Low-light image restoration with deep learning techniques

Arthur Lecert

COLLEGE DOCTORAL BRETAGNE
MATHS, TELECOMS
INFORMATIQUE, SIGNAL
SYSTEMES, ELECTRONIQUE

Université de Rennes

Inria

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE Nº 601
Mathématiques, Télécommunications, Informatique, Signal, Systèmes, Électronique
Spécialité : *Signal, Image, Vision*

Par
Arthur LECERT

Restauration d'images à faible luminosité à l'aide de méthodes d'apprentissage profond.

Thèse présentée et soutenue à INRIA Rennes - Bretagne Atlantique, le 20/12/2023
Unité de recherche : SIROCCO, INRIA

Rapporteurs avant soutenance :

Enrico MAGLI	Full professor, Politecnico di Torino
Thomas OBERLIN	Associate professor, ISAE-SUPAERO

Composition du Jury :

Présidente :	Laetitia CHAPEL	Full professor, Institut Agro Rennes-Angers
Rapporteurs :	Enrico MAGLI	Full professor, Politecnico di Torino
	Thomas OBERLIN	Associate professor, ISAE-SUPAERO
Examinateurs :	Laetitia CHAPEL	Full professor, Institut Agro Rennes-Angers
	Raoul de CHARETTE	Researcher, INRIA
Dir. de thèse :	Christine GUILLEMOT	Research Director, INRIA
Co-encadrante. de thèse :	Aline ROUMY	Research Director, INRIA

Invité(s) :

Renaud FRAISSE	Imaging Systems Senior Expert, Airbus Defence and Space

Résumé en français **i**
Contexte . . . i
La restauration d'images de nuit . . . ii
Plan détaillé . . . v

Introduction **1**
Application context . . . 1
The restoration of low-light images . . . 2
Identifying the degradations . . . 2
Outdoor images . . . 3
Preserving the integrity of the image . . . 4
Detailed outline . . . 4

Publications **7**

1 State-of-the-art on the restoration of low-light images **9**
1.1 Model-based algorithms . . . 10
1.1.1 Contrast enhancement . . . 10
1.1.2 LIME . . . 12
1.2 Deep learning approaches . . . 14
1.2.1 Supervised methods . . . 14
1.2.2 Unsupervised methods . . . 19
1.3 Metrics . . . 25
1.3.1 Reference-based metrics . . . 26
1.3.2 No-reference metrics . . . 29
1.4 Chapter conclusion . . . 33

2 Background on the original Retinex model **35**
2.1 Perceptual interpretation . . . 36
2.2 Physics of light . . . 39
2.2.1 Horn's interpretation . . . 39
2.2.2 Reversing the camera image processing pipeline . . . 41
2.2.3 Color constancy . . . 41
2.3 Chapter conclusion . . . 42

3 A new regularization for Retinex decomposition **45**

3.1 Initial problem statement . . . 46
3.2 The proposed prior . . . 47
3.2.1 The trivial solution and the exposure prior . . . 47
3.2.2 New problem formulation . . . 48
3.3 Experiments . . . 48
3.3.1 Restoration of the components . . . 48
3.3.2 Implementation details . . . 49
3.3.3 Quantitative comparison . . . 49
3.3.4 Qualitative results . . . 51
3.3.5 Ablation and hyperparameter study . . . 52
3.4 Chapter conclusion . . . 53

4 GAN architecture leveraging a Retinex model with colored illumination 57
4.1 Improvements to the Retinex model . . . 59
4.1.1 A colored illumination . . . 59
4.1.2 New priors . . . 61
4.2 The Architecture . . . 62
4.2.1 Architecture choices . . . 62
4.2.2 The loss terms . . . 64
4.2.3 Visualization and restoration of the components . . . 66
4.3 Results . . . 70
4.3.1 Metrics & Evaluation methodology . . . 70
4.3.2 Implementation details . . . 71
4.3.3 Ablation study . . . 73
4.3.4 Qualitative comparison . . . 73
4.3.5 Guarantees on a restoration without hallucination . . . 81
4.4 Chapter conclusion . . . 83

5 Retinex theory and solving Schrödinger Bridges with Diffusion models 85
5.1 Introduction to the Monge-Kantorovich problem . . . 87
5.2 Link between the Retinex theory and the Schrödinger Bridge problem . . . 88
5.2.1 Entropy-regularized optimal transport and the static Schrödinger Bridge . . . 88
5.2.2 Link to the Retinex theory . . . 89
5.2.3 The dynamic form of the Schrödinger bridge . . . 90

5.3 The Schrödinger Bridge with score-based generative models 91
5.3.1 The formulation of the problem 91
5.3.2 Procedures to reduce the bias in the solutions 95
5.3.3 Regularizations for the Schrödinger Bridge 97
5.4 Results 102
5.4.1 IMF 102
5.4.2 CUT Scheme 104
5.4.3 Adding the Retinex priors 104
5.5 Chapter conclusion 108

Conclusion **111**
Summary 111
Perspectives 112

Bibliography **115**

RÉSUMÉ EN FRANÇAIS

Contexte

Les techniques basées sur l'apprentissage automatique se répandent de nos jours dans les domaines de la vision par ordinateur et du traitement du signal depuis les travaux révolutionnaires impliquant des réseaux de neurones convolutifs. Ces derniers furent entraînés à l'aide de la combinaison d'algorithmes de rétropropagation du gradient et de descente de gradient stochastique. L'une de leurs premières applications fût la reconnaissance de chiffres et de caractères manuscrits (voir le réseau d'origine LeNet dans le papier [LeCun *et al.* 1989] par exemple). Ces algorithmes d'apprentissage profond se retrouvent aujourd'hui dans presque tous les défis, notamment lorsque de larges quantités de données peuvent être collectées. Un défi majeur pour ces applications reste leur utilisation dans des conditions de visibilité extrême. En effet, la capture d'images de nuit continue de poser un problème important pour la navigation de véhicules autonomes ou bien encore dans les applications liées à l'imagerie satellitaire.

Dans ce premier cas, ces systèmes de navigation reposent sur des traitements de haut niveau tels que la classification d'objets, mais aussi leur segmentation. Différentes modalités peuvent être employées pour leur fonctionnement (e.g. des caméras standards, des capteurs LIDAR, ...). En pratique, il arrive que des véhicules autonomes roulent de nuit et ne réussissent pas à identifier correctement leur environnement. Les algorithmes de traitement d'images ne sont dans ce cas pas conçus de manière assez robuste pour fonctionner lorsque l'acquisition est effectuée avec un nombre limité de photons. Récemment, de nombreuses villes réduisent l'éclairage public comme Toulouse ou Lyon dans un souci écologique et n'allument que les feux tricolores uniquement. Ces sources lumineuses col-

orées causent des problèmes identifiés comme *déviations de couleur* par la littérature qui s'ajoute au problème du rapport signal à bruit très faible.

La restauration d'images de nuit

Dans cette thèse, nous cherchons à définir de nouvelles méthodes basées sur l'apprentissage profond pour restaurer les images prises en extérieur à très faible luminosité et plus précisément des images nocturnes. Nous nous concentrons sur les images RVB et donc sur le spectre visible.

Le problème soulève naturellement plusieurs questions : *est-il possible de définir les dégradations induites par une acquisition réalisée avec un très faible nombre de photons ? Est-il possible de les modéliser puis les inverser ? Est-il possible de constituer des paires d'images dégradées/de vérité terrain ? Est-il possible de corriger les images à faible luminosité en préservant l'intégrité de la scène représentée ?*

Dans la suite du manuscrit, nous visons à répondre à ces questions. Le contexte du problème conduit à diverses contraintes que nous décrivons ci-dessous. Pour caractériser l'effet de la capture d'une image nocturne, nous étudions tout d'abord un exemple tiré du jeu de données Waymo [Sun *et al.* 2020] et illustré par Fig. 1. Nous appliquons naïvement une correction gamma pour faire ressortir les dégradations dans les parties les plus sombres de l'image.

Parmi les divers types de dégradations associés aux images de nuit, le faible rapport signal à bruit est sûrement le plus étudié dans la littérature. Différents modèles existent pour simuler ce bruit photonique intrinsèque à la vision à faible luminosité (voir [Wei *et al.* 2021] par exemple ou [Brooks *et al.* 2019] pour une approximation). La nature quantique du photon induit un bruit inévitable dans la sortie qui suit une distribution de Poisson. Un deuxième type de dégradation provient du fait que le taux d'humidité augmente pendant la nuit. Ceci produit un effet de flou sur les images, ainsi que du brouillard provenant de la dispersion des rayons lumineux par la vapeur d'eau [Narasimhan *et al.* 2003; Li *et al.* 2015]. Ce phénomène peut accentuer l'effet de lueur autour des sources lumineuses dans la scène.

Figure 1: De gauche à droite : un exemple d'image nocturne que nous voudrions restaurer et une version de cette image après correction gamma pour faire ressortir les dégradations induites dans les zones sombres. Source : jeu de données Waymo [Sun *et al.* 2020].

Nous allons nous concentrer dans cette thèse sur un troisième aspect, les déviations de couleur, problème aussi peu étudié dans la littérature. Le bruit nocturne étant omniprésent dans les images, il interviendra dans nos modèles, bien que notre priorité ne soit pas de le traiter. Concernant la Fig. 1, la couleur des arbres dans l'arrière-plan doit être récupérée dans le processus de restauration pour un résultat satisfaisant. La couleur peut être d'une nature essentielle dans un processus de classification ou détection d'objets par exemple, elle ne doit donc pas être ignorée. Ce manque de couleur n'est pas seulement dû à la très faible intensité des sources lumineuses de la scène, mais aussi au fait que, la nuit, la plupart de ces sources de lumière artificielle en extérieur sont colorées et non blanches. Ces rayons incidents interagissent avec les matériaux de la scène selon leurs spectres de reflectance respectifs. Des exemples de spectres de réflectance sont illustrés sur la Fig. 2. Leur capacité de réflexion n'est pas toujours élevée sur l'ensemble du spectre visible. Les "vraies couleurs" des objets peuvent être noyées dans les dégradations. Cette figure illustre une des difficultés majeures du problème. Chaque matériau a son propre spectre de reflectance associé et peut absorber certaines bandes de fréquences au lieu de réfléchir tous les rayons incidents. La modélisation de ce phénomène n'est donc pas aisée sachant qu'il est très difficile d'estimer les matériaux dans une scène de nuit. Il existe dans la littérature des tentatives de modélisation de cette propriété sous la forme d'un problème de tone mapping [Thompson *et al.* 2002], mais elles sont limitées et ne rendent pas compte de l'ensemble du phénomène. Un cas extrême est lorsque les sources lumineuses n'ont qu'une émittance spectrale avec seulement quelques pics. Dans ce cas, le capteur de la caméra ne capture qu'une information minimale de la réflectance réelle

(i.e. dans l'ensemble du spectre visible).

