.

Les couches convolutives du réseau Hourglass fonctionnent comme des extracteurs de caractéristiques. Elles reconnaissent les modèles spatiaux dans les données et extraient les caractéristiques importantes des cartes de caractéristiques. Ces couches reconnaissent les bords, les coins, les textures et d'autres motifs visuels essentiels à la compréhension de l'image d'entrée à l'aide de filtres convolutifs.

3.2.1.2 Couches sous-échantillonnage (downsampling layers)

Après chaque Couche convolution on trouve des couches de sous-échantillonnage également appelées couches de déconvolution pour réduire la taille des cartes de caractéristiques tout en conservant les informations importantes. Ces couches combinent des informations provenant de voisinages locaux pour extraire des représentations de plus haut niveau et se concentrer sur les caractéristiques pertinentes grâce à des techniques de sous-échantillonnage telles que le regroupement maximal (max pooling) et le regroupement moyen (average pooling).

Pour la technique de regroupement maximal sélectionnent la valeur maximale au sein des régions, tandis que le sous-échantillonnage spatial réduit les dimensions et élargit le champ réceptif. Les couches de mise en commun (pooling layers) peuvent être utilisées à différentes échelles pour capturer des informations multi-échelles.

Les fonctions d'activation non linéaires, telles que Re-LU, suivent les couches de mise en commun pour introduire la non-linéarité et capturer des relations complexes.

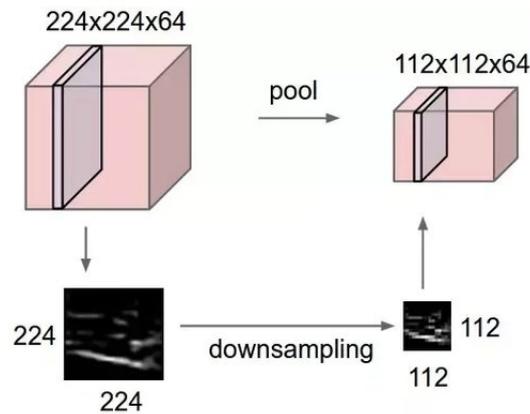


FIGURE 3.3 – Schéma explicatif de principe de couches sous-échantillonnage (downsampling layers)[3]

3.2.1.3 Couches sur-échantillonnage (upsampling layers)

Les couches de sur-échantillonnage du réseau Hourglass sont essentielles pour augmenter la taille des cartes de caractéristiques et récupérer les détails spatiaux perdus lors du sous-échantillonnage. Elles utilisent des techniques telles que la déconvolution et les filtres de sur-échantillonnage apprenables pour réaliser la multiplication et l'agrégation

par pixel.

Le réseau peut modifier la résolution en utilisant des facteurs de sur-échantillonnage fixes ou pouvant être appris, ce qui lui confère une certaine souplesse. Le réseau protège les informations à haute résolution tout en incorporant les caractéristiques échantillonnées en mélangeant les couches d'échantillonnage avec les connexions de saut.

Cette fusion d'informations multi-échelles améliore la capacité du réseau à collecter le contexte global ainsi que les caractéristiques à grain fin, ce qui se traduit par une amélioration des performances dans des tâches telles que la localisation d'objets et la segmentation sémantique.

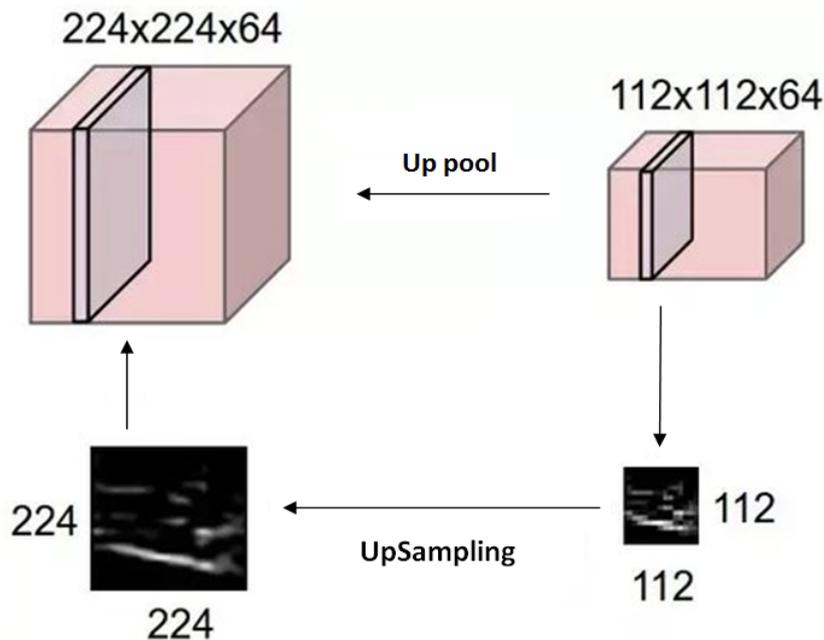


FIGURE 3.4 – Schéma explicatif de principe de couches sur-échantillonnage (upsampling layers). [3]

3.2.1.4 Sauts de connexion

Les sauts de connexion sont des liens directs entre les parties de l'encodeur et décodeur du réseau, ce qui permet aux données d'éviter certaines couches intermédiaires.

La fonction des connexions de saut dans le réseau Hourglass est d'accélérer le flux d'informations des niveaux antérieurs de l'encodeur vers les couches équivalentes du décodeur. Ces liens permettent au réseau de collecter des données contextuelles locales et globales, aussi ces liens contribuent à améliorer l'efficacité et la précision globales

du modèle en conservant les données à grain fin de l'encodeur et en les intégrant aux informations sémantiques de haut niveau du décodeur

Les connexions de saut du réseau sont généralement construites à l'aide de procédures de sommation par éléments . La mise en œuvre d'une connexion de saut peut varier en fonction des besoins uniques de la tâche et de la conception architecturale.

3.3 Proposition d'une application utilisant le réseau Hourglass pour le ré-éclairage des images portrait basé sur l'apprentissage profond

Le ré-éclairage d'un portrait [7] unique est un problème qui consiste à utiliser des techniques d'apprentissage profond pour modifier les conditions d'éclairage d'une photographie de portrait unique. L'objectif de cette application est de modifier l'éclairage d'une image faciale tout en conservant son aspect général et sa cohérence visuelle.

Pour atteindre le but , premièrement on va crée un ensemble de données .

3.3.1 Génération de base de données (Deep Portrait Relighting Dataset "DPR")

La création d'un ensemble de données approprié est une étape cruciale dans l'entraînement du réseau Hourglass . Un ensemble de données bien préparé garantit que le réseau apprend et se généralise efficacement, Deep Portrait Relighting (DPR) c'est l'ensemble de données de réseau Hourglass . Il est destiné à fournir une collection de photos de portrait à grande échelle et haute résolution 1024x1024, qui sont obtenues dans des paramètres naturels plutôt que dans des réglages contrôlés. pour la génération de DPR on utilisons un algorithme basé sur Ratio-Image.

L'ensemble de données DPR est développé en plus du groupe de données CelebA, notamment la version de haute résolution Celeb A-HQ. CelebA 3.5est un ensemble de données populaires dans le domaine de la vision informatique pour les applications de

reconnaissance faciale. Il se compose de 30 000 images de visage en résolution 1024 x1024, ces images ont été annotées avec des caractéristiques différentes telles que les points de repère du visage, les identités et les qualités comme l'âge.



FIGURE 3.5 – Exemples de la base CelebA-HQ [26]

Après la suppression des photos sur lesquelles le détecteur de points de repère n'a pas réussi à reconnaître les points de repère, l'ensemble de DPR final contient 27 627 images. Parce qu'elles comportent des annotations de points de repère précises et fiables, ces photos sont considérées comme idéales pour le ré éclairage de portraits. ensuite on va généré 138 135 photos relit en choisissant au hasard 5 conditions d'éclairage pour chacune de ces images à partir un algorithme de détection de l'éclairage qui est déjà entraîné à l'aide du collection d'images capturées sous différents éclairages .

le détecteur de points de repère (voir la figure 3.6)est un algorithme ou un modèle de vision par ordinateur conçu pour localiser et identifier des points spécifiques ou des points de repère sur un visage humain, Ces points de repère correspondent généralement à des caractéristiques faciales clés telles que les yeux, le nez, la bouche et d'autres structures faciales proéminentes .la détection des points de repère est presque entièrement résolue par des méthodes d'apprentissage automatique qui sont entraînées sur un ensemble de données d'images annotées à la main.

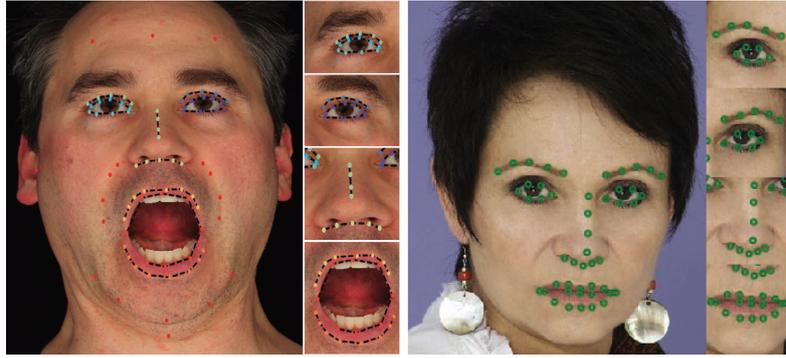


FIGURE 3.6 – Détection de points de repère sur un visage

L'ensemble de données DPR offre une meilleure qualité d'annotations en filtrant les photos dont les détections de points de repère ont échoué, ce qui est essentiel pour le ré-éclairage de portraits qui s'appuient sur des points de repère faciaux corrects.

3.3.2 Architecture de réseau Hourglass

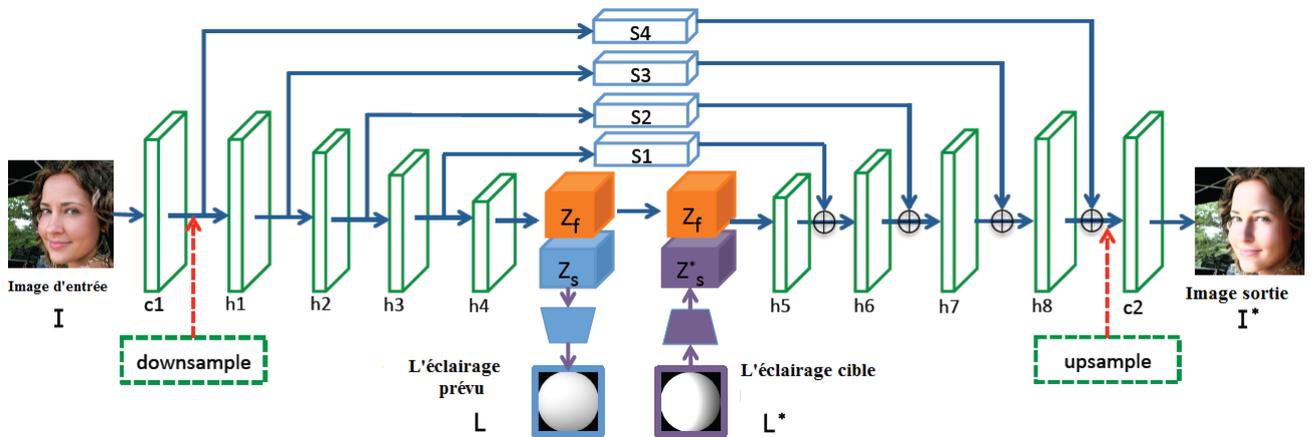


FIGURE 3.7 – La structure de réseau Hourglass

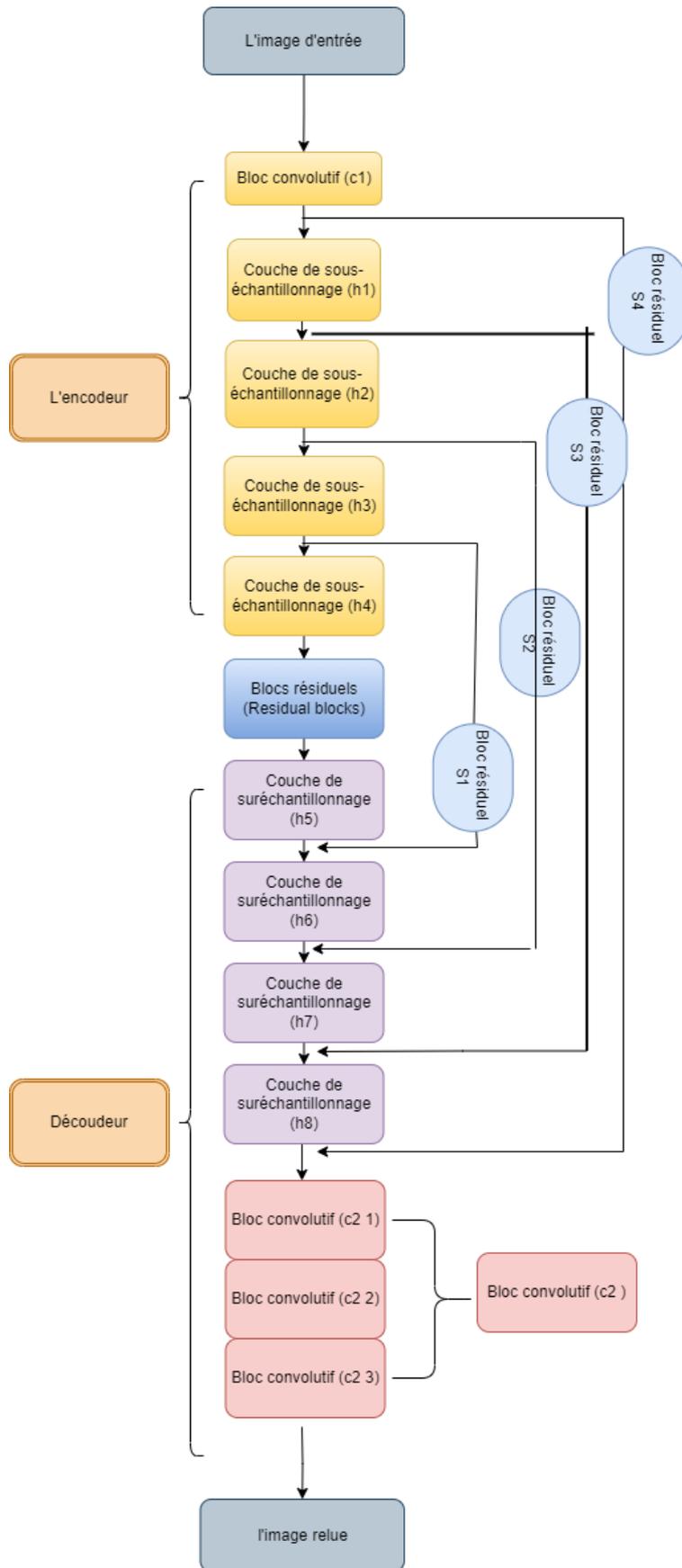


FIGURE 3.8 – Schéma explicatif du contenu de réseau Hourglass

L'objectif de cette application est de modifier l'éclairage d'une image faciale tout en conservant son aspect général et sa cohérence visuelle.

Les solutions traditionnelles de rééclairage de portraits exigent des modifications manuelles sophistiquées ou reposent sur un ensemble restreint de paramètres d'éclairage spécifiés. Les techniques d'apprentissage profond, quant à elles, utilisent des réseaux neuronaux pour apprendre la correspondance entre les photos de visage en entrée et leurs équivalents éclairés.

L'architecture du réseau Hourglass a été présentée comme une solution pour le rééclairage approfondi de portraits individuels.

mon proposition de réseau Hourglass se compose de deux parties : un encodeur qui extrait les caractéristiques de l'image du visage d'entrée et un décodeur qui crée l'image du visage ré éclairé, avec quatre les connexions de saut sont utilisées pour connecter les fonctionnalités à différents échelles dans la partie encodeur à leur échelle correspondante dans la partie décodeur .

Dans notre application L'encodeur prend en entrée une image de visage (I) et une condition d'éclairage cible (L^*). ensuite les caractéristiques de visage extraites par les couches convolutives dans l'encodeur sont divisées en deux parties : les caractéristiques du visage (Z_f) et les caractéristiques de l'éclairage (Z_s).

Les caractéristiques du visage (Z_f) reflètent les qualités de l'image du visage qui sont indépendantes des conditions d'éclairage. Ces caractéristiques capturent la structure intrinsèque du visage, la texture et d'autres caractéristiques faciales qui ne sont pas affectées par les changements de lumière. Le réseau peut se concentrer sur la modification de l'éclairage tout en conservant les aspects vitaux du visage en séparant les éléments du visage.

Les caractéristiques d'éclairage (Z_s) de l'image du visage recueillent des informations sur les conditions d'éclairage. Ces caractéristiques représentent la distribution géographique et l'intensité de l'éclairage sur le visage. Un réseau de régression de l'éclairage est utilisé pour prévoir l'éclairage (L) de l'image du visage d'entrée. Ce réseau utilise les caractéristiques d'éclairage (Z_s) comme données d'entrée pour estimer l'état de l'éclairage.

rage.

L'éclairage cible (L) est mis en correspondance avec les caractéristiques d'éclairage (Z_s) à inclure dans le processus de ré éclairage. Cette correspondance garantit que le réseau dispose des informations d'éclairage nécessaires tout au long du processus de ré éclairage.il peut modifier l'éclairage de la photo du visage en associant l'éclairage cible aux attributs d'éclairage.

3.3.3 Entraînement de réseau Hourglass

Dans notre proposition , on décrivons l'approche de préparation des données pour l'entraînement du réseau Hourglass pour le ré éclairage profond de portraits uniques dans la section 3.3.1 . on commençons par l'ensemble de données CelebA-HQ pour sélectionner une image source I_s et une image cible I_t au hasard à partir des cinq images relit créées pour chaque image dans le groupe de données CelebA-HQ. L'éclairage équivalent de SH (Harmonie sphérique) est également choisi pour les images source et cible (I_s et I_t).pour génère l'image relit I_t et l'éclairage prédit I_s . L'image relit de référence I_t et l'éclairage de référence I_s sont utilisés pour superviser le processus d'apprentissage c'est à dire l'image relue générée I_t est comparée à l'image relue de référence I_t .

Entre l'image de relit produite I_t et L'image relit de référence I_t .aussi L'éclairage de référence I_s représente les conditions d'éclairage réelles de l'image source I_s . La sortie du réseau comprend l'éclairage prédit L^*_s , qui est comparé à l'éclairage réel L_s au cours de l'apprentissage on applique deux fonctions de perte(Loss functions) :

1. **La perte L1** : est calculée la différence de pixels entre l'image relue produite I_t et l'image relue de référence I_t , en orientant le réseau vers la création d'images relues qui correspondent étroitement aux conditions d'éclairage prévues.

$$\mathcal{L}_I = \frac{1}{N_I} (\|\mathbf{I}_t - \mathbf{I}_t^*\|_1 + \|\nabla \mathbf{I}_t - \nabla \mathbf{I}_t^*\|_1) + (\mathbf{L}_s - \mathbf{L}_s^*)^2,$$

2. **La perte L2** : pour l'éclairage prédit L est calculée entre l'éclairage prédit I_s et l'éclairage réel I_s . Cette perte mesure la différence entre l'éclairage prédit et

l'éclairage réel, ce qui permet au réseau d'apprendre à estimer les conditions d'éclairage avec précision.

Alors les images produites par la technique de l'image de rapport peuvent contenir des erreurs dues à une mesure incorrecte de la normalité du visage ou de l'éclairage. donc on utilisons une perte GAN (L-GAN) pour surmonter ce problème et améliorer la qualité des images produites.

$$\mathcal{L}_{GAN} = \mathbb{E}_{\mathbf{I}}(1 - D(\mathbf{I}))^2 + \mathbb{E}_{\mathbf{I}_s} D(G(\mathbf{I}_s, \mathbf{L}_t))^2$$

où :

\mathbf{I} est l'image réelle, G et D représentent respectivement notre réseau de ré-éclairage et notre discriminateur. et le discriminateur respectivement. Nous utilisons 1 comme étiquette pour les images réelles et 0 pour les images truquées .