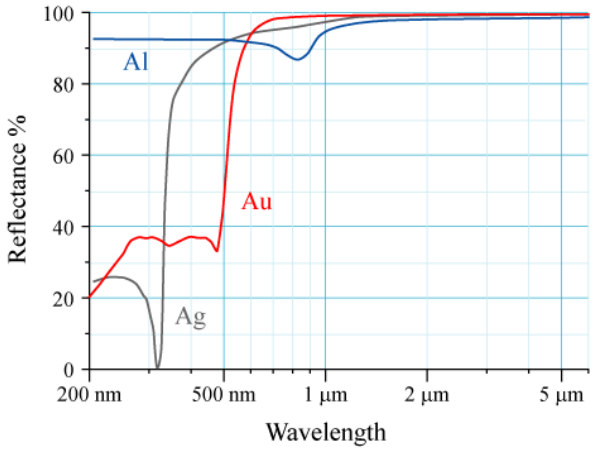


Figure 2: Exemples de spectres de réflectance de métaux. Source : *Wikipedia*

Dans ce manuscrit, nous cherchons à restaurer des images d'extérieur. Ce problème se différencie de la restauration d'images en intérieur pour plusieurs raisons qui impliquent la nécessité de développer de nouvelles approches spécifiques à ce contexte. Premièrement, les images contiennent des scènes complexes avec une grande profondeur de champ. Deuxièmement, il y a une difficulté intrinsèque à collecter des paires d'images jour/nuit en nombre suffisant pour des méthodes basées sur l'apprentissage profond. Nous détaillons les jeux de données les plus courants pour la restauration en basse lumière et leurs caractéristiques dans la Section 1.2.1. En résumé, les jeux de données existants avec des paires ne contiennent que des images d'intérieur ou simulent la dégradation à faible luminosité en modifiant certains paramètres d'exposition (par exemple l'ISO). Cela ne permet pas de modéliser la véritable dégradation nocturne, comme nous l'avons vu précédemment. C'est pourquoi nous avons décidé d'étudier en priorité des approches non supervisées et d'entraîner nos réseaux sur des distributions empiriques valides dans le chapitre 4.

L'une des principales propriétés du processus de restauration que nous souhaitons définir est la préservation de l'information contenue dans l'image d'entrée. Aucune hallucination ne doit remplacer son contenu en ajoutant des objets là où ils ne sont pas censés se trouver. Ceci nécessite un contrôle particulier des méthodes génératives. Les dégradations sont supposées être déterministes, à l'exception du bruit.

Plan détaillé

Pour situer nos contributions dans le contexte de la littérature, nous commençons par passer en revue un ensemble de méthodes de l'état de l'art concernant la restauration d'images à faible luminosité dans le chapitre 1. Nous présentons les méthodes basées sur des a priori sur les caractéristiques d'images bien éclairées en améliorant le contraste des images. Après les méthodes basées sur l'égalisation d'histogrammes, nous introduisons le lecteur à des méthodes qui restaurent les images en les décomposant à l'aide d'un modèle physique telles que LIME. Ensuite, nous abordons les méthodes d'apprentissage profond à la fois celles supervisées mais aussi les plus récents essais de correction d'images sans supervision. Enfin, à la fin du chapitre, la question délicate mais néanmoins essentielle des métriques que nous pouvons utiliser est traitée. Nous listons les métriques existantes avec leurs avantages et désavantages dans notre contexte.

La théorie Retinex est particulièrement bien adaptée au cas de la restauration d'images de nuit puisqu'elle permet une décomposition des images en séparant l'illumination de la scène de son contenu. Différents modèles ont été adoptés dans la littérature. Nous les présentons en détail dans le chapitre 2. Ceci nous permettra de proposer un nouveau modèle dans le chapitre 4. Dans ce deuxième chapitre, l'histoire des premières interprétations de la théorie Retinex est rappelée dans un premier temps. Après l'interprétation originale perceptuelle, nous nous concentrons sur l'interprétation physique de Horn. Nous détaillons son lien avec notre problème et les différentes propriétés que nous pouvons considérer pour la restauration résultant des fortes dégradations. Nous mettons l'accent sur le fait que le modèle de décomposition est seulement valide sur les données brutes en entrée avant les opérations non-linéaires de la pipeline de formation d'image standard RVB. Il faut ainsi les inverser avant de pouvoir décomposer l'image. Ensuite, nous montrons le lien avec le problème de constance de couleur. Retrouver les "vraies" couleurs de la scène est équivalent à une décomposition Retinex parfaite avec une illumination en couleur.

Le chapitre 3 est notre premier chapitre de contribution. Nous analysons les caractéristiques des images à faible luminosité et leur composante d'illumination selon la théorie Retinex et identifions une solution triviale qui n'a pas été prise en compte par les méthodes non supervisées de l'état de l'art. Le défi vient du fait que la solution triviale ne peut pas être complètement éliminée de l'ensemble réalisable car elle correspond à une

solution valide lorsque l'image contient une source lumineuse ou une zone surexposée. Pour résoudre ce problème, nous proposons un nouveau terme de régularisation qui conserve les solutions plausibles dans l'ensemble de solutions explorées. Pour démontrer l'efficacité de l'a priori proposé, nous menons des expériences en utilisant des a priori avec des réseaux de neurones convolutifs dans un cadre similaire au travail récent RetinexDIP [Zhao *et al.* 2021] et une étude d'ablation approfondie. Enfin, nous n'observons plus d'artefacts de halo dans l'image restaurée. Pour toutes les métriques, sauf une, notre approche non supervisée donne des résultats aussi bons que l'état de l'art supervisé, ce qui indique le potentiel de ce cadre pour l'amélioration des images à faible luminosité.

Dans le chapitre 4, nous étudions la restauration d'images à faible luminosité avec des scènes extérieures sans vérité terrain. Jusqu'à présent, les approches dans la littérature ont évité d'utiliser le modèle de décomposition Retinex de manière non supervisée ou ont ajouté des a priori contraignants sur les composantes recherchées. Nous proposons ici de retirer la contrainte d'une illumination lisse en niveaux de gris du modèle Retinex. En effet, selon la physique de la lumière, ce modèle devrait considérer une illumination colorée. À partir de ce nouveau modèle de décomposition, nous formulons une nouvelle architecture basée sur l'apprentissage profond et inspirée par le transfert de style. Notre méthode permet de visualiser l'illumination (i.e. un style complexe ayant les mêmes dimensions qu'une image) et la réflectance (i.e. le contenu). Notre approche permet d'obtenir des composantes plus réalistes par rapport à l'état de l'art, c'est-à-dire sans artefact, sans amplification du bruit et sans hallucination, avec une restauration simple pour chacune des composantes. Ce problème est en effet difficile car non seulement nous visons à résoudre en même temps un problème de séparation de sources corrélées et un problème sous-déterminé, où le nombre d'observations est plus faible que le nombre de composantes à estimer. L'approche proposée garantit qu'aucune hallucination n'est ajoutée en sortie.

Enfin, dans le chapitre 5, nous reformulons le problème de restauration d'images de nuit dans le cadre des méthodes de transport optimal. Ce cadre nous permet de formaliser la restauration à appliquer tout en répondant à notre critère de conservation de la scène dans l'image. Nous mettons en évidence le lien entre notre objectif de restauration d'images de nuit à l'aide du modèle de décomposition Retinex et le problème du pont de Schrödinger à travers l'information mutuelle. Cette solution dynamique de transport

optimal régularisée par un a priori entropique est calculée grâce à un modèle de diffusion. Diverses sources de biais sont présentes dans le calcul de cette solution. Afin de les minimiser, les approches existantes se concentrent sur l'algorithme de calcul ou bien sur des termes de régularisation. Un algorithme utilisant deux processus de diffusion codépendents est rappelé ainsi que deux a priori permettant de maximiser l'information mutuelle sur l'ensemble du chemin de transport. Ceci doit permettre en théorie de mieux conserver l'information dans l'image d'entrée pour une meilleure restauration. Les deux a priori consistent en un terme de régularisation adversaire et un terme d'apprentissage contrastif. Nous proposons en plus de cela de prendre en compte les spécificités physiques de notre problème à travers le modèle Retinex pour améliorer davantage les solutions trouvées.

INTRODUCTION

Application context

Learning-based techniques are becoming increasingly popular in computer vision and signal processing since the groundbreaking works involving convolutional neural networks. They were trained with backpropagation algorithms and stochastic gradient descent. One of their first applications was handwritten number and character recognition (see the original LeNet in [LeCun *et al.* 1989] for instance). Nowadays, we can find deep learning algorithms in almost all computer vision and signal processing challenges particularly when large quantities of data can be collected. A major challenge remaining for these applications is their use in extreme visibility conditions. Indeed, capturing images at night still cause a major problem for the navigation of autonomous vehicles, or in satellite imaging.

In the first case, these navigation systems rely on high-level tasks such as object classification and segmentation. Different modalities can be used for their operation (e.g. standard cameras, LIDAR sensors, ...). In practice, autonomous vehicles sometimes drive at night and fail to identify their surroundings correctly. The image processing algorithms were not robustly designed to operate when acquisition is carried out with a limited number of photons. Recently, many towns and cities, such as Toulouse and Lyon, are reducing their street lighting in an attempt to reduce their carbon footprints and only keep the traffic lights on. These colored light sources are the cause of problems identified in the literature as *color deviations*, which in addition to the very low signal-to-noise ratio.

The restoration of low-light images

In this thesis, we seek to define novel deep learning-based methods to restore low-light and more precisely night outdoor images. We focus on RGB images and thus the visible spectrum.

The problem naturally raises several questions: *can we define the degradations induced by a capture carried out with a very low number of photons? Can we model the degradations then correct them? Can we constitute pairs of degraded/ground-truth images? Can we correct low-light images while preserving their information?*

We aim at addressing these questions in the rest of the manuscript. The context of the problem leads to various constraints that we describe below.

Identifying the degradations

To characterize the effect of capturing an image at night, we first study an example from the Waymo dataset [Sun *et al.* 2020] depicted in Fig. 3. We apply a naive gamma correction to bring out the degradations in the darkest parts of the image.

Figure 3: From left to right: example of a night image that we would like to restore and a gamma corrected version of the image showing the high level of noise in the dark regions. Source: Waymo dataset [Sun *et al.* 2020]

Among the various types of degradation associated with night images, the low signal-to-noise ratio is certainly the most studied in the literature. Diverse realistic models have been designed to simulate this photon shot noise intrinsic to low-light images (see

[Wei *et al.* 2021] for instance or [Brooks *et al.* 2019] for an approximation). The quantum nature of the photon induces an unavoidable noise in the resulting output following a Poisson distribution. A second type of degradation arises from the fact that humidity levels increase at night. This produces a blurred effect on the images, as well as fog due to the light rays being scattered by the water vapor [Narasimhan *et al.* 2003; Li *et al.* 2015]. This phenomenon can accentuate the glowing property of the light sources in the scene.

In this thesis, we will focus on a third aspect, color deviations, a problem which received very limited attention from the community. As photon shot noise is omnipresent in night images, it will be included in our models, although our priority is not to deal with it. Regarding Fig. 3, the color of the trees in the background has to be recovered in the restoration process. The color can be of essential nature in the classification process for example, so it should not be ignored. This lack of color not only comes from the very low intensity of the reflected light but also from the fact that at nighttime, most of the outdoor artificial light sources are colored and not white. These incident rays interact with the materials in the scene according to their respective reflectance spectrum. Examples of such reflectance spectra are illustrated in Fig. 4. Their reflectance value is not always high for the entire visible spectrum. The "true colors" of the objects may be buried in the degradations. This figure illustrates one of the major difficulties of the problem. Each material has its own reflectance spectrum and may absorb certain frequency bands instead of reflecting all the incident rays. Modelling this phenomenon is therefore complex while estimating the materials in the scene is also difficult. Some attempts to model this property as a tone mapping problem exist [Thompson *et al.* 2002] but are limited and do not capture the whole phenomenon. In the worst case scenario, the light sources only have spectral emittance with only few pikes. Then, the camera sensor only captures minimal information about the true reflectance (i.e. in the whole visible spectrum).

Outdoor images

In this manuscript, we seek to restore outdoor images. This problem is different from restoring indoor images for several reasons which imply that new approaches specific to this context are required. First, the images contain complex scenes with a wide depth-of-field range. Secondly, there is an intrinsic difficulty in collecting a sufficient number of day/night image pairs for deep learning-based methods. We detail the most common

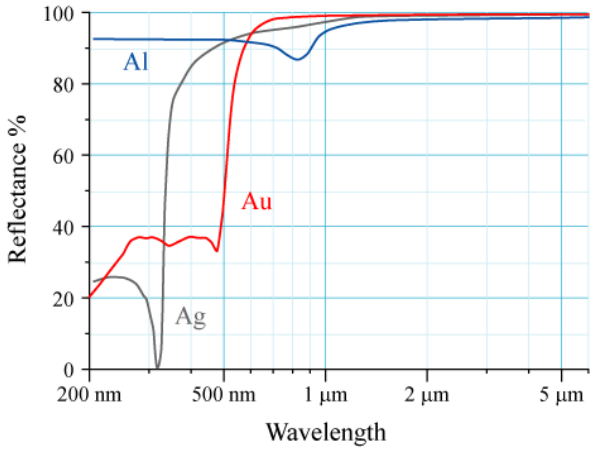


Figure 4: Examples of metal reflectance spectra. Source: *Wikipedia*

datasets for low-light restoration and their characteristics in Section 1.2.1. In short, the existing datasets with pairs only contain indoor images or simulate the low-light degradation by changing some exposure parameters (e.g. ISO). This does not model the true night degradation as discussed in the previous section. That is why we decided to study unsupervised approaches and train our networks on valid empirical distributions in Chapter 4.

Preserving the integrity of the image

One of the key properties of the restoration process that we want to define is the preservation of the information in the input image. No hallucination should replace its content adding objects where they are not supposed to be. Extra attention is therefore paid to the usage of generative models. The degradations are assumed to be deterministic apart from the noise.

Detailed outline

To situate our contributions in the context of the literature we start by reviewing a part of the state-of-the-art methods regarding the restoration of low-light images in Chapter 1. We present the contrast enhancement methods based on priors on normal-light images characteristics. Afterward, we introduce the reader to methods that restore low-light images by decomposing them using a physical model, such as LIME [Guo *et al.* 2017].

Then, we introduce the deep learning-based approaches both supervised and the more recent attempts to correct a low-light image without supervision. Finally, at the end of the chapter, the delicate though crucial question of the suitable metrics we can use is treated. We list the existing metrics as well as their pros and cons in our context.

The Retinex theory is particularly well suited to the restoration of nighttime images, as it can be used to decompose an image into the illumination of the scene and its reflectance. Various models have been adopted in the literature. We introduce them in Chapter 2. This will enable us to propose a new model in Chapter 4. In this second chapter, the history of the first interpretations of the Retinex theory is recalled. After the original perceptual interpretation, we focus on Horn's physical interpretation. We detail how it is related to our problem and the different properties we can consider for the restoration resulting from the heavy degradations. We emphasize on the necessity to linearize the images before the application of the Retinex model. The decomposition model is only valid on the raw data before the nonlinear operations of the camera image processing pipeline. Then, we show the link with the Color constancy problem. Recovering the "true" colors of the scene is equivalent to the perfect Retinex decomposition with a colored illumination.

Chapter 3 is our first contribution chapter. We analyze the characteristics of low-light images and their illumination component according to the Retinex theory and identify a trivial solution not taken into consideration by the previous unsupervised state-of-the-art methods. The challenge comes from the fact that the trivial solution cannot be completely eliminated from the feasible set as it corresponds to the true solution when the low-light image contains a light source or an overexposed area. To address this issue, we propose a new regularization term which only remove absurd solutions and keep plausible ones in the set. To demonstrate the efficiency of the proposed prior, we conduct our experiments using deep image priors in a framework similar to the recent work RetinexDIP [Zhao *et al.* 2021] and an in-depth ablation study. Finally, we observe no more halo artifacts in the restored image. For all but one metric, our unsupervised approach gives results as good as the supervised state-of-the-art indicating the potential of this framework for low-light image enhancement.

In Chapter 4, we study the restoration of low-light images with outdoor scenes without ground truth. Until now, approaches in the literature have avoided using the Retinex

decomposition model in an unsupervised way or have added constraining priors on the searched components. We propose here to relax the constraint of a smooth grayscale illumination of the Retinex model. Indeed, according to the physics of light, it should include a colored illumination. Resulting from this new decomposition model, we formulate a new deep learning-based architecture inspired by the style transfer methods. Our method enables us to visualize the illumination (i.e. a complex style with the same dimensions as an image) and the reflectance (i.e. the content). It achieves more visually pleasing components compared to the state-of-the-art i.e. without artifact, without noise amplification and without hallucination with a simple restoration for each of the components. This problem is indeed difficult because not only we seek to solve a correlated source separation problem and a nonlinear inverse problem at the same time, but also to solve an underdetermined problem, where the number of observations is smaller than the number of components to estimate. The proposed approach ensures that no hallucination is added.

Finally, in Chapter 5, we reformulate the restoration of night images in the optimal transport framework. In this framework, we do not need to determine the degradation to formalize the restoration process. Moreover, it meets our criterion of preserving the integrity of the scene in the image. We highlight the link between our objective of restoring low-light images with the Retinex decomposition and the problem of the Schrödinger bridge through mutual information. This dynamical optimal transport solution regularized with an entropy prior is computed with a diffusion model. Multiple sources of bias are present in the calculation of this solution. To mitigate them, existing approaches improve the algorithm to compute the transport plans or on regularization terms for these computations. An algorithm using two codependent diffusion processes is recalled as well as two priors maximizing the mutual information over the entire transport path. In theory, this should make it possible to better conserve the information in the input image for better restoration. The two a priori consist of a regularization term and a contrastive learning term. In addition, we propose to take into account the physical specificities of our problem through the Retinex model to further improve the solutions found.

PUBLICATIONS

International conferences

A. Lecert, R. Fraisse, A. Roumy, and C. Guillemot, "A new regularization for retinex decomposition of low-light images," in 2022 IEEE international conference on image processing (ICIP), Oct. 2022, pp. 906–910. doi: 10.1109/ICIP46576.2022.9897893.

International journals

A. Lecert, A. Roumy, R. Fraisse, and C. Guillemot, "GAN architecture leveraging a retinex model with colored illumination for low-light image restoration," IEEE Access, vol. 11, pp. 84574–84588, 2023, doi: 10.1109/ACCESS.2023.3301614.

National conferences

Arthur Lecert, Aline Roumy, Renaud Fraisse, Christine Guillemot (2023). Illumination colorée et information mutuelle pour le modèle Retinex en restauration d'images à faible luminosité. In GRETSI 2023, Colloque Gretsi (Symposium of the French Research group on signal and image processing).

In preparation

Arthur Lecert, Renaud Fraisse, Aline Roumy, Christine Guillemot. Restoring low-light images with Retinex Schrödinger Bridges

Chapter 1

State-of-the-art on the restoration of low-light images

1.1 Model-based algorithms . . . 10
1.1.1 Contrast enhancement . . . 10
The gamma correction . . . 10
Histogram equalization & variants . . . 11
1.1.2 LIME . . . 12
1.2 Deep learning approaches . . . 14
1.2.1 Supervised methods . . . 14
Typical datasets . . . 14
Retinex-based methods with handcrafted priors . . . 17
1.2.2 Unsupervised methods . . . 19
Tone mapping . . . 19
Style transfer methods . . . 20
Regular GANs . . . 23
Deep Image prior & RetinexDIP . . . 24
1.3 Metrics . . . 25
1.3.1 Reference-based metrics . . . 26
Pixel-wise metrics . . . 26
Perceptual metrics . . . 28
1.3.2 No-reference metrics . . . 29
Intensity-based metrics . . . 29
Perceptual metrics . . . 31
1.4 Chapter conclusion . . . 33

In this chapter, we review relevant state-of-the-art methods on the problem of restoring low-light images. We do not intend it to be exhaustive but aim at covering the main directions taken by the community. The first approaches we review employ a prior model (i.e. in its most general definition), a preconceived idea, on the desired images. For instance, the contrast enhancement-based methods inherently consider that a restored image has a flat/spread histogram. Other methods consider a physical model to decompose the image and restore the components individually. We then review both recent supervised and unsupervised deep learning-based approaches as they are currently the state-of-the-art in the literature. Finally, we recall the different metrics used in most of the quantitative studies we could find on the restoration of low-light images.

1.1 Model-based algorithms

1.1.1 Contrast enhancement

The problem of restoring low-light images has been traditionally cast as a contrast enhancement problem. This approximation made it possible to develop algorithms combining speed and effectiveness.

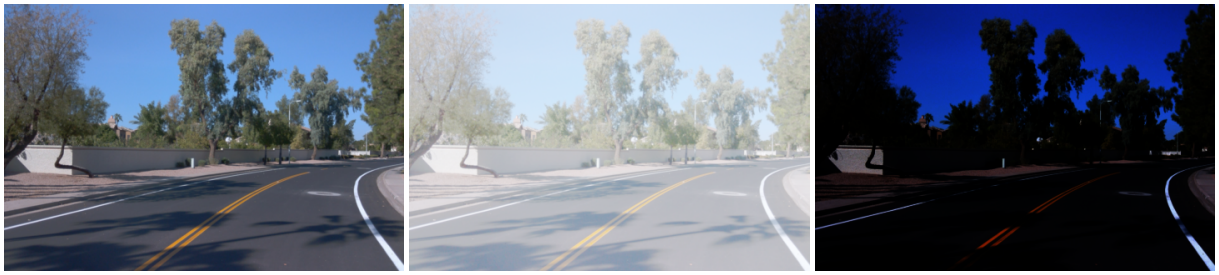
The gamma correction

One of them, the gamma correction, consists of a simple nonlinear operation applied to an image $I \in [0,1]^{3n}$ where n is the number of pixels. It boils down to the following equation:

$$I_{\text{out}} = I_{\text{in}}^{\gamma} \tag{1.1}$$

This tone mapping function was first use to encode an image more efficiently, reducing the bandwidth required to transmit an image based on the physical properties of the human eye. Indeed, since humans better distinguish between dark tones than brighter tones, one can allocate fewer or more bits depending on the region of the image. The image maintains the same perceptual quality after the process. This process is illustrated on Fig. 1.1 with *gamma encoding* ($\gamma > 1$) and *gamma decoding* ($\gamma < 1$). Here, the "restoration" is only a side effect of the decoding process. It boosts low intensity values of the pixels in the image. This preserves the integrity of the scene in the image but raises a question on finding which gamma value to apply and to which pixels. The simple gamma correction process does not take into account the different effects of the low-light degradation on the

capture of the image. The output image is a coarse noisy approximation of the restored image.



$\gamma = 1$ $\gamma = 0.25$ $\gamma = 4$

Figure 1.1: Example of a gamma encoding/decoding an image. Source: Waymo dataset [Sun *et al.* 2020]

Histogram equalization & variants

Increasing the contrast of an image can be done by stretching its histogram to flatten it. This process is called histogram equalization. An illustration is shown in Fig. 1.2. Flattening the histogram better distribute the intensity values of the pixels across the range used to display an image. It results that if the image is mostly dark, the objects in these areas will be more easily recognizable in the resulting image. However, artifacts might be added by processing the image globally and not considering the local differences specific to each image.

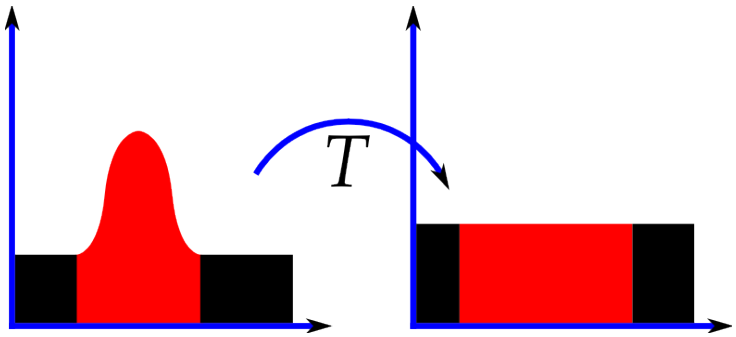


Figure 1.2: Illustration of the effect of histogram equalization. Source: *Wikipedia*

To address this issue, adaptive histogram equalization methods (AHE) have been developed historically for use in aircraft cockpit displays [Ketcham *et al.* 1974; Hummel *et al.* 1977]. They restore the images according to the histograms of each pixel and its neighborhood. The most simple variant is a square neighborhood as depicted in Fig. 1.3.