Alors,La perte globale L pour notre réseau est une combinaison linéaire des pertes mentionnées dans cette section :

$$L = L_I + L_{GAN} + \lambda L_F$$

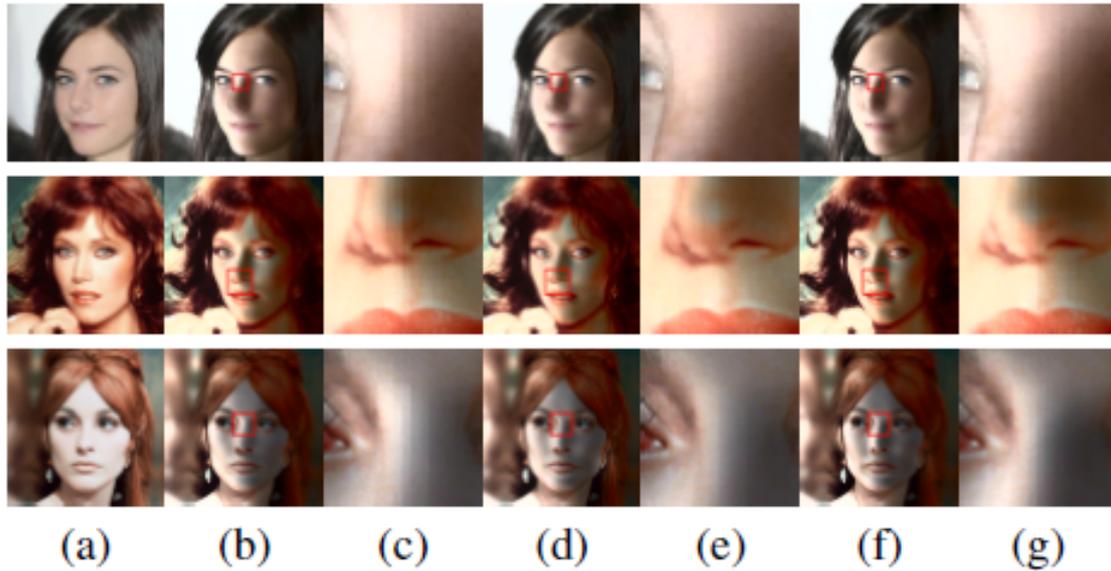


FIGURE 3.9 – (a) montre l'image d'entrée, (b), (d) et (f) sont des images générées en utilisant L_I , $L_I + L_{GAN}$ et $L_I + L_{GAN} + L_{GAN}$. images générées en utilisant L_I , $L_I + L_{GAN}$ et $L_I + L_{GAN} + L_f$ respectivement. L_f respectivement ; (c), (e) et (g) sont la région du rectangle rouge de (b), (d) et (f) respectivement. Notez le bord au milieu du bruit généré par L_I .

Les quatre sauts de connexion (S1 S2 S3 S4) qui relient les étapes correspondantes entre les parties encodeur et décodeur (voir la figure 3.7) permettent l'extraction et du traitement des caractéristiques au sein du réseau lorsque les données d'entrée passent par le réseau, les caractéristiques sont sous-échantillonnées et sur-échantillonnées à chaque étape, ce qui permet d'extraire des informations à plusieurs échelles. Ces connexions permettent de conserver et d'intégrer les informations de bas niveau et de haut niveau, améliorant ainsi la capacité du réseau à capturer les détails les plus fins et le contexte global (voir la figure 3.8).

Les caractéristiques récupérées à chaque étape (S1, S2, S3, S4) sont fusionnées et utilisées pour améliorer le résultat final. Le réseau peut rassembler et interpréter efficacement des informations multi-échelles en utilisant la représentation hiérarchique des caractéristiques à plusieurs tailles, ce qui se traduit par de meilleures performances dans des tâches telles que l'identification d'objets, la prédiction de la posture et la segmentation sémantique.

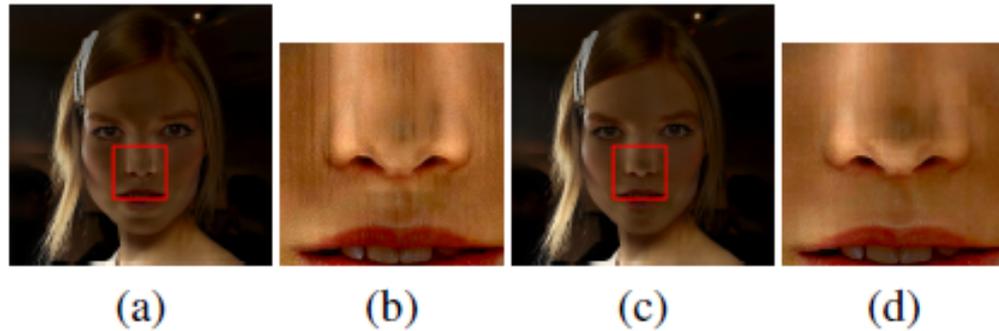


FIGURE 3.10 – (a) sortie du réseau Hourglass, (b) région rectangulaire de (a), (c) sortie du réseau Hourglass. région rectangulaire de (a), (c) sortie du réseau Hourglass avec (d) région rectangulaire de (c). Nous augmentons l'intensité. Augmentation de l'intensité des pixels de (b) et (d) pour une meilleure visualisation.

la technique d'entraînement par saut préserve plus d'informations sur les visages dans la couche de caractéristiques que le réseau de Hourglass standard. dans la figure 3.9 on montre également que l'apprentissage par saut peut améliorer la qualité des résultats produits en minimisant les artefacts dans la zone du nez par exemple.

Pendant la phase d'entraînement du réseau Hourglass, les photos réelles de l'ensemble de données FFHQ sont utilisées comme référence ou images "réelles" dans la perte GAN. L'ensemble de données FFHQ est un ensemble de données de haute qualité qui comprend une variété de visages humains photographiés dans différents contextes d'éclairage.

Lorsque des photos de l'ensemble de données FFHQ sont utilisées dans la perte GAN, le réseau peut comparer et aligner la distribution des parcelles d'images locales des images relues produites avec la distribution des parcelles des images réelles. Cet alignement garantit que les photos produites sont aussi proches que possible de l'apparence naturelle des photos réelles, y compris les changements d'illumination trouvés dans l'ensemble de données FFHQ.

Une perte de correspondance des caractéristiques qui est ajoutée à la perte globale (L) après dix époques (le nombre total de époques quatorze au cours de la formation). L est également proposée pour augmenter la précision des images d'une même personne

dans des conditions d'éclairage différentes comme suit :

$$\mathcal{L}_F = \frac{1}{N_F} (\mathbf{Z}_{f1} - \mathbf{Z}_{f2})^2$$

où :

I_{s1} et I_{s2} , et N_F est le nombre d'éléments dans Z_f , I_{s1} et I_{s2} , N_F est le nombre d'éléments dans Z_f .

L'optimiseur Adam est utilisé pour former le réseau. cette algorithme d'optimisation combine les avantages de deux autres algorithmes d'optimisation, **AdaGrad** et **RM-Sprop**, pour optimiser efficacement les réseaux neuronaux profonds. Il maintient des taux d'apprentissage adaptatifs pour chaque paramètre et utilise des estimations des premiers et seconds moments des gradients pour mettre à jour les paramètres de réseau hourglass.

Dans cette proposition on choisit d'utiliser les LSGANs (Least Squares GAN) au lieu des GANs réguliers pour améliorer la qualité d'image relie qui peut être flou.car :

1. les LSGAN sont capables de générer une qualité supérieure des images que des GANs réguliers. Deuxièmement,
2. les LSGAN sont plus stables pendant Le processus d'apprentissage.
3. LSGAN (Least Squares GAN) se concentre principalement sur la formulation de la perte plutôt que d'introduire des changements architecture aux spécifiques par rapport aux GAN traditionnels.

3.4 LSGANs(Réseaux adversariaux génératifs à moindres carrés)

Les GAN ordinaires utilisent la fonction de perte d'entropie croisée sigmoïde lorsqu'ils considèrent le discriminateur comme un classificateur. Comme indiqué dans la section 1, lorsque le générateur est mis à jour, cette fonction de perte entraîne une disparition des gradients pour les échantillons qui se trouvent du bon côté de la limite

de décision, mais qui sont encore éloignés des vraies données. Les réseaux adverbiaux génératifs des moindres carrés (LSGAN) sont proposés comme solution à ce problème. Supposons que nous utilisons le schéma de codage a-b pour le discriminateur, a et b représentant respectivement les étiquettes des données fausses et authentiques. Les fonctions objectives des LSGAN sont les suivantes :

Pour traiter spécifiquement ces artefacts locaux, dans cette proposition on utilise un GAN Patch ,on utilise un schéma explique son principe (voir la figure 3.11)

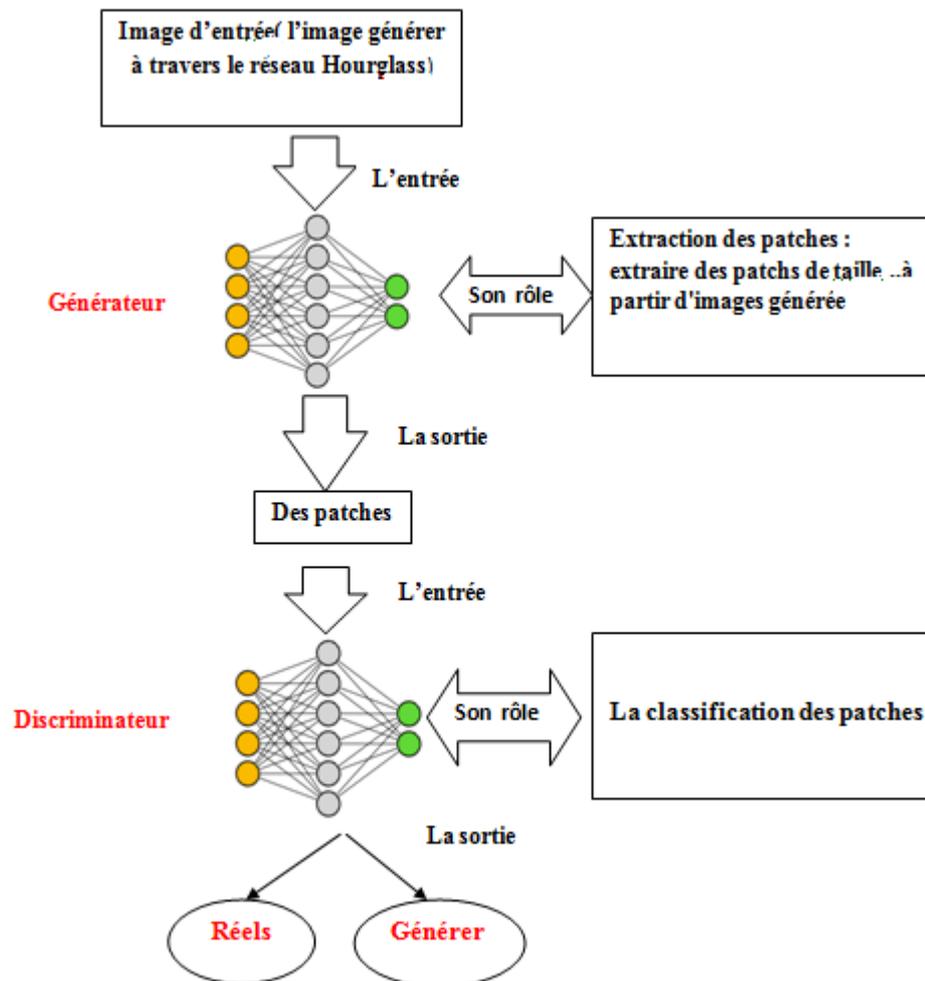


FIGURE 3.11 – Schéma explicatif d'un patch GAN

LSGAN se concentre sur l'évaluation et l'amélioration de la qualité des patches locaux de l'image. Au lieu de considérer l'ensemble de l'image. dans le schéma, le discriminateur du GAN Patch évalue le réalisme des patches individuels dans les images générées (0) et réelles(1).