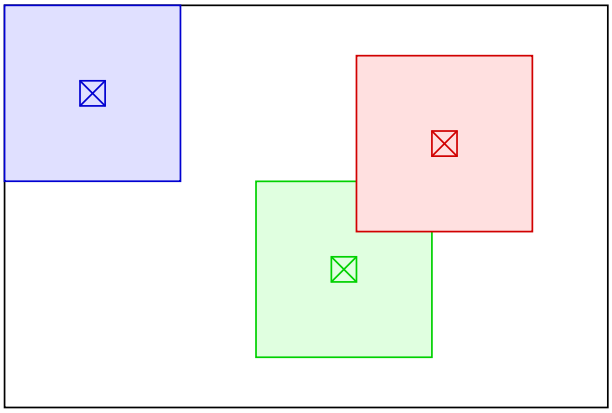


Figure 1.3: Example of a type of neighborhood used in adaptive histogram methods. Source: *Wikipedia*

However, adaptive histogram equalization might increase the intensity of pixels in flat regions more than required. To tackle this downside, a contrast limited adaptive histogram equalization (CLAHE [Pizer *et al.* 1990]) was developed. A clipping value is set to prevent the amplification of values over the limit as shown in Fig. 1.4.

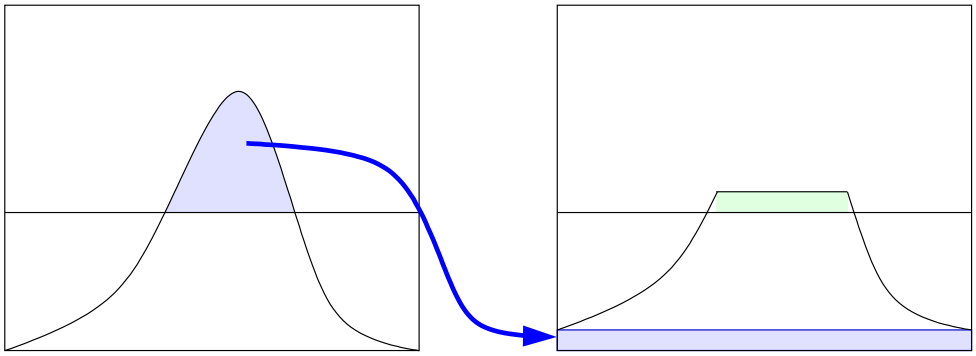
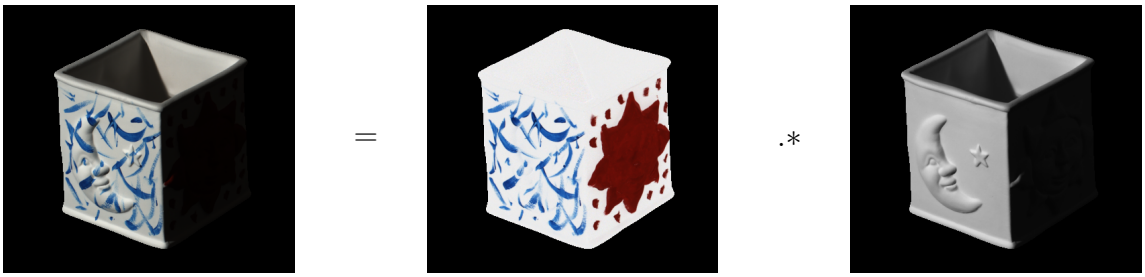


Figure 1.4: Illustration of clipping the amplification of the pixel values. Source: *Wikipedia*

1.1.2 LIME

LIME [Guo *et al.* 2017] is an approach that estimates the illumination of the scene to restore a low-light image. Thus, it is well-suited for our problem. At its core, the authors follow a common interpretation of a decomposition model obeying the Retinex theory [Land *et al.* 1977; Barrow *et al.* 1978] detailed in Chapter 2. Since we employ the same model in our methods, a comparison with the baseline algorithm is natural.

According to this interpretation, an image $I \in \mathbb{R}^{3n}$ is a noisy observation of the product



Original image = The reflectance .* The illumination

Figure 1.5: Illustration of the target components of the common interpretation found in the literature of the Retinex decomposition model. The images come from the MIT intrinsic dataset [Grosse *et al.* 2009].

of two components

$$I = L. * R + \eta \tag{1.2}$$

where $.*$ is the element-wise product, $L \in \mathbb{R}^n$ the grayscale illumination map, $R \in \mathbb{R}^{3n}$ the diffuse reflectance component, and η the additive Gaussian noise. The illumination map is thus repeated for each color channel of the reflectance. This model can be seen as a scale mixture since one pixel is the product of a scaling factor (i.e. the illumination) applied to all three RGB components of the reflectance. The illumination component contains the lightness information of the scene including shadows or light sources. The reflectance consists of the intrinsic color of the elements of the scene regardless of the exposure conditions. Any specularity component is ignored in this model. Fig. 1.5 illustrates the target components of the decomposition.

In addition to this model, they define a simple but powerful prior to estimate the illumination component as a first guess:

$$\tilde{L} = \max_{c \in \{R,G,B\}} I_c \tag{1.3}$$

Then, the authors drift from this solution by smoothing the illumination. Indeed, they argue that in a ideal decomposition, the illumination map should not contain any texture details but still keep the structure of the scene. They solve the following optimization problem:

$$\hat{L} = \underset{L}{\operatorname{argmin}} \|\tilde{L} - L\|_F^2 + \alpha \|W. * \nabla L\|_1 \tag{1.4}$$

where ∇L is the gradient of illumination both vertically and horizontally, W a weighting

matrix. In the comparisons we make with LIME, we use the third strategy to determine W. Afterward, the authors obtain the reflectance thanks to the decomposition model (1.2) (i.e. $\hat{R} = I./(\hat{L} + \varepsilon)$, $./$ being the element-wise division). They restore the illumination with a gamma correction ($\gamma = 0.8$). Due to the illumination being smooth, all high frequencies end up in the reflectance. Therefore, they denoise the reflectance component with BM3D [Dabov *et al.* 2007]. The output image is finally reconstructed recombining the components.

1.2 Deep learning approaches

In the last decades, significant breakthroughs in inverse problems were achieved thanks to learning-based methods and especially deep neural networks. We sum up in the next sections approaches trained in a supervised or unsupervised fashion to restore low-light images.

1.2.1 Supervised methods

The first and most intuitive way to address an inverse problem with deep learning networks is to try to build a dataset of pairs of original/degraded signals to learn to reverse the degradation mapping between the two sets.

Typical datasets

Obtaining night/day image pairs of the same outdoor scene is very challenging. Thus, existing datasets do not contain true day/night image pairs but rather estimated ones. For instance, in the database FiveK [Bychkovsky *et al.* 2011], the normal-light target image is made by human experts. Other datasets consider pairs with modified exposure parameter (e.g. ISO) of the scene, which is different from a night/day illumination (e.g. NPE [Wang *et al.* 2013], LIME dataset which is called HDR by the original authors [Sen *et al.* 2012], VV [Vonikakis *et al.* 2018], MEF [Ma *et al.* 2015], LOL [Wei *et al.* 2018], DICM which is a mix of images from USC-SIPI [Weber *et al.* 2018] and a True Color Kodak images database [Kodak *et al.* 1999] or the multi exposure dataset from [Afifi *et al.* 2021]). Therefore, methods trained on these datasets learn to compensate the exposure parameter (e.g. ISO) change during the capture, but not the lighting change (e.g. colored streetlamps during the night versus white sun lighting at daytime). As a first step, we choose to work

with the LOL dataset [Wei *et al.* 2018] in Chapter 3. This facilitates the comparison with state-of-the-arts methods. Fig. 1.6 depicts examples of paired images from the LOL dataset [Wei *et al.* 2018].

Figure 1.6: Examples of paired images from the LOL dataset [Wei *et al.* 2018]. From left to right: the ground-truth image and the degraded version.

Beyond the LOL dataset, we also considered the Waymo dataset [Sun *et al.* 2020] to train our network in Chapter 4. This dataset does not contain any night/day image pairs of the same scene but provides true day and true night outdoor scenes made for autonomous vehicles. The other advantage of this dataset is that it provides much more images than the LOL dataset [Wei *et al.* 2018]. It contains 128 093 of normal-light images and 15 419 low-light images compared to the 500 previous pairs. Fig. 1.7 depicts examples of images from the training set of the Waymo dataset [Sun *et al.* 2020].

There exist datasets providing ground-truths components close to the Retinex decomposition such as the MIT dataset [Grosse *et al.* 2009]. The authors directly capture "intrinsic images" with a clever method. They build a dataset consisting of the original

Figure 1.7: Examples of non-paired images from the Waymo dataset [Sun *et al.* 2020]. From left to right: a day and a night image.

image with specularities, the diffuse version (i.e. without specularities) and the shading component. Knowing these components they compute the reflectance and the specularity component. Some of the components are similar to the ones in the Retinex model Equation (1.2) commonly found in the literature. The illumination component in this interpretation corresponds to the shading component in the MIT dataset. The link between the dataset and the model is depicted in Fig. 1.5. This dataset has two major drawbacks, however. It consists of only 15 images which greatly reduce the possibilities to train a deep neural network on it. Moreover, it only contains the ground-truth normal-light Retinex components but not their degraded counterparts in low-light conditions. Thus, it cannot be used to determine the low-light degradations. It can be used instead for a source separation problem in daylight which is not our objective here. Also, the images only represent simple scenes of one object maximum and not complex scenes that we can find in outdoor images.

Retinex-based methods with handcrafted priors

Wei *et al.* [Wei *et al.* 2018] were the first to propose a Retinex-based method to restore low-light images with deep neural networks. They dubbed their model Retinex-Net. Their enhancement pipeline is a three-steps process. They first decompose the input image according to the common interpretation of the Retinex theory (1.2), correct each component respectively and then reconstruct the restored image. The whole process is illustrated in Fig. 1.8. Having access to the ground-truth, the authors can define strong priors to alleviate the difficulty of the decomposition. They first use a reconstruction prior to force the two components to follow the Retinex model:

$$\mathcal{L}_{\text{recon}} = \|L. * R_{\text{low}} - I_{\text{normal}}\|_1. \tag{1.5}$$

In the previous equation, R_{low} is the low-light reflectance and I_{normal} is the normal-light version of the image i.e. the ground-truth restored image. Since the reflectance component is supposed to be equal in this interpretation between the low and the normal-light version of the image, they impose a consistency prior on the reflectance such that:

$$\mathcal{L}_{\text{RC}} = \|R_{\text{low}} - R_{\text{normal}}\|_1. \tag{1.6}$$

Building on the idea that the grayscale illumination map should be smooth while still keeping the structure of the scene, a final prior is added:

$$\mathcal{L}_{\text{IS}} = \|\nabla L_{\text{low}}. * \exp(-\lambda \nabla R_{\text{low}})\|_1 + \|\nabla L_{\text{normal}}. * \exp(-\lambda \nabla R_{\text{normal}})\|_1 \tag{1.7}$$

where ∇L the gradient both vertically and horizontally of the illumination L, λ a weight to balance the terms. The extracted reflectance is denoised using BM3D [Dabov *et al.* 2007] before the reconstruction.

Shortly afterwards, Zhang *et al.* [Zhang *et al.* 2019] proposed an improved deep neural architecture based on the same three-steps pipeline and the same decomposition model (1.2). Fig. 1.9 depicts framework of the restoration process. The contributions of their work can be sum up in the following manner. They reformulate the illumination smoothing prior (1.7) as:

$$\mathcal{L}_{\text{IS}} = \left\| \frac{\nabla L}{\max(\nabla I, \epsilon)} \right\|_1. \tag{1.8}$$

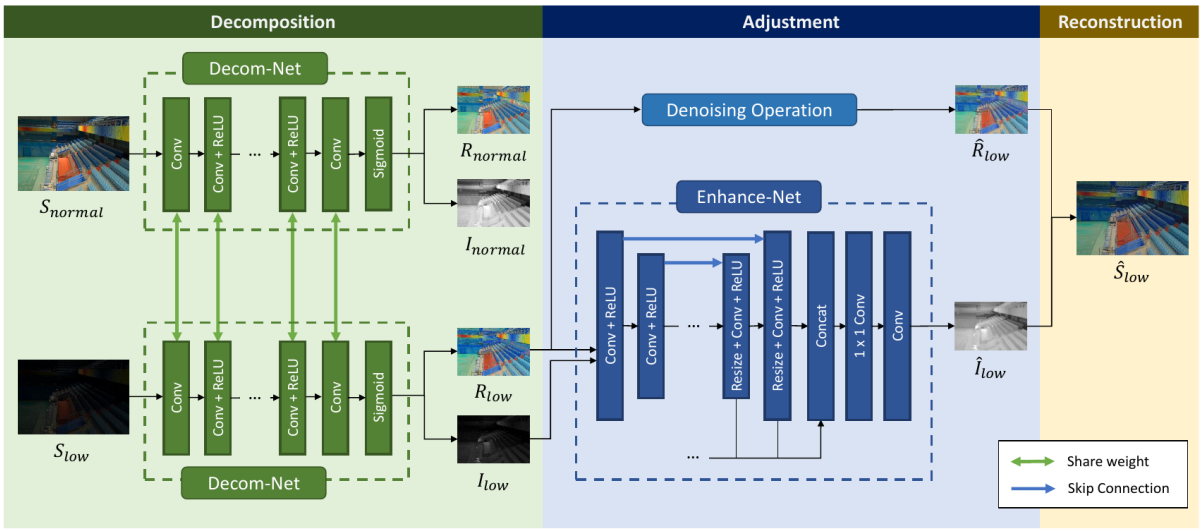


Figure 1.8: Framework for the Retinex-Net proposed in [Wei *et al.* 2018].

If ∇I has high values, the penalty will be small to smooth the surface whereas low values will give a high penalty to force the illumination component L to be close. They restore the reflectance with an additional network instead of only denoising the component with BM3D [Dabov *et al.* 2007].

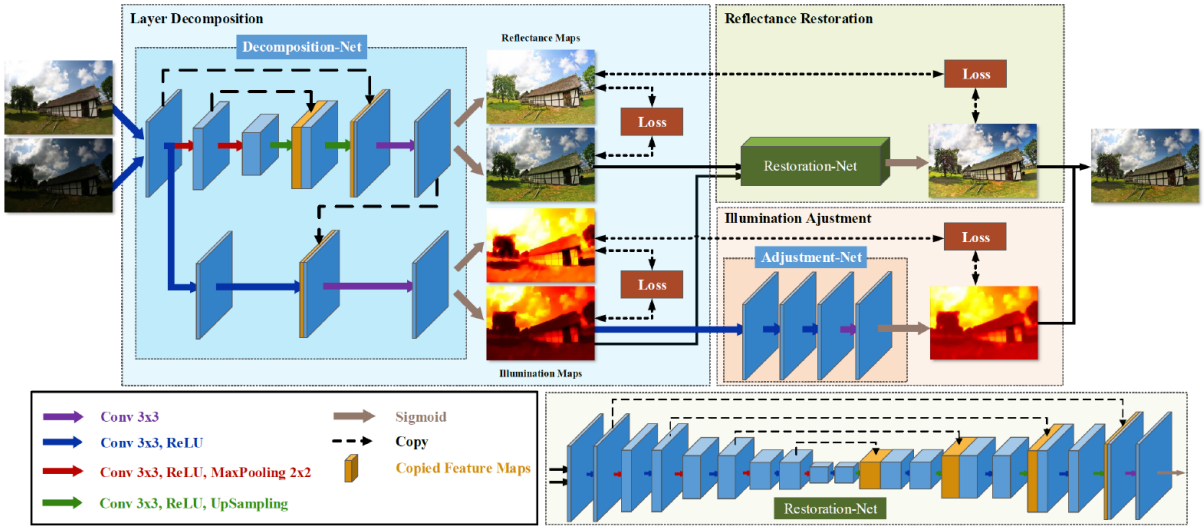


Figure 1.9: Framework for the KinD network proposed in [Zhang *et al.* 2019].

The state-of-the-art methods for Retinex decomposition are based on deep neural networks trained on datasets providing paired images in an end-to-end manner. While they can efficiently decompose an image with powerful supervised priors knowing the ground-

truth, relying on a dataset remains a problem. Indeed, as argued in Section 1.2.1, these approaches only correct underexposed images while night images for instance suffer from more difficulties. Thus, when these networks are applied to natural out-of-distribution images, the restored images may contain artifacts or color deviations.

1.2.2 Unsupervised methods

Unfortunately the outdoor degradation that we want to reverse makes it challenging to constitute a relevant dataset. Recent works on the same problem depart from supervised learning and focus on unsupervised learning instead.

Tone mapping

A whole set of approaches cast the problem of restoring low-light images as an optimization problem in which the goal is to find the best tone mapping function (or curve function) to map the low-light images to the normal-light domain. It can be seen as trying to find the best parameters for a more general family of functions than the gamma correction Section 1.1.1.

In their work [Li *et al.* 2020], Li *et al.* proposes the Zero-DCE method shown in Fig. 1.10. At its core, it is an iterative algorithm which applies successively n times the pixel-wise quadratic curve defined as:

$$LE_n(x) = LE_{n-1}(x) + \mathcal{A}_n(x)LE_{n-1}(x)(1 - LE_{n-1}(x)). \tag{1.9}$$

For the first step, the input is the low-light image I while $\mathcal{A}_n(x)$ is the output parameter map of a neural network for the light-enhancement curves. The authors also introduce a prior to smooth the illumination imposing the parameter value to be locally equal in the output.

This method present multiple pros. Indeed, it is lightweight and fast and does not add hallucination. However, the low-light noise is increased instead and it requires a careful selection of the training data such as a dataset containing multi-exposure unpaired images. Moreover, it is designed to restore underexposed images which is different from our objective. We highlight these differences in Chapter 4.

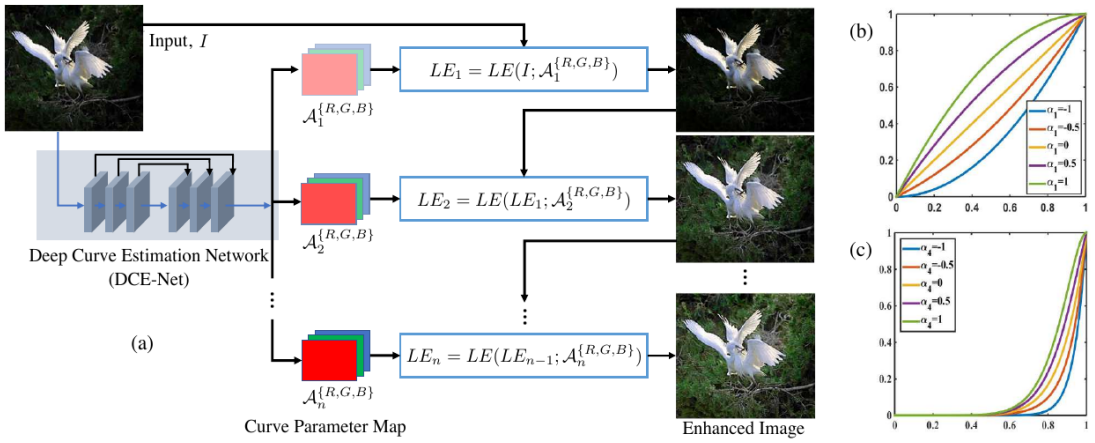


Figure 1.10: Framework for the Zero-DCE network proposed in [Li *et al.* 2020].

Style transfer methods

Style transfer is a very general problem where one tries to apply the "style" of an image to another while preserving the assumed domain-invariant "content" of the scene. Here, we will focus on generative models which recombine a normal-light style with a content from a low-light image to restore it.

MUNIT [Huang *et al.* 2018] is a method based on generative adversarial networks (GANs) [Goodfellow *et al.* 2014]. The authors manage to extract the style from the content thanks to defining clever loss terms and by assuming that the style component follows a Gaussian distribution.

This assumption originates from the work of Ulyanov *et al.* [Ulyanov *et al.* 2017] with Instance normalization (1.10) as a new operation for neural networks. The authors discovered that the spatial mean and spatial variance of an image on each of its color channels carried information about its style. Normalizing an input with respect to these parameters helps to remove the style.

$$I_{\text{out}} = \frac{I_{\text{in}} - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{1.10}$$

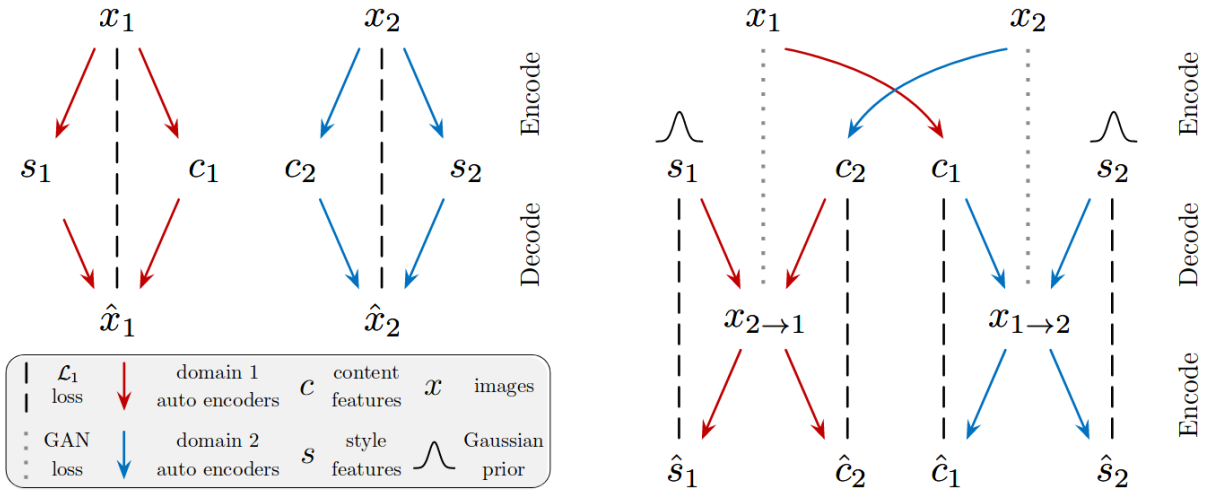
$$\mu = \frac{1}{HW} \sum_H \sum_W I_{\text{in}} \tag{1.11}$$

$$\sigma = \frac{1}{HW} \sum_H \sum_W (I_{\text{in}} - \mu)^2 \tag{1.12}$$

Building on this idea, Huang *et al.* [Huang *et al.* 2017] defined the Adaptive Instance normalization (1.13). It efficiently transfers the style of one image to another by aligning the channel-wise mean and variance of the former to the latter. It also does not introduce additional parameters to learn.

$$\text{AdaIN}(I_1, I_2) = \sigma(I_2)\left(\frac{I_1 - \mu(I_1)}{\sigma(I_1)}\right) + \mu(I_2) \tag{1.13}$$

MUNIT [Huang *et al.* 2018] is thus conceived around this idea. To learn the mapping functions between two domains with empirical distributions (i.e. datasets), Huang *et al.* train two autoencoders (red and blue respectively in Fig. 1.11). The loss terms ensure that they can reconstruct the original images while being able to transfer the style to the other image. Finally the discriminators force the generated images to set on the image manifold, creating plausible outputs.