3.5 Conclusion

Nous avons présenté dans ce chapitre motivations qui nous ont poussées à intégrer l'apprentissage profond dans le domaine du rendu, par l'exposé d'une technique de rééclairage appliquée aux portraits à image unique. À cette fin, cette approche basée sur la physique crée un ensemble de données à grande échelle et de haute qualité conçu exclusivement pour le ré-éclairage de portrait en profondeur. Cet ensemble de données sera utilisé comme données d'entraînement pour le système.

À l'aide de l'ensemble de données fourni, une architecture de réseau en sablier (Hourglass network) [15] est formée selon la méthode proposée. Ce réseau prend en entrée une image de portrait source et une condition d'éclairage cible et produit une image de portrait rallumée. Le réseau en sablier est un modèle bien connu qui est souvent utilisé pour la segmentation d'images et l'estimation de la pose. Dans le chapitre suivant nous allons expliquer la mise en œuvre de l'approche proposée, ainsi que la présentation de quelques résultats. newpage

Chapitre 4

Mise en œuvre et résultats

4.1 Introduction

Dans ce chapitre nous allons décrire les objectifs des différentes étapes de réalisation de notre application, pour cela nous allons commencer par décrire la manière avec laquelle nous avons créé la base de données DPR que nous avons utilisée pour l'entraînement des réseaux pour le rééclairage des photos de portrait, ensuite nous présentons la mise en œuvre du réseau Hourglass et de son entraînement.

La deuxième partie de ce chapitre sera consacrée à définir le processus de réalisation de notre application, pour cela nous commencerons par décrire les outils (matériels et logiciels) ainsi que les plateformes que nous avons utilisées.

Nous terminons ce chapitre par la présentation des résultats, ainsi qu'une analyse quantitative et qualitative sur les images produites par notre application.

4.2 Préparation de base de données (DPR)

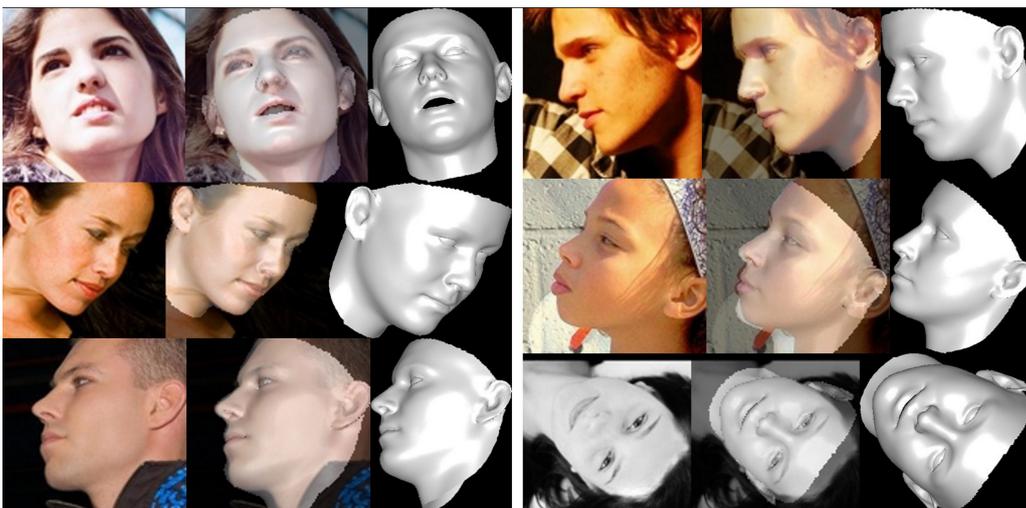
Pour créer l'ensemble de données DPR spécifiquement pour le rééclairage des portraits, on utilise l'ensemble de données CelebA-HQ qui est une version étendue de l'ensemble de données CelebA et contient 30 000 photos de visages à une résolution de 1024x1024 pixels comme base. Cependant, certaines images sont exclues de l'ensemble de données DPR si le détecteur de points de repère ne parvient pas à les détecter. Cette

étape permet de s'assurer que l'ensemble de données contient des images dont les repères faciaux sont fiables et détectés avec précision, ce qui est essentiel pour le processus de ré-éclairage. En conséquence, l'ensemble de données DPR se compose de 27 627 images après suppression de celles dont les points de repère n'ont pas été détectés.

L'approche globale présentée combine l'astuce de l'image de rapport avec l'estimation de la normale et les techniques d'affinement pour générer l'ensemble de donnée DPR . En manipulant les composantes de réflectance, de normalité et d'éclairage d'une image de visage.

Voila un sous ensemble de DPR générer avec cette méthode .

utilisent la méthode 3DDFA pour estimer les normales des images de visage restantes. La méthode 3DDFA ajuste un modèle 3D morphable (3DMM) aux images faciales et produit des paramètres de forme qui peuvent être utilisés pour générer des images normales de portrait à des résolutions arbitraires. Il est probable que 3DDFA soit appliqué aux 27 627 images de visage pour estimer les normales correspondantes. les normales estimées peuvent ne pas s'aligner avec précision sur les images de portrait en raison des limites de la modélisation des variations de la géométrie du visage avec le 3DMM, donc on propose un algorithme d'affinement des normales basé sur la méthode "As-Rigid-As-Possible" (ARAP). Cet algorithme aligne les normales estimées sur les images de portrait afin d'éviter les artefacts dans les images relues qui en résultent.



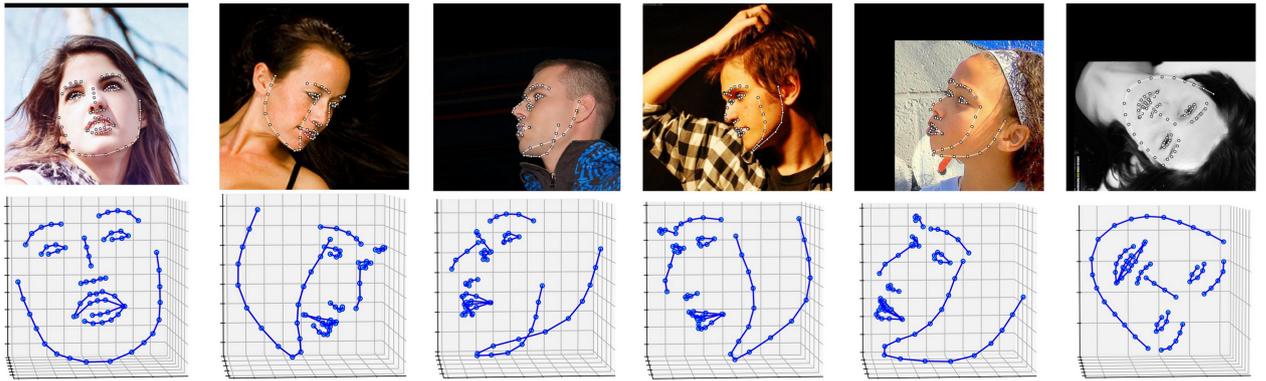


FIGURE 4.1 – Exemples d’application 3DFFA

Ensuite, pour chacune des 27 627 images de l’ensemble de données DPR en choisissant cinq conditions d’éclairage au hasard à partir d’un autre ensemble de données d’éclairage antérieur. Cet ensemble de données contient probablement une collection de configurations ou de styles d’éclairage prédéfinis. Ces paramètres d’éclairage sont appliqués aux images originales pour créer un total de 138 135 images de visages éclairés (27 627 images x 5 paramètres d’éclairage) voir la figure 4.1 qui représente un sous ensemble de cette DPR générer par cette méthode . Ce spectre de conditions d’éclairage garantit que l’ensemble de données DPR comprend une variété de scénarios de ré éclairage permettant une formation et une évaluation approfondies du modèle d’apprentissage profond pour le ré éclairage de portraits.

La première chose qu’ils font est de créer un produit élément par élément de la réflectance (R), de la normale (N) et de l’éclairage (L) pour représenter une image du visage, désignée par I . Ils basent leur expression de l’image du visage sur $I = R * f(N, L)$, où f représente la fonction d’ombrage de Lambertian, sur le modèle de réflectance de Lambertian pour les visages humains.

La disponibilité de l’ensemble de données DPR, avec sa grande échelle et sa haute résolution permet de développer et tester des techniques d’apprentissage profond liées à l’éclairage de portraits à image unique. Il permet d’explorer et de faire progresser les algorithmes et les méthodes de ré éclairage en offrant diverses images faciales et les variations de ré éclairage correspondantes à des fins de formation et d’évaluation.

4.3 Aperçu de la mise en œuvre

La classe suivante définit l'architecture du réseau Hourglass proposée.

```

class HourglassNet_1024(nn.Module):
    """
    basic idea: low layers are shared, upper layers are different
    lighting should be estimated from the inner most layer
    NOTE: we split the bottle neck layer into albedo, normal and lighting
    """
    def __init__(self, model_512, baseFilter = 16, gray=True):
        super(HourglassNet_1024, self).__init__()
        self.model_512 = model_512
        self.ncLight = 27 # number of channels for input to lighting network
        self.baseFilter = baseFilter

        self.ncPre = self.baseFilter # number of channels for pre-convolution

        self.pre_conv = nn.Conv2d(1, self.ncPre, kernel_size=5, stride=1, padding=2)
        self.pre_bn = nn.BatchNorm2d(self.ncPre)
        self.downSample = nn.MaxPool2d(kernel_size=2, stride=2)

        self.upSample = nn.Upsample(scale_factor=2, mode='nearest')

        self.conv_1 = nn.Conv2d(self.ncPre, self.ncPre, kernel_size=3, stride=1, padding=1)
        self.bn_1 = nn.BatchNorm2d(self.ncPre)
        self.conv_2 = nn.Conv2d(self.ncPre, self.ncPre, kernel_size=1, stride=1, padding=0)
        self.bn_2 = nn.BatchNorm2d(self.ncPre)
        self.conv_3 = nn.Conv2d(self.ncPre, self.ncPre, kernel_size=1, stride=1, padding=0)
        self.bn_3 = nn.BatchNorm2d(self.ncPre)

        self.output = nn.Conv2d(self.ncPre, 1, kernel_size=1, stride=1, padding=0)

```

La classe est initialisée avec deux paramètres : baseFilter et gray.

baseFilter détermine le nombre de filtres de base utilisés dans le réseau, et gray est un booléen indiquant si les images d'entrée sont en niveaux de gris ou en couleur. le nombre de canaux de sortie du réseau d'éclairage est fixé à 9 (entrée en niveaux de gris) ou à 27 (entrée en couleur).