(a) Within-domain reconstruction

(b) Cross-domain translation

Figure 1.11: Illustration of the reconstruction and cross-domain loss terms in MUNIT [Huang *et al.* 2018].

Pizzati *et al.* extend the MUNIT approach with their CoMoGAN method [Pizzati *et al.* 2021]. A physics model is added to guide the network to continuously translate an image on a functional manifold to another domain. We focus on their example of a cyclic target with a style transfer from Day to Dusk ⇒ Night ⇒ Dawn and Day again. The process is illustrated on Fig. 1.12. To achieve that, the authors propose Functional Instance

normalization:

$$\mathrm{FIN}(I, \phi) = \frac{I - \mu(I)}{\sigma(I)} f_\gamma(\phi) + f_\beta(\phi) \tag{1.14}$$

$$f_\gamma(\phi) = a_\gamma \cos(\phi) + b_\gamma \tag{1.15}$$

$$f_\beta(\phi) = a_\beta \sin(\phi) + b_\beta \tag{1.16}$$

where $a_\gamma, a_\beta, b_\gamma, b_\beta$ are learnable parameters and ϕ is an input parameter for the physics model. The chosen physics model implemented in the paper naively simulate the low-light degradation as a tone-mapping function. The core idea of this method is to learn both modeled features (e.g. the global tone of night images) and non-modeled features (e.g. light sources at night).

Thus, the model takes as input a day image and hallucinates features in the output scene to make a plausible night image. It is very useful as a data augmentation approach for high-level tasks (e.g. classification or detection). However, it was not designed to take as input a night image and therefore cannot restore it. Moreover, it does not preserve the integrity of the scene since it adds non-modeled objects to the scene.

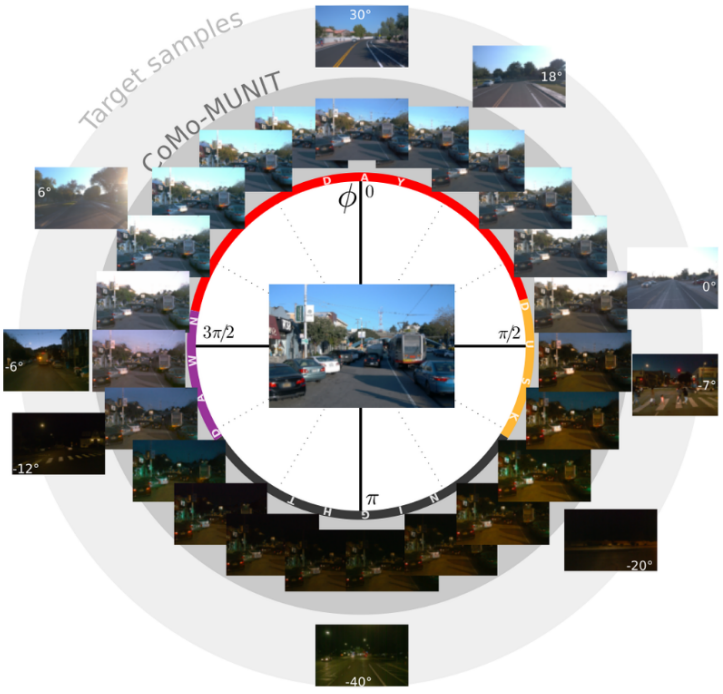


Figure 1.12: Illustration of the style transfer process in CoMoGAN [Pizzati *et al.* 2021].

Pizzati *et al.* improved their previous work for few-shots learning shortly afterwards. They designed the ManiFest framework [Pizzati *et al.* 2022]. It is depicted in Fig. 1.13. The main idea behind is to learn to extract the style from the input image as well as target examples. Then, the network learns to weight the linearly interpolated style representations in the latent space such that the output image belongs to the target manifold. A residual image is also generated if required as a complement to the translated image. It is useful in the case where there is only very few true images from the target domain and synthetic images are available in great quantity. They combine both a patch discriminator and a style loss to impose these constraints. The style loss is computed as a minimization of the error between the extracted mean and variance (i.e. style statistics) of the features from a pretrained VGG network of a true target image and the input image.

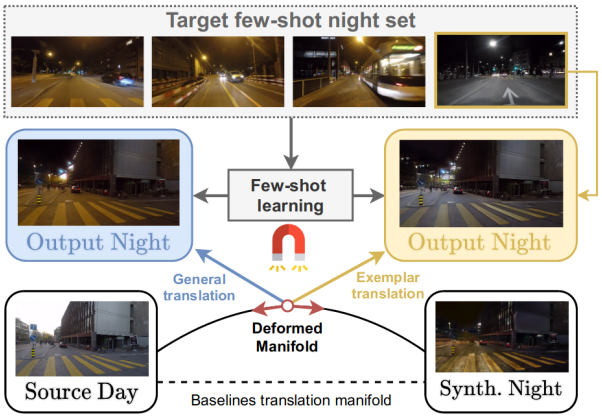


Figure 1.13: Illustration of the ManiFest framework [Pizzati *et al.* 2022].

In our context, the illumination can be seen as a complex style of the image and the reflectance as the content that we want to keep. On the contrary the style transfer methods describe their style as a mean and variance parameter only. We develop that idea in Chapter 4. The main disadvantage of the style transfer methods in our case is that they hallucinate parts of the output image. On the contrary, we want to preserve the integrity of the scene during the restoration.

Regular GANs

Regular GANs have also been explored to restore low-light images. Jiang *et al.* implemented the EnlightenGAN approach [Jiang *et al.* 2021]. They restore underexposed images with a generator, a patch discriminator and a regular discriminator. They regularize the process concatenating an illumination map to the extracted features in the skip

attention modules. This illumination map uses a different definition from the Retinex theory. The authors define it as the negative luma channel $1 - Y'$ in $Y'UV$ color space. That way, the network focuses on the dark parts of the image. This is similar to the part of the literature considering the restoration of low-light images as dehazing the negative low-light image [Galdran *et al.* 2018].

The output of EnlightenGAN [Jiang *et al.* 2021] is a pleasant and plausible image. However, elements, which were not present in the original image may be introduced as it hallucinates. Another drawback is that it requires a careful selection of the training data such as a manual inspection to remove images of medium brightness.

Still, few methods try to decompose an image following the Retinex model in an unsupervised fashion. RetinexGAN [Ma *et al.* 2021] is one of them. They train a GAN-based architecture to extract the Retinex components following the decomposition model (1.2) combined with the LIME prior to estimate the illumination map (1.3). Finally, the output image is computed by reconstructing the extracted reflectance with the negative illumination component.

Deep Image prior & RetinexDIP

In 2018, a new type of general image prior for inverse problems was discovered by Ulyanov *et al.* [Ulyanov *et al.* 2018]. The authors found that an untrained deep convolutional neural network taking a random noise as input can be utilized as a prior to solve optimization problems. Indeed, optimizing through the parameter space of the neural network instead of the image space is a strong prior since the network converges faster on natural images. It might be due to the fact that natural images contain a lot of self-similarity and that the convolution acts like a patch regularization. This Deep Image prior (DIP) results in an alternative way to generative models such as GANs. The main drawback is that one has to retrain the network each time you want to restore an image. Avoiding overfitting the degradation of the image is also a challenge for these methods. For instance, if we consider a noisy image I, a CNN with parameters θ, a random input noise z:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \|I - T_\theta(z)\|_2^2. \tag{1.17}$$

$$T_\theta : z \sim \mathcal{U}(0,1)^{3n} \mapsto T_\theta(z) \in \mathbb{R}^{3n}. \tag{1.18}$$

In that case, the DIP will generate a denoised version of that image or in other words restore it.

A recent work by Zhao et al. dubbed RetinexDIP [Zhao *et al.* 2021] combined the Retinex decomposition with deep image priors to generate the two intrinsic components of the image. Then, they enhance the low-light image by restoring the illumination with a gamma correction. The authors designed an illumination consistency prior to constrain the generated illumination to be close to the LIME prior (1.3), a first plausible guess but still an approximation. This is a hard constraint on the set of reachable solutions and therefore the deep image priors cannot explore as many solutions as they should. Moreover, they still follow the decomposition model (1.2) as shown on Fig. 1.14.

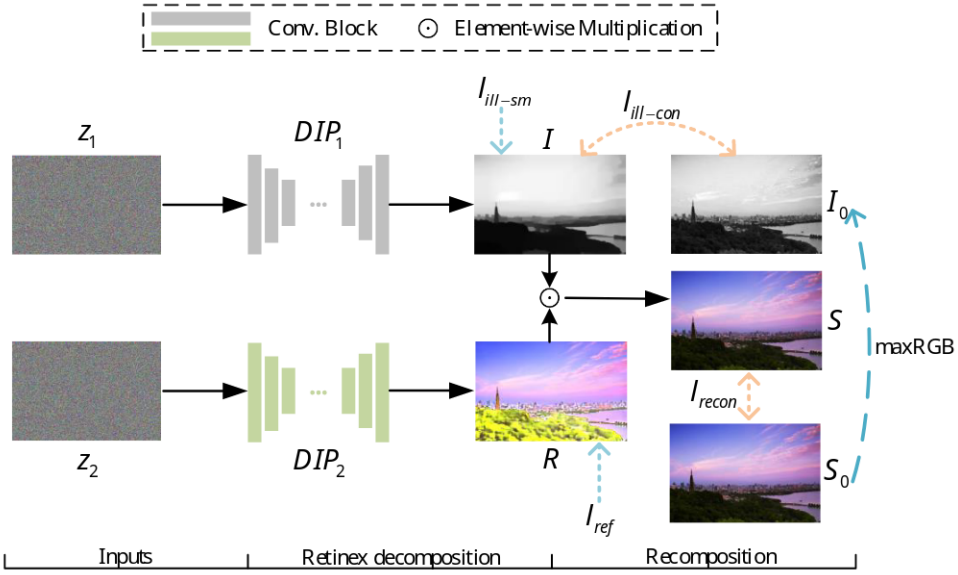


Figure 1.14: Illustration of the RetinexDIP framework [Zhao *et al.* 2021].

1.3 Metrics

In this section, we aim at recalling the metrics in the low-light literature, what they measure and the challenges regarding our context. Image quality assessment methods can generally be classified into two categories: metrics which require the ground-truth clean image and those which do not.

1.3.1 Reference-based metrics

Pixel-wise metrics

We begin by well-known pixel-wise metrics. Let I be the degraded image or the output corrected image, I^* the ground-truth image. To ease the notation, images and vectors are used interchangeably.

A simple measure of the error on the output image can be used such as the Mean Square Error (MSE)

$$\text{MSE}(I, I^*) = \frac{1}{n}\sum_{i=1}^{n}[I_i - I_i^*]^2. \tag{1.19}$$

The biggest problem with the MSE is that it does not differentiate image errors as humans do. Indeed, a slightly noisy image leads to a higher error than a blurred image which is not desirable. It is sensitive to outlier values but favors smooth approximations of the ground-truth image. Therefore, it has been traditionally used in denoising applications.