Avant l'image d'entrée passe dans tous les quatre blocs de réseau Hourglass(HG0,HG1,HG2,HG3), premièrement elle passe par une couche de pré convolution 2D avec 1 canal d'entrée, 16 canaux de sortie ,un filtre 5x5 , un pas de 1 et d'un remplissage de 2 pour facilite l'extraction d'informations de l'image portrait et génère une carte de caractéristiques .ensuite on applique une couche de normalisation par lot pour normalise les activations avec une moyenne et un écart-type calculés sur le mini lot et les renvoie en tant que sortie normalisée afin de accélérer la convergence du modèle pendant

l'entraînement.

Après, on appliquons l'opération de sous-échantillonnage en utilisant le pooling max sur la carte des caractéristique 2D avec une taille de noyau de 2 et un pas de 2 pour réduit les dimensions spatiales de la carte . ensuite on appliquons l'opération de sur-échantillonnage en utilisant l'interpolation du plus proche voisin avec le facteur d'échelle=2 pour doubler les dimensions spatiales de la sortie(la carte des caractéristique) de la précédent opération (sous-échantillonnage) .

	h1	h2	h3	h4	h5	h6	h7	h8	s1	s2	s3	S4	c1	C2-1	C2-2	C2-3	C2-0
Nombre du canal d'entrée	16	16	32	64	15	64	32	6	64	32	16	16	1	1	16	16	16
Nombre du canal de sortie	16	32	64	155	64	32	16	16	64	32	16	16	16	16	16	16	1
Taille de filtre	3	3	3	3	3	3	3	3	3	3	3	3	5	3	1	1	1

FIGURE 4.2 – Détails sur chaque bloc de réseau Hourglass.[28]

4.4 Entraînement de Hourglass

Après la génération le DPR ,on utilise l'ensemble de données Multi-PIE pour une évaluation quantitative de le modèle proposée . L'ensemble de données Multi-PIE comprend des photos de la même personne sous différents éclairages qui peuvent être utilisés comme paire d'images source et cible.Chaque photographie Multi-PIE est éclairée par une source lumineuse ponctuelle dominante, alors que la plupart des images de portraits "dans la nature" ont un éclairage diffus.

Nous produisons des photos avec 7 paramètres d'éclairage en faisant la moyenne de 3 à 4 images de visage originales provenant de Multi-PIE, ce qui donne des images avec un éclairage plus réaliste et diffus.

Il y a quatorze époques de formation en tout pour le réseau. Un passage complet sur l'ensemble de l'ensemble des données de formation est appelé une époque. La valeur du paramètre est fixée à 0,5. Le réseau est initialement formé pendant cinq époques sans aucune connexion sautée. La fonction de saut de connexions d'un réseau neuronal

permet aux données de passer par des couches spécifiques, ce qui améliore le flux de gradient pendant la formation. Les connexions sautées sont progressivement introduites après les cinq premières époques, une à la fois, jusqu'à ce que toutes les couches sautées soient présentes.

4.5 Dispositif expérimental

L'évaluation des performances et de l'efficacité de l'application proposée repose en grande partie sur le dispositif expérimental. Il s'agit de spécifier les configurations matérielles et logicielles utilisées pour les tests, l'entraînement et la validation de l'application.

4.5.1 Configuration du matérielle

Les ressources informatiques utilisées pour faire fonctionner l'application logicielle mise en œuvre sont incluses dans la configuration matérielle. Elle contient des informations sur les spécifications du système informatique, notamment : **Un CPU multi-cœur** puissant est utilisé pour le prétraitement des données, la construction du réseau Hourglass et la gestion de l'entraînement dans l'architecture matérielle proposée, **Les GPU** à grande capacité de mémoire pour accélérer les procédures de formation afin d'exploiter leurs capacités de traitement parallèle en temps réel, aussi on utilise **une mémoire (RAM)** suffisante pour stocker et traiter les données d'apprentissage, les représentations intermédiaires et les paramètres du réseau.

4.5.2 Configuration du logicielle

Pour l'implémentation du réseau dans la configuration logicielle pour le relighting des images de portraits à l'aide du réseau Hourglass, un cadre d'apprentissage profond tel que TensorFlow, PyTorch, et Keras est utilisé.

Pour la préparation, des bibliothèques (packages) de traitement d'image tels que OpenCV est utilisée.

La bibliothèque SHTOOLS est utilisée pour les transformations harmoniques sphé-

riques, et des utilitaires tels que "utils-normal.py" aident à la coordination de la conversion du système.

4.5.2.1 PyTorch

L'utilisation de PyTorch 1.0.0 aide à construire et entraîner facilement le modèle Hourglass pour le ré éclairage des portraits. Elle fournit une interface de haut niveau pour concevoir l'architecture du réseau, contrôler le chargement et le prétraitement des données et améliorer les paramètres du modèle grâce à des méthodes d'optimisation basées sur le gradient.

Les couches convolutives, les fonctions d'activation, les fonctions de perte et les algorithmes d'optimisation ne sont que quelques-unes des caractéristiques et fonctions intégrées que PyTorch fournit pour les applications typiques d'apprentissage profond.



4.5.2.2 TensorFlow

Dans le contexte de la mise en œuvre du réseau Hourglass pour le rééclairage de portraits, TensorFlow offre une plateforme flexible et efficace pour la construction d'architectures de réseaux neuronaux. Il permet aux développeurs de définir et d'entraîner des modèles complexes grâce à son API de haut niveau, qui simplifie le processus de construction et de configuration des couches du réseau, le traitement des données et l'optimisation des paramètres du modèle.



4.5.2.3 Keras

En offrant des fonctionnalités pratiques pour la préparation des données, le traitement par lots et la validation des modèles, Keras rationalise encore davantage la procédure de formation. Grâce à keras on peut rapidement pré traiter les données d'entrée, les diviser en ensembles de formation et de validation et de test et former le réseau Hourglass.



4.5.2.4 OpenCV

On utilise OpenCV 4.0.0 pour manipuler les images comprend le chargement, le redimensionnement et le recadrage des images, ainsi que la conversion des espaces colorimétriques. Ces fonctions sont essentielles pour préparer les photos avant de les introduire dans le réseau Hourglass.

En outre, OpenCV offre une variété de méthodes de filtrage et d'amélioration des images, notamment le débruitage, l'accentuation et le lissage. Pour améliorer la qualité et l'attrait esthétique des photos de portraits éclairés, ces techniques peuvent être appliquées à la sortie du réseau Hourglass.

4.5.2.5 Shtools

La représentation des conditions d'éclairage sous forme d'harmoniques sphériques est gérée à l'aide de SHTOOLS. Parce qu'elles offrent un moyen condensé et efficace de communiquer les informations d'éclairage, les harmoniques sphériques sont fréquemment employées pour représenter l'environnement lumineux dans les applications de relighting.



4.6 Mesures d'évaluation

Dans cette proposition deux mesures pour évaluer la qualité de l'image relue générée par rapport à l'image de référence. Ces mesures sont conçues pour tenir compte de l'ambiguïté inhérente aux conditions d'éclairage et pour évaluer la cohérence de l'image générée avec l'éclairage cible.

On utilise une erreur quadratique moyenne invariante **Si-MSE** (scale invariant Mean Squared Error) pour mesurer l'erreur entre l'image générée et l'image de référence.

$$\text{Si-MSE} = \frac{1}{N_I} \min_{\alpha} (\mathbf{I}_t - \alpha * \mathbf{I}_t^*)^2,$$

Pour évaluer quantitativement à quel point l'image générée représente les conditions d'éclairage souhaitées et pour extraire des informations sur l'éclairage des images créées I_t et des images cibles $I*_t$ en utilise le réseau de neurone profond **SfsNet**.

Puis ,on calculons la distance L2 invariante à l'échelle (**Si-L2**) pour mesurer la différence entre les informations d'éclairage extraites de l'image produite (L_t) et de l'image cible ($L*_t$). La distance Si-L2 est une métrique qui quantifie la similarité de deux vecteurs tout en étant insensible à l'échelle. Elle est déterminée en calculant la norme L2 de la différence entre les deux vecteurs et en la normalisant par les amplitudes de leurs moyennes.

L'efficacité de la perte de GAN et de la perte de caractéristiques est illustrée par les résultats quantitatifs et qualitatifs de notre réseau formé avec différentes configurations : LI, LI + IGAN, et LI + LGAN + Lf (modèle complet). Le tableau 1 et la figure 8 décrivent ces résultats.

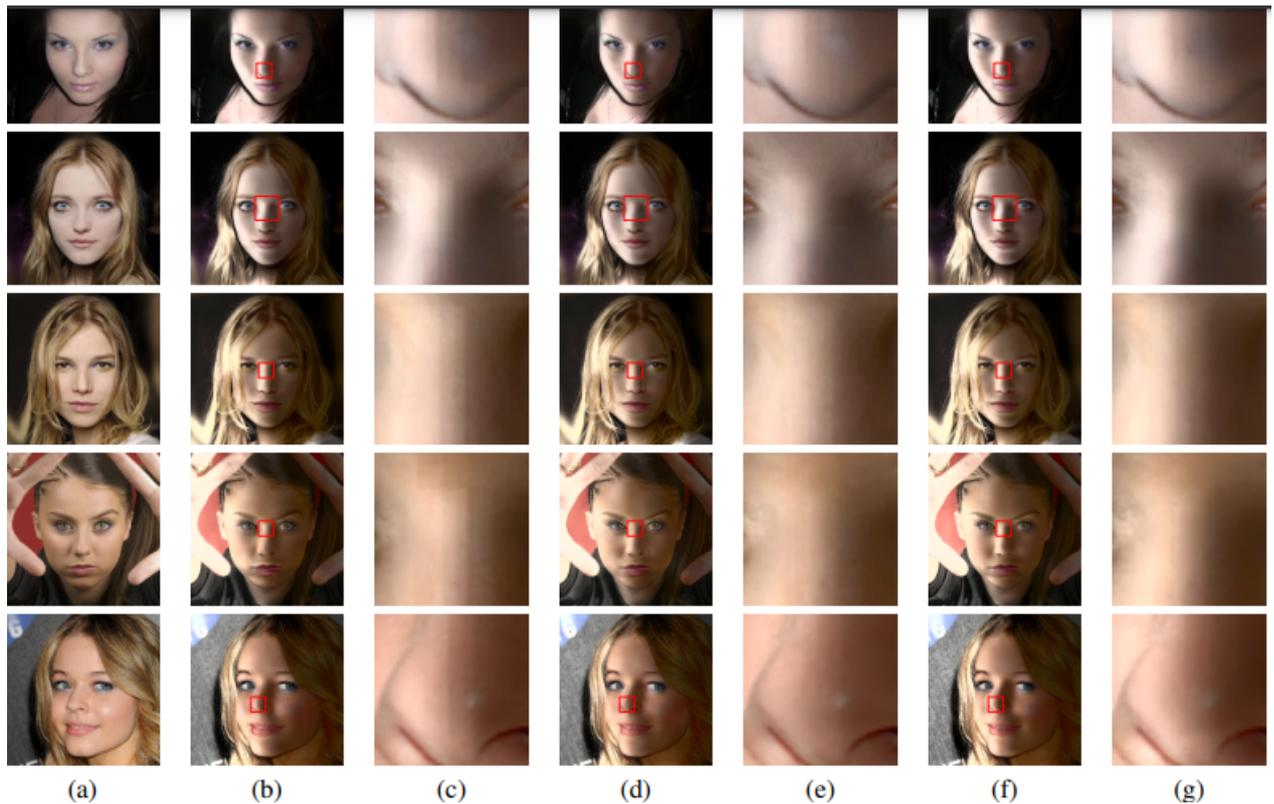


FIGURE 4.3 – Résultats comparatives d'évaluation

(a) l'image d'entrée, (b), (d) et (f) sont des images générées en utilisant LI, LI + L_{GANet} LI + L_{GAN} + L_f . respectivement ; (c), (e) et (g) sont la région du rectangle rouge de (b), (d) et (f) respectivement.

Notez le bord au milieu du bruit généré à l'aide de LI.

Les résultats de l'étude d'ablation sur l'ensemble de données Multi-PIE sont présentés dans un tableau suivant :

TABLE 4.1 – Étude d'ablation sur l'ensemble de données Multi-PIE

	Si-MSE	Si- L_2
L_I	0.00504	0.1307
$L_I + L_{GAN}$	0.00658	0.1686
$L_I + L_{GAN} + L_f$	0.00590	0.1444

Les valeurs plus faibles pour Si-MSE indiquent une meilleure performance en termes de similarité structurelle et pour Une distance Si-L2 implique une meilleure précision

dans la représentation des conditions d'éclairage requises, tandis qu'une distance plus grande indique une plus grande divergence par rapport aux conditions d'éclairage souhaitées.

Voilà une comparaison entre la méthode proposée et les méthodes de l'état de l'art selon l'ensemble de donnée Si-MSE . le meilleur résultat donné par la méthode proposée

TABLE 4.2 – Ensemble de données multi-PIE d'évaluation

	Si-MSE	Si- L_2
Li et al.	0.01322	0.3939
Shih et al.	0.01513	0.3415
Shu et al	0.00659	0.1593
La méthode proposée	0.00590	0.1444

4.7 Résultats

4.7.1 Résultats sur des images frontales

l'image originale :



Nous testons les images frontales sur ce modèle (le réseau Hourglass) avec les sources de lumière son intensité 95 et son couleur blanc et on obtenons les résultats suivantes.

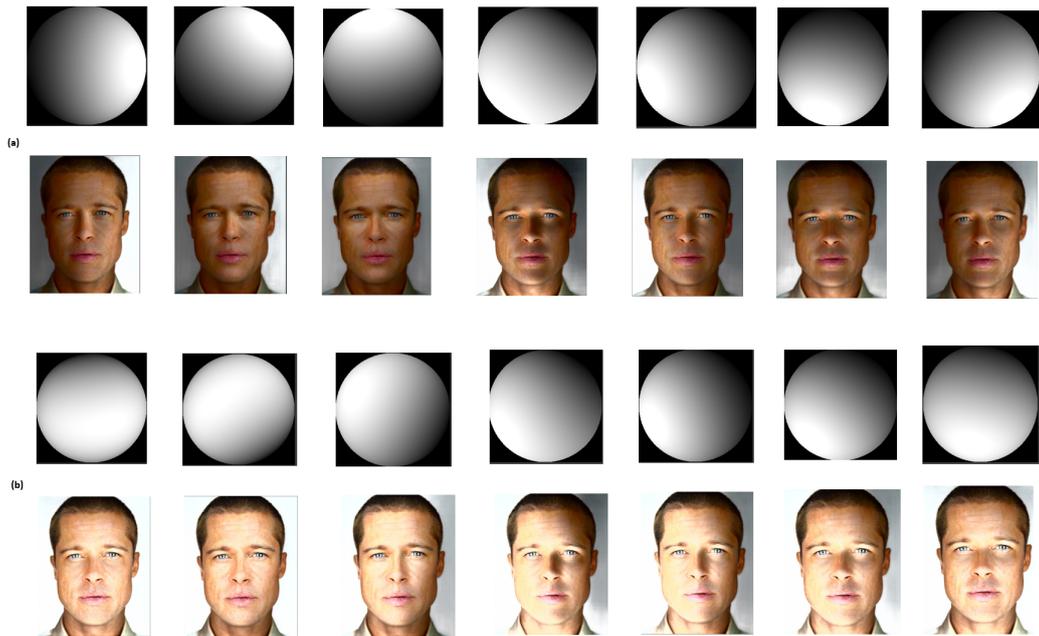
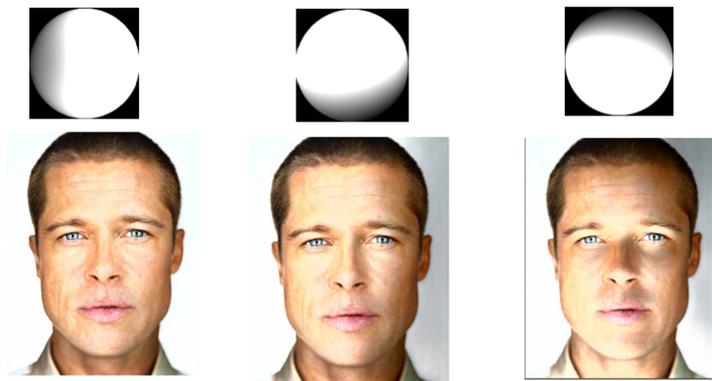


FIGURE 4.4 – Résultats sur des images frontales

(a) représente les résultats de l'image frontale avec une seule source de lumière.
 (b) (a) représente les résultats de l'image frontale avec une deux sources de lumière,
 la première sources est déplacer chaque fois et la deuxième source est stable .

Ensuite nous changeons l'intensité de la lumière en 10 et voila les résultats.



Nous démontrons que cette application donne des résultats réalistes et acceptables pour les images frontales.

4.7.2 Résultats sur des images non frontales

La méthode proposée produit de bons résultats lorsqu'elle traite des postures faciales non frontales. Cela signifie que le système peut évaluer et simuler correctement les conditions d'éclairage même lorsque le visage n'est pas dans une posture frontale conventionnelle. Pour produire des résultats esthétiques, l'approche prend très probablement en considération les différences d'angles et d'intensités d'éclairage générées par les positions non frontales.

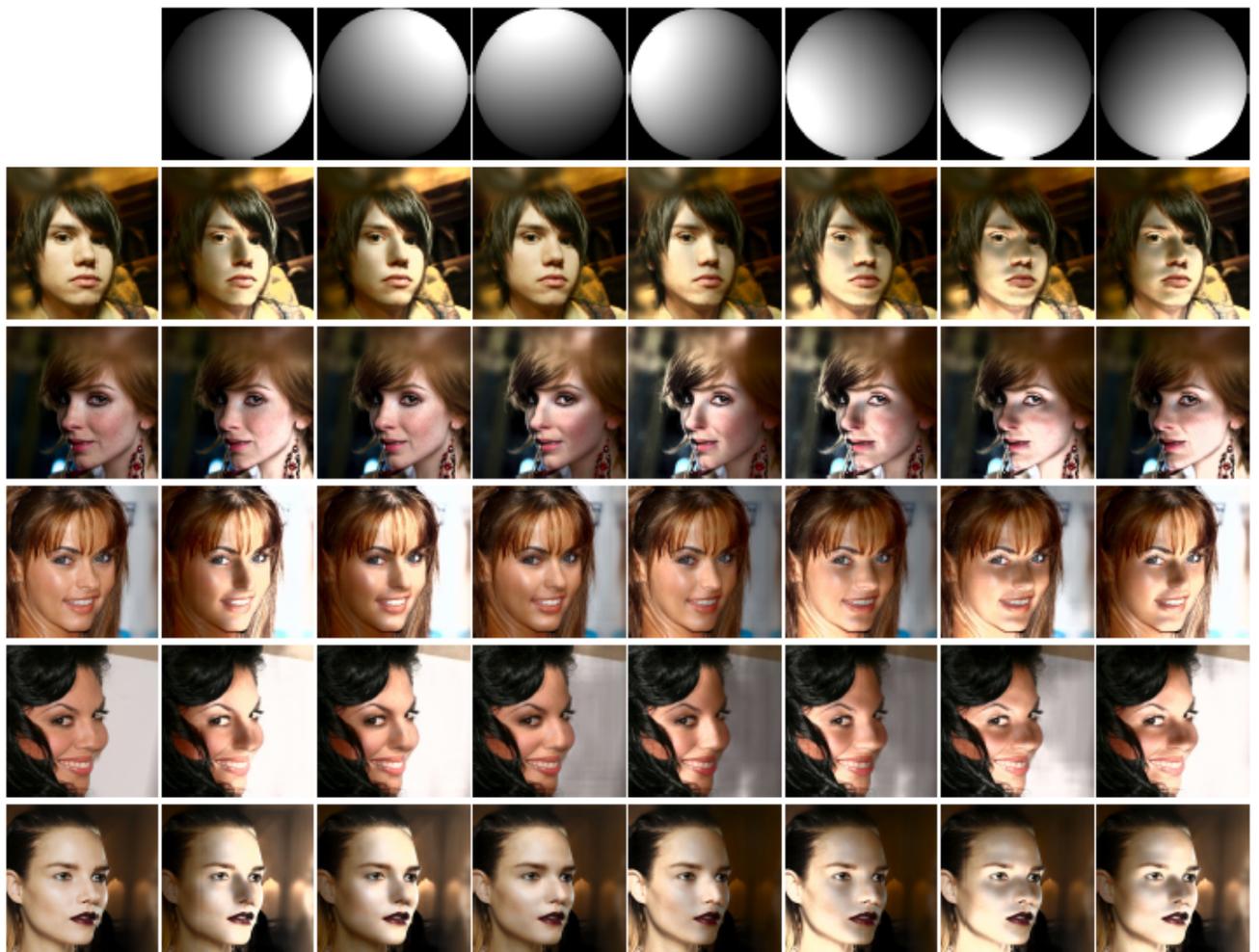


FIGURE 4.5 – Résultats sur des images non frontales

4.7.3 Résultats sur des images difficiles

La figure 4.6 montre comment l'approche a fonctionné sur des photos difficiles. L'expression "images difficiles" fait généralement référence à des photos présentant des caractéristiques

téristiques complexes, telles que des occlusions (parties du visage masquées ou cachées) ou des cosmétiques sur le visage. Selon le passage, l’algorithme proposé résout ces problèmes de manière adéquate et donne de bons résultats. Cela montre que le système peut prendre en compte les occlusions et les différences d’aspect de la peau causées par les cosmétiques, ce qui lui permet de fournir des représentations lumineuses appropriées en dépit de ces difficultés.