The Mean Absolute Error (MAE) can be used instead replacing the squared term in (1.19) with its absolute. On the contrary to the MSE, it favors sharper images but it is not differentiable at 0. This makes it harder to numerically optimize. The Huber loss [Huber *et al.* 1964] combines the best of both adding a threshold parameter δ,

$$\mathcal{L}_\delta(I, I^*) = \begin{cases} \frac{1}{2}(I - I^*)^2, & \text{if } |I - I^*| \leq \delta, \\ \delta \cdot (|I - I^*| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases} \tag{1.20}$$

The Peak Signal-to-Noise Ratio (PSNR) measures the quality of an image with respect to the ground-truth as the ratio of powers between the signals. It is defined as a normalized MSE taking into account the maximum possible value d

$$\text{PSNR}(I, I^*) = 10\log_{10}\left(\frac{d^2}{\text{MSE}(I, I^*)}\right) \tag{1.21}$$

$$d = 2^b - 1 \tag{1.22}$$

where b is the number of bits over which the images are encoded (e.g. 255 for 8-bits images). Its unit is the decibel *dB*.

In their work [Wang *et al.* 2004], Wang *et al.* defines the Structural Similarity Index Measure (SSIM). This metric is designed to better correlate with the human visual system to estimate the quality of an image. It considers the luminance, contrast and structure of the images to do so. It boils down to

$$\mathrm{SSIM}(I, I^*) = \frac{(2\mu_I\mu_{I^*} + c_1)(2\sigma_{II^*} + c_2)}{(\mu_I^2 + \mu_{I^*}^2 + c_1)(\sigma_I^2 + \sigma_{I^*}^2 + c_2)} \tag{1.23}$$

$$c_1 = (k_1 d)^2 \tag{1.24}$$

$$c_2 = (k_2 d)^2 \tag{1.25}$$

where (c_1, c_2) are two stabilization variables, d the dynamic range like in (1.22), (μ_I, μ_{I^*}) and (σ_I, σ_{I^*}) the mean and standard deviation of the images, σ_{II^*} the cross-correlation. The default values are $(k_1 = 0.01, k_2 = 0.03)$.

In an attempt to measure the low-light degradations in an image, Wang *et al.* define the Lightness Order Error (LOE) in [Wang *et al.* 2013]. At its core, the metric follows the Retinex decomposition model (1.2) including a smooth grayscale illumination component. Both illumination maps from the input image and the ground-truth are extracted thanks to the LIME prior (1.3) [Guo *et al.* 2017]. However, this time, the smoothness property is introduced by using down-sampling before the metric is computed while in LIME they smooth the illumination by solving an optimization problem imposing a sparsity constraint on the gradient of the illumination. Then, a logical XOR operator is applied to compare each pixel *relative order* from the input image and the reference. It penalizes the difference between the two images. Intuitively, one seeks to minimize this metric such that the *relative order* of each pixel is preserved between the input image and the ground-truth. The LOE metric can be defined as follows,

$$\mathrm{LOE}(I, I^*) = \frac{1}{n}\sum_{x=1}^{n}\sum_{y=1}^{n}\Big(U(Q_I(x), Q_I(y)) \oplus U(Q_{I^*}(x), Q_{I^*}(y))\Big) \tag{1.26}$$

$$U(x, y) = \begin{cases} 1, & \text{if } x \geq y, \\ 0, & \text{otherwise.} \end{cases} \tag{1.27}$$

$$Q_I(x) = \max_{c \in \{R,G,B\}} I(x) \tag{1.28}$$

n being the number of pixels, $\oplus$ the exclusive-or operator, $Q_I(x) \in \mathbb{R}$. the estimation

of the illumination with (1.3). In practice, the heavy calculations are prevented thanks to a preprocessing step in the form of a down-sampling process. The biggest flaw of this metric is that it still consider the strict constraints of the smooth grayscale illumination. This does not hold regarding the physics of light as discussed in Chapter 4.

Perceptual metrics

To reduce the gap between the human assessment of image quality, perceptual metrics were developed. They do not follow a pixel-wise approach to the problem but rather come from the study of neural networks trained for a certain task. In these methods, the authors identify that the features neural networks extract are closely correlated to the human perception of the distortions.

The Learned Perceptual Image Patch Similarity (LPIPS) metric [Zhang *et al.* 2018] is the most famous of them all. It consists of a neural network trained on ImageNet [Deng *et al.* 2009] to classify objects into certain categories (i.e. other tasks lead to the same conclusion according to their studies). Two types of networks are studied: VGG [Simonyan *et al.* 2015] and AlexNet [Krizhevsky *et al.* 2012]. The authors only collect features from the convolution layers of these architectures. Doing so, they discard the "heads" of the networks (i.e. making the decision on which class to assign to the input image). Let f_θ be their "backbone" CNN layers and $f_\theta(I) = \hat{y}_I$ the output feature representations. Then, this metric boils down to

$$\text{LPIPS}(I, I^*) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}^l_{Ihw} - \hat{y}^l_{I^*hw})\|_2^2 \tag{1.29}$$

where l is the index of the layer, (H_l, W_l) the spatial dimensions of the feature maps, w_l a vector scaling each layer according to its importance, $(\hat{y}^l_{Ihw}, \hat{y}^l_{I^*hw})$ the feature stacks after a forward pass on the input images (I, I^*) respectively. To learn w_l, the authors freeze the weights of the networks, add linear layers and train them on TID2013 [Ponomarenko *et al.* 2015], a dataset for evaluation of full-reference image metrics. They call that step "perceptual calibration".

Regarding our problem, no dataset containing true paired degraded/ground-truth images exist. Therefore, we cannot apply these metrics. The LPIPS metric [Zhang *et al.* 2018] has also been used to compare distributions in an attempt to alleviate this requirement

[Huang *et al.* 2018]. It was applied on a large set of samples from each domain with the idea that the expectation of this metric would lead to an unbiased estimator of the perceptual quality of images.

1.3.2 No-reference metrics

Intensity-based metrics

The Signal-to-Noise Ratio (SNR) is the most basic image quality assessment method. It has been defined in several ways in the literature. We focus on the definition without reference:

$$\mathrm{SNR}(I) = \frac{\mu}{\sigma} \tag{1.30}$$

where (μ, σ) are the mean and standard deviation of the input signal. Its simplicity leads to inefficiency in the determination of the quality of degraded images which explains why it is not used in practice especially in the low-light restoration literature.

Hassen *et al.* noticed in their work [Hassen *et al.* 2013] that the Local Phase Coherence (LPC) could be used as an image sharpness measure. They extend their idea and propose the LPC-based Sharpness Index (LPC-SI). In practice, they first use log-Gabor filters at N scales and M orientations. Let c_{ijk} be the complex coefficient at the scale i, the orientation j and spatial location k. If k is the location of a sharp edge, the authors note that the phase of the wavelet coefficients in the spatial neighborhood as well as across the different scales should have a linear relationship. Any blurry effect breaks this behavior and can thus be used to measure the degree of the distortion. This relation is depicted in Fig. 1.15.

The authors find that computing these coefficients at 3 scales and using the phase of the coarse scale neighbors to predict the phase of one point in the finest scale is an effective metric for sharpness. If the phases are perfectly aligned, the LPC-SI is maximized and

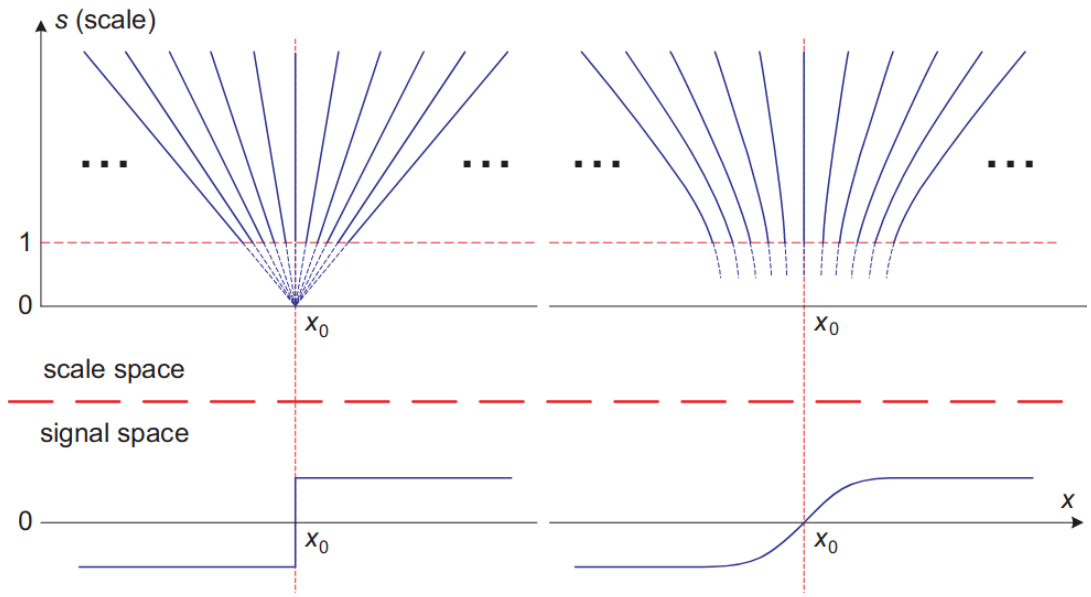


Figure 1.15: Illustration of the phase congruency at the core of the LPC-SI metric [Hassen *et al.* 2013]. In a blurred image, the phase of wavelet coefficients at different scales in the spatial neighborhood of a sharp edge break their linear relationship.

its value is 1. The LPC-SI metric is defined in the paper as follows,

$$S_{\text{LPC}}^{j,k} = \frac{\mathcal{R}\left(\prod_{i=1}^{N} c_{ijk}^{w_i}\right)}{\left|\prod_{i=1}^{N} c_{ijk}^{w_i}\right|} \tag{1.31}$$

$$S_{\text{LPC}}^{k} = \frac{\sum_{j=1}^{M} |c_{1jk}| S_{\text{LPC}}^{j,k}}{\sum_{j=1}^{M} |c_{1jk}| + C} \tag{1.32}$$

where $\mathcal{R}(\cdot)$ is the real part of a complex number, w is a weight vector to ease the computations of the metric, C a constant for the stability of the division. The weight vector w is found by solving a least-square optimization problem. S_{LPC}^{k} is the output spatial LPC-SI map for an image with K pixels. The authors state that since the human visual system is heavily biased on the sharpest features in an image, the metric should reflect this phenomenon. Thus, they define u_k a final weight and β_k the decaying speed of the

weight u_k leading to the final equations

$$S_{\text{LPC}} = \frac{\sum_{k=1}^{K} u_k S_{\text{LPC}}^k}{\sum_{k=1}^{K} u_k} \tag{1.33}$$

$$u_k = \exp\Big[-\frac{1}{\beta_k}\Big(\frac{k-1}{K-1}\Big)\Big]. \tag{1.34}$$

Perceptual metrics

Mittal *et al.* develop a blind (i.e. without reference) image quality assessment method in their work [Mittal *et al.* 2013] called Natural Image Quality Evaluator (NIQE). Natural scene statistics are extracted from a corpus of 125 images and a multivariate Gaussian density is fitted to these features. The score is computed as the distance of between this density and the one obtained from the statistics of the input image. It is defined as follows

$$\text{NIQE}(\mu_{\mathbf{1}}, \mu_{\mathbf{2}}, \mathbf{\Sigma_1}, \mathbf{\Sigma_2}) = \sqrt{\Big((\mu_{\mathbf{1}} - \mu_{\mathbf{2}})^\top \Big(\frac{\mathbf{\Sigma_1} + \mathbf{\Sigma_2}}{2}\Big)^{-1} (\mu_{\mathbf{1}} - \mu_{\mathbf{2}})\Big)} \tag{1.35}$$

where (μ_1, μ_2) are the mean vectors and (Σ_1, Σ_2) the covariance matrices of the corpus model and the one of the input image. In practice however, this metric has poor performance and neural methods achieve a higher correlation with human judgements as studied in the NTIRE 2022 challenge [Gu *et al.* 2022]. The authors conducted their studies on the PIPAL dataset [Gu *et al.* 2020].