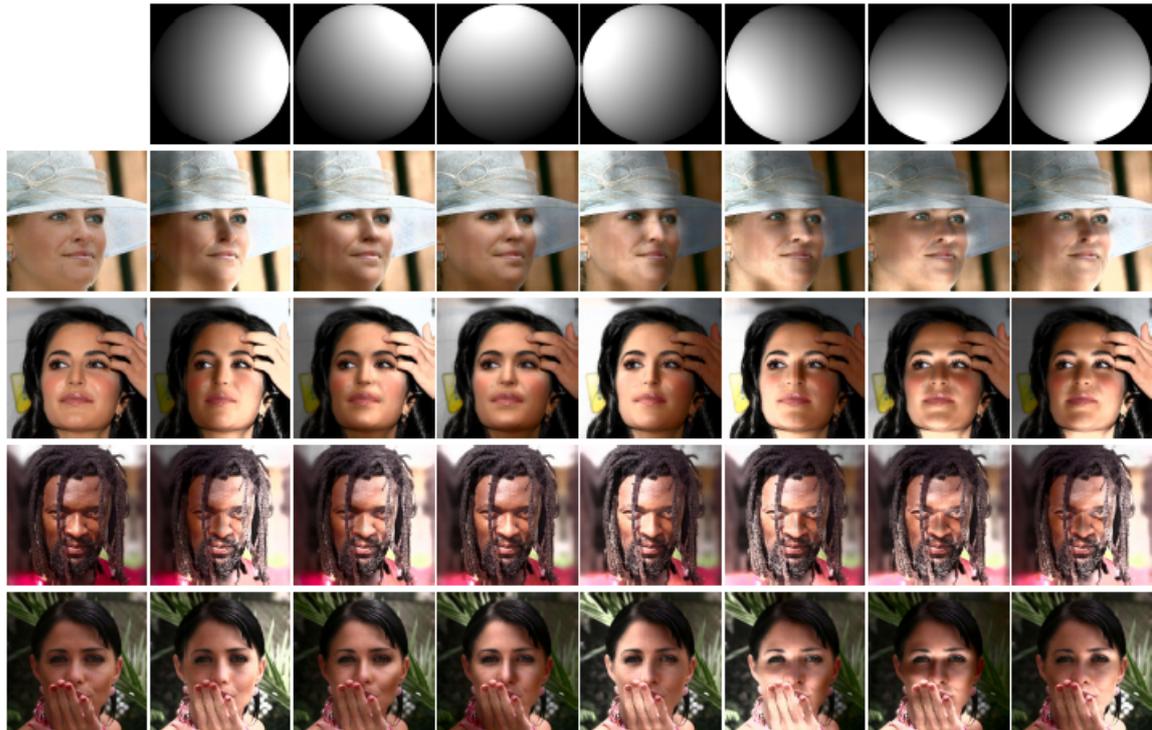


FIGURE 4.6 – Résultats sur des images difficiles.

La première ligne de la figure montre l’éclairage SH cible, La première colonne des autres lignes montre l’image d’entrée, les autres colonnes montrent les images éclairées par la méthode proposée sous l’éclairage SH cible.

Cette méthode peut traiter les visages avec occlusions.

4.7.4 Résultats de données DPR de haute résolution

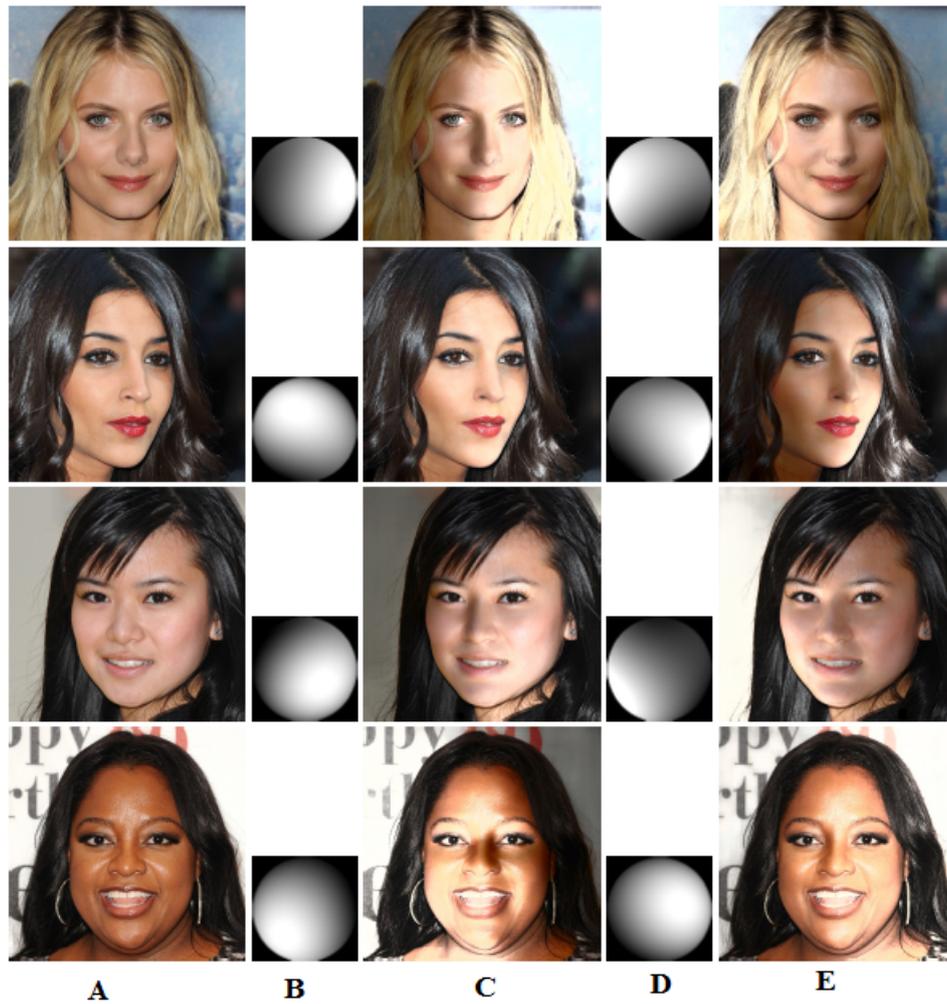


FIGURE 4.7 – Résultats sur l’ensemble de données DPR.

(A) est l’image d’entrée, (B) et (D) sont l’éclairage SH cible, (C) et (E) sont les images éclairées par la méthode proposée.

4.7.5 Résultats de données de portraits de Flickr

Nous testons le modèle formé sur l’ensemble de données de portraits de Flickr et il donne des résultats acceptables et réalistes.

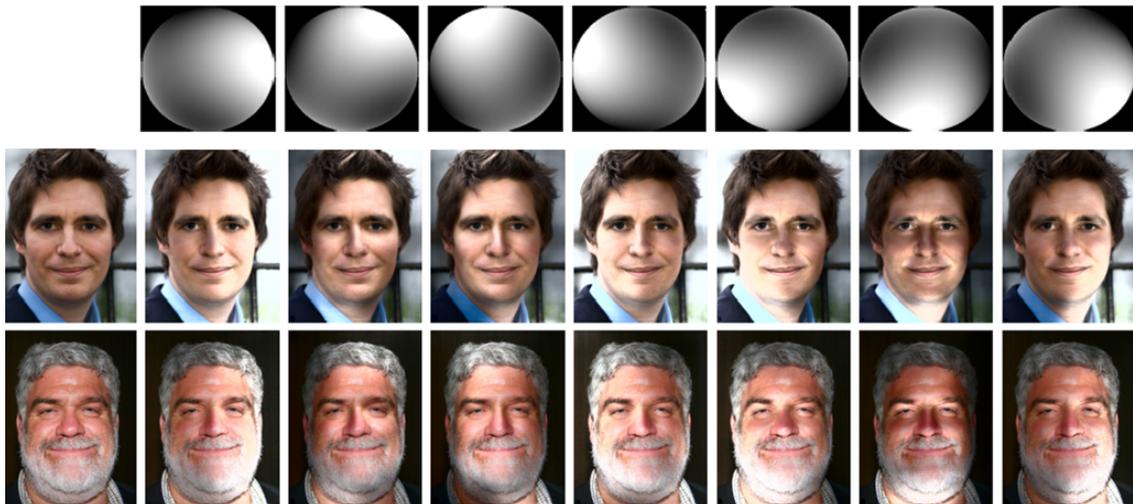


FIGURE 4.8 – Résultats sur l’ensemble de données de portraits de Flickr.

4.7.6 Mauvais Résultats

Comme nous représentons l’éclairage à l’aide des harmoniques sphériques, l’approche ne peut pas modéliser les ombres portées. Cela nécessiterait une représentation de l’éclairage utilisant le traçage des rayons.

la technique proposée est que les résultats obtenus sont affectés par les ombres dans les photographies de portraits avec des ombres fortes, comme illustré dans la deuxième ligne de la figure 4.9. on observe que les ombres portées causées par les lunettes ne changent pas lorsque les conditions d’éclairage varient. Cette lacune indique que la méthode, probablement basée sur les harmoniques sphériques, ne tient pas compte de la nature dynamique des ombres portées.

Dans le monde réel, les sources lumineuses peuvent changer de position et d’intensité, ce qui entraîne des variations des ombres portées. Si ces changements ne sont pas représentés avec précision, les rendus risquent d’être moins réalistes.

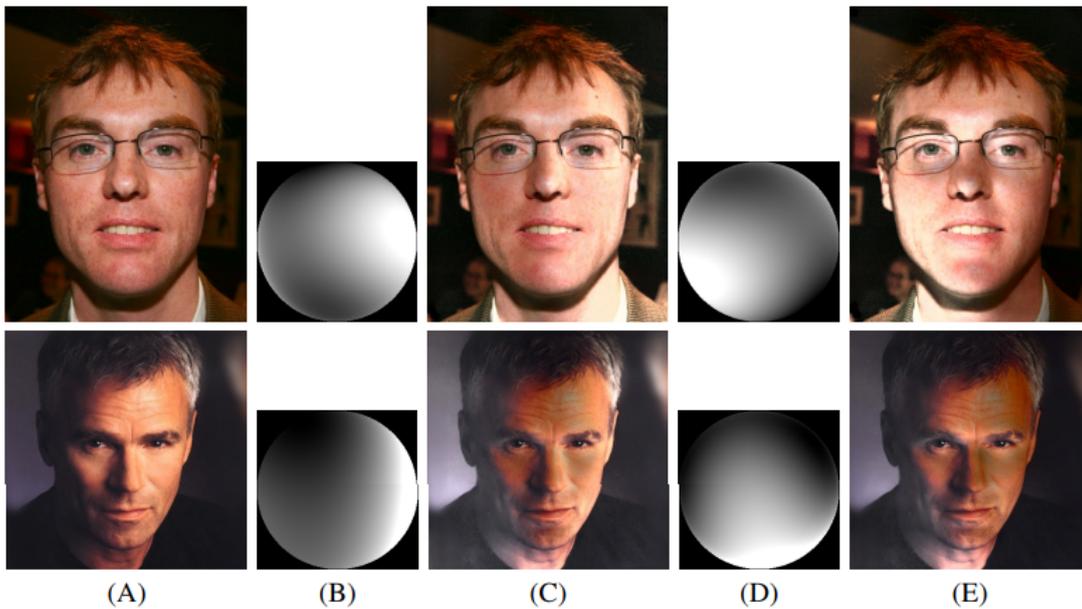


FIGURE 4.9 – Mauvais résultats.

(A) est l'image d'entrée, (B) et (D) sont l'éclairage de la cible SH, (C) et (E) sont des images éclairées par la méthode proposée.

4.8 Conclusion

Dans ce chapitre nous nous sommes concentré sur la mise en œuvre et les résultats du ré éclairage des photos de portraits à l'aide d'un réseau Hourglass à partir d'une seule image d'entrée. Pour cela on utilise l'apprentissage profond avec une fonction de perte pour calculer la différence entre l'entrée originale et la sortie reconstruite.

Le chapitre commence par une discussion sur la préparation de la base de données DPR, qui est divisée en trois sections : test, entraînement et validation. On utilise ensuite cette base pour faire la phase de test. Cependant, comme il n'y a pas de véritable vérité terrain (Ground truth) accessible, on ne peut pas utiliser cet ensemble de données pour évaluer l'exactitude des photos ré éclairées, alors nous utilisons l'ensemble de données Multi-PIE pour l'évaluation quantitative afin de résoudre cette restriction.

Pour les résultats obtenus, nous voyons que cette application donne des résultats acceptables pour les images frontal , non frontal et les images de l'ensemble de données de portraits de Flickr. newpage

Chapitre 5

Conclusion générale

Le rendu neuronal est un domaine innovant à l'intersection de l'infographie et de l'apprentissage profond. Il vise à générer des images très réalistes et détaillées en exploitant la puissance des réseaux neuronaux. À la différence des méthodes de rendu traditionnelles qui reposent sur des modèles mathématiques et des algorithmes complexes, le rendu neuronal utilise des réseaux neuronaux pour apprendre directement et approximativement la correspondance complexe entre les informations de la scène et les images rendues.

Dans ce mémoire nous nous sommes intéressés à résoudre le problème de la modification des conditions d'éclairage d'un portrait sans utiliser de nombreuses images ou des configurations compliquées. Les limites mentionnées précédemment nous ont motivées à utiliser une nouvelle approche basée sur l'apprentissage profond à base de réseaux neuronaux convolutifs (CNN).

Notre principal objectif étant de produire des portraits éclairés très réalistes qui améliorent l'expérience visuelle du spectateur en maîtrisant la capture de détails infimes et d'effets d'éclairage complexes. Nous utilisons les résultats de [28] qui appliquent dans une première phase une méthode de ré-éclairage de portrait basée sur la physique pour générer un ensemble de données de ré-éclairage de portrait à grande échelle, de haute qualité et "dans la nature" (DPR). Ensuite nous avons entraîné un réseau Hourglass à l'aide de l'ensemble de données DPR obtenu dans la phase précédente pour la tâche de ré-éclairage des portraits. Celui-ci prend en entrée une image source et un éclairage cible et

génère l'image cible. Il prédit également l'éclairage SH pour l'image source en utilisant les caractéristiques de la couche goulot d'étranglement pour démêler les informations d'éclairage de l'image source.

L'étude effectuée dans ce mémoire nous a permis de traiter un problème de rendu graphique, en utilisant des techniques d'apprentissage, nous pouvons considérer que notre application est un point de départ pour l'exploration et l'intégration des CNN (divers types) dans le rendu. Plusieurs extensions sont à prévoir, comme :

1. Diversité des ensembles de données : Pour améliorer la généralisation dans le rééclairage de portraits, il est essentiel de les entraîner sur des ensembles de données diversifiés qui englobent un large éventail de conditions d'éclairage, de sujets et de poses. Ceci permettra d'améliorer la capacité du réseau à gérer différents scénarios de rééclairage et à améliorer ses performances globales.
2. Rééclairage en temps réel : Le rééclairage en temps réel est un objectif nécessaire dans les applications interactives comme la vidéoconférence, la réalité virtuelle et la réalité augmentée. La solution consiste à optimiser l'architecture du Hourglass et son processus d'apprentissage.
3. Traitement des effets d'éclairage complexes : Les portraits contiennent souvent des effets d'éclairage complexes, tels que la diffusion sous la surface, les ombres portées et les détails à haute fréquence. L'amélioration des Hourglass pour capturer et modéliser efficacement ces phénomènes d'éclairage complexes peut conduire à des portraits éclairés plus réalistes.
4. Séparer les facteurs d'éclairage : La séparation des effets de l'éclairage des autres attributs de l'image, tels que la pose, l'expression et l'identité, est une tâche difficile dans le rééclairage des portraits, ce qui permet d'obtenir des effets de rééclairage plus précis et mieux contrôlés.
5. Contrôle de l'utilisateur et interactivité : Il est important de donner aux utilisateurs des commandes et des interfaces intuitives pour qu'ils puissent régler de manière interactive les paramètres d'éclairage et explorer différentes options de rééclairage. La conception d'interfaces conviviales permettant de manipuler en

temps réel les propriétés de l'éclairage, telles que la direction, l'intensité, la couleur et les effets ambiants, peut améliorer l'expérience de l'utilisateur et favoriser l'exploration créative.

6. Évaluation de la qualité perceptuelle : Il est essentiel de mettre au point des mesures et des méthodes d'évaluation robustes pour évaluer la qualité perceptive des images relit. Les mesures d'évaluation traditionnelles telles que PSNR et SSIM peuvent ne pas rendre pleinement compte de la qualité visuelle subjective perçue par les humains.

Bibliographie

- [1] Comprendre les réseaux de neurones gan. http://studiomaven.org/Context_Overview_of_Ray-Traced_Rendering.html, . 2023.
- [2] Comprendre les réseaux de neurones gan. <https://www.kaggle.com/code/theblackmamba31/generating-fake-faces-using-gan>, . 2023.
- [3] Nadeem Akhtar and U. Ragavendran. Interpretation of intelligence in cnn-pooling processes : a methodological survey. *Neural Computing and Applications*, 32, 02 2020. doi : 10.1007/s00521-019-04296-5.
- [4] Oswald Aldrian and William AP Smith. Inverse rendering of faces with a 3d morphable model. *IEEE transactions on pattern analysis and machine intelligence*, 35 (5) :1080–1093, 2012.
- [5] Eliel Arnold. Comprendre les réseaux de neurones gan. <https://www.arnold.fr/leclairage-dans-les-bureaux>. 2023.
- [6] Ali Azawii, Sufyan Al-Janabi, and Belal Al-Khateeb. Survey on intrusion detection systems based on deep learning. *Periodicals of Engineering and Natural Sciences (PEN)*, 7 :1074–1095, 09 2019. doi : 10.21533/pen.v7i3.635.
- [7] Quentin Bammey. Experiments on deep single-image portrait relighting. *Image Processing On Line*, 12 :420–434, 2022.
- [8] Jeremy. Birn. *Digital Lighting & Rendering*. Voices That Matter. New Riders, 3rd edition, 2014. ISBN 9780321928986. URL <https://books.google.dz/books?id=pevrAQAAQBAJ>.

- [9] Youcef Djeriri. *Les Réseaux de Neurones Artificiels*. 09 2017.
- [10] Youcef Djeriri. *Les Réseaux de Neurones Artificiels*. Université Djillali Liabès de Sidi Bel-Abbès, 09 2017.
- [11] Clément Douarre and Rousseau David. Introduction à l'apprentissage profond (deep learning) de l'intelligence artificielle. <https://culturesciencesphysique.ens-lyon.fr/ressource/IA-apprentissage-Rousseau.xml>, 2021. Consulté le 10 juin 2023.
- [12] Jeff Heaton. Ian goodfellow, yoshua bengio, and aaron courville : Deep learning : The mit press, 2016, 800 pp, isbn : 0262035618. *Genetic Programming and Evolvable Machines*, 19(1-2) :305–307, 2018.
- [13] Talbi Ilyes. Comprendre les réseaux de neurones gan. <https://larevueia.fr/comprendre-les-reseaux-de-neurones-gan/>, 2023. Consulté le 10 juin 2023.
- [14] Peng Liu, Wai Lok Woo, and S.s Dlay. One colored image based 2.5d human face reconstruction. 01 2009.
- [15] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *Computer Vision–ECCV 2016 : 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pages 483–499. Springer, 2016.
- [16] Duc-Phong Nguyen, Tan-Nhu Nguyen, Stéphanie Dakpé, Marie-Christine Ho Ba Tho, and Tien-Tuan Dao. Fast 3d face reconstruction from a single image using different deep learning approaches for facial palsy patients. *Bioengineering*, 9(11) :619, 2022.
- [17] Romain Pacanowski. *Modes de représentation pour l'éclairage en synthèse d'images*. Ph.d. thesis, labri, Université Bordeaux 1 ; and Département d'Informatique et Recherche Opérationnelle, Université de Montréal, September 2009.

- [18] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. Post-production facial performance relighting using reflectance transfer. *ACM Transactions on Graphics (TOG)*, 26(3) :52–es, 2007.
- [19] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering : From Theory to Implementation*. Elsevier Science, 2016. ISBN 9780128007099. URL <https://books.google.dz/books?id=iNMVBQAAQBAJ>.
- [20] Amnon Shashua and Tammy Riklin-Raviv. The quotient image : Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2) :129–139, 2001.
- [21] François X Sillion and Claude Puech. *Radiosity and global illumination*, volume 11. Morgan Kaufmann Publishers, 1994.
- [22] Mohammad Taye. Theoretical understanding of convolutional neural network : Concepts, architectures, applications, future directions. *Computation*, 11 :52, 03 2023. doi : 10.3390/computation11030052.
- [23] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. *Computer Graphics Forum*, 41(2) :703–735, 2022.
- [24] Mateusz Trokielewicz and Adam Czajka. Data-driven segmentation of post-mortem iris images. pages 1–7, 06 2018. doi : 10.1109/IWBF.2018.8401558.
- [25] Zhen Wen, Zicheng Liu, and Thomas S Huang. Face relighting with radiance environment maps. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–158. IEEE, 2003.
- [26] ABDENNOUR YAHIAOUI. Système de discrimination visages/faux visages par réseaux de neurones convolutifs (cnn). 2021.

- [27] Huo Yingge, Imran Ali, and Kang-Yoon Lee. Deep neural networks on chip - a survey. pages 589–592, 02 2020. doi : 10.1109/BigComp48618.2020.00016.
- [28] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7194–7202, 2019.