The Inception Score (IS) was developed in [Salimans *et al.* 2016]. It uses an Inception model [Szegedy *et al.* 2016] trained to assign the right class to an object among more than 20,000 categories in the ImageNet dataset [Deng *et al.* 2009]. It boils down to

$$\text{IS} = \exp\Big(\mathbb{E}_{I\sim p(I)}[\text{KL}(p(y|I)||p(y))]\Big) \tag{1.36}$$

where y is the label assigned by the network. A higher IS is supposed to mean the generated image is of higher quality. Indeed, the authors seek a generated image which the Inception network can classify with high probability (i.e. $p(y|I)$ should have low entropy) and a generator network which can generate a set of images as diverse as possible (i.e. $p(y)$ should have high entropy).

Huang *et al.* build on the Inception Score and propose the Conditional Inception Score (CIS) [Huang *et al.* 2018]. The previous metric is improved in the case of image-to-image translation. If we consider two domains with label 1 and 2 respectively. Let $I_{1\to 2}$ be the image translated from the first to the second domain thanks to a generative approach. To the contrary of the IS, the CIS measures the diversity of resulting images conditioned on a single input image. This enables it to measure the difference between a model which always generates the same output or a model which really learned the desired output distribution. It is defined as follows

$$\text{CIS} = \mathbb{E}_{I_1\sim p(I_1)}\Big[\mathbb{E}_{I_{1\to 2}\sim p(I_{1\to 2}|I_1)}[\text{KL}(p(y_2|I_{1\to 2})||p(y_2|I_1))]\Big]. \tag{1.37}$$

The Fréchet Inception Distance (FID) [Heusel *et al.* 2017] is a perceptual metric close to IS and CIS. The authors introduce the statistics of the ground-truth images in the computations. The FID metric consists of performing a forward pass with the pretrained Inception model on the set of generated images as well as the one of the reference images. The Inception model is once again a classifier trained on ImageNet. Then, multivariate Gaussians are fitted to the features of the last layer before the classification. The two Gaussians are finally compared with the Fréchet distance [Dowson *et al.* 1982] (i.e. the 2-Wasserstein distance between multivariate Gaussians) which leads to this final formula

$$\text{FID}(\mathcal{N}(\mu_1,\Sigma_1),\mathcal{N}(\mu_2,\Sigma_2)) = \|\mu_1-\mu_2\|_2^2 + \text{Tr}\left(\Sigma_1+\Sigma_2-2(\Sigma_1\Sigma_2)^{\frac{1}{2}}\right). \tag{1.38}$$

The last layer of the Inception model is assumed to contain meaningful and complementary characteristics to the pixel level ones due to higher-level latent representations.

Regarding our problem, metrics based on pretrained networks raise a number of issues in terms of bias. Indeed, as the networks have only been trained on day images, the learned feature representations are not independent of the illumination of the scene which will result in a biased metric. This means these metrics are only relevant in the case where the final goal is to approximate the daylight distribution of images. We only follow this objective in Chapter 5 as the optimal transport framework guarantees to preserve the information in the image. On the contrary, retraining perceptual metrics on night images is also tricky. We can only assess the classes of objects in images where the degradations are not significant.

This lack of suitable metric prevented us to explore in-depth quantitative studies in the first chapters. We focused on one property that we sought in the restored images that is to say the sharpness. However, this can not be the only way of measuring the quality of the restoration process since it does not take into account all the low-light degradations. Therefore, it also drove us to heavily employ qualitative comparisons with the state-of-the-art methods. The degradations are easily noticeable in the images.

1.4 Chapter conclusion

To conclude this chapter, the first drawback observed in approaches following at the same time the Retinex interpretation in (1.2) and the prior in (1.3) is that they only explore the direct neighborhood of this approximation of the illumination. LIME [Guo *et al.* 2017], RetinexGAN [Ma *et al.* 2021] and our first work RetinexDIP [Lecert *et al.* 2022] described in Chapter 3 all fall into this category. These iterative methods are highly dependent on their initialization. A second drawback is that in all the Retinex-based methods, the prior of a smooth grayscale illumination is used in order to reduce the number of degrees of freedom and facilitate the search for solutions to this inverse problem. We address these issues in Chapter 4. We propose diverse improvements to the Retinex decomposition model following the physics of light.

To get a better understanding of the Retinex theory and the numerous interpretations of it, we summarize its history in Chapter 2. We recall the original perceptual interpretation made by Land *et al.* . Then, we focus on Horn's interpretation since he considered a physical model which is more suitable in the context of computer vision. We also review some priors defined in the literature and explain the choices we made regarding the settings of the model. Indeed, since it is considerably hard to estimate the geometry of the scene in a low-light image, we cannot consider it in our context. Besides, we highlight the validity of the decomposition model with respect to the image processing pipeline. Finally, we also link the Retinex theory to the Color constancy problem.

Chapter 2

Background on the original Retinex model

2.1 Perceptual interpretation 36
2.2 Physics of light 39
2.2.1 Horn's interpretation 39
2.2.2 Reversing the camera image processing pipeline 41
2.2.3 Color constancy 41
2.3 Chapter conclusion 42

In this chapter, we review the history of the Retinex theory, the first hypotheses that lead to it, how it differs in the context of computer vision and the physics of light. We sum up some of the main intuitions of this theory regarding its perceptual interpretation in the following sections but we then focus on the aspect of interest in the context of computer vision.

Figure 2.1: Illustration of a typical Mondrian-like array of rectangular color shapes used in Land's and McCann's color experiments. Source: Provenzi *et al.* [Provenzi *et al.* 2017]

2.1 Perceptual interpretation

The Retinex theory goes back to the fundamental work by Edwin H. Land [Land *et al.* 1964; Land *et al.* 1977] as a framework to study the human visual system. It is a portmanteau term to represent the ensemble of biological mechanisms that convert light flux (i.e. input signal in the system) into the lightness sensation interpreted by the cortex. Edwin and his collaborator John J. McCann conducted a series of experiments taking place in a Mondrian-like world similar to the representation in Fig. 2.1 to explain the different properties of the human vision [Land *et al.* 1971]. Thus, the end goal of this collection of analyses was to define a model subject to optical illusions such as Fig. 2.2.

As a result from their experiments, they found that there is no predictable connection between the intensity of light rays at various wavelengths and the color sensations associated with the scene. The perceived colors only depend on the wavelength and not the intensity as long as the cone cells are used (photopic vision i.e. daylight vision). This is no longer valid in the scotopic vision (i.e. night vision). In very low-light, the rods

process the light instead. The authors concluded that the cortex must discard the illumination and that it is able to find an independent biologic equivalent correlated with the reflectance. Therefore, the cortex has to be involved in this system to estimate the colors under different conditions. Depending on the set of pigments absorbing the light ray (i.e. either at long, middle or short wavelength in the visible spectrum), an "image" of the scene is formed. The three images are compared by the cortex to produce the final color sensation.

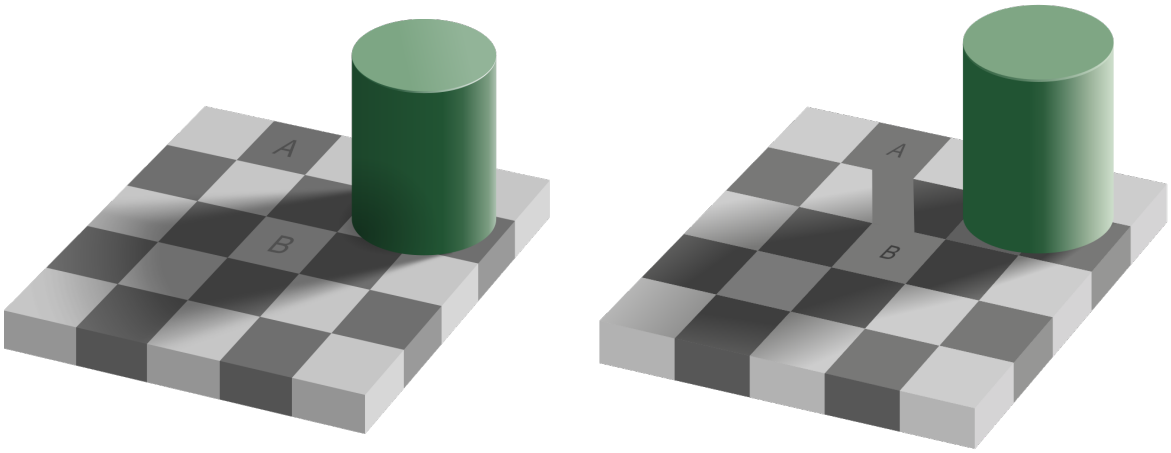


Figure 2.2: The well-known checker illusion. A and B are the same shade. Source: *Wikipedia*.

Moreover, Edwin and McCann proved that the color perceived by a human depends on its surroundings. This spatial dependency is the core idea of this first perceptual computational model of the Retinex theory.

After the color matching experiments, Land and McCann defined a first path-based algorithm [Land *et al.* 1971] to obtain the reflectance. It processes each RGB channel individually to mimic the three types of cones (i.e. long, middle and shortwaves). For each channel, it compares the intensity of the pixels over piecewise constant paths to take into account the spatial locality property of the Retinex theory. It performs a multiplicative chain of ratios like in Fig. 2.3 and then applies different non-linear operations such as a threshold and a reset mechanism.

A variety of algorithms were conceived on spatial locality to recover the human color perception in different situations. They can be categorized in path-based, center/surround

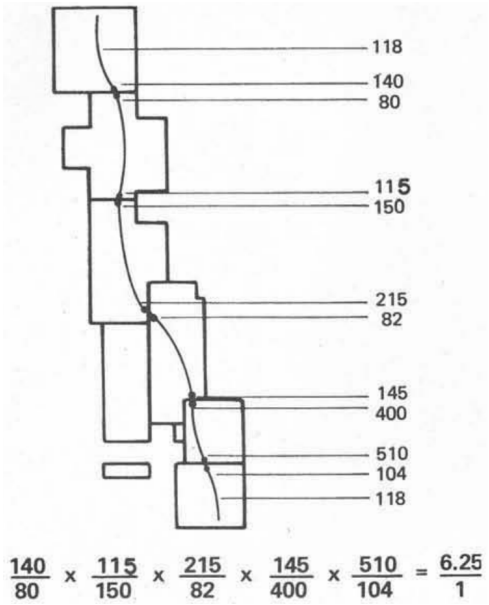


Figure 2.3: Example of the multiplicative chain of ratios operation performed by the original Retinex algorithm [Land *et al.* 1971].

and finally variational algorithms. We refer the reader to [Provenzi *et al.* 2017] for a more in-depth review of methods based on the perceptual interpretation. Throughout its history illustrated in Fig. 2.4, the Retinex theory has been interpreted in many different ways but the algorithms that have resulted can all be categorized as contrast enhancement methods. Land and McCann have listed a number of phenomenons a model has to follow to mimic the human visual system. In the end, however, they have not defined a computational model that we can use in computer vision and signal processing algorithms.

Horn rectified this problem as described in the next section. He proposes a different interpretation of this theory that includes the properties of light. The state-of-the-art approaches for the restoration of low-light images is built around this interpretation. Since we do not seek to mimic the human vision but study the Retinex theory in the context of computer vision, we will focus on it in the rest of the chapter.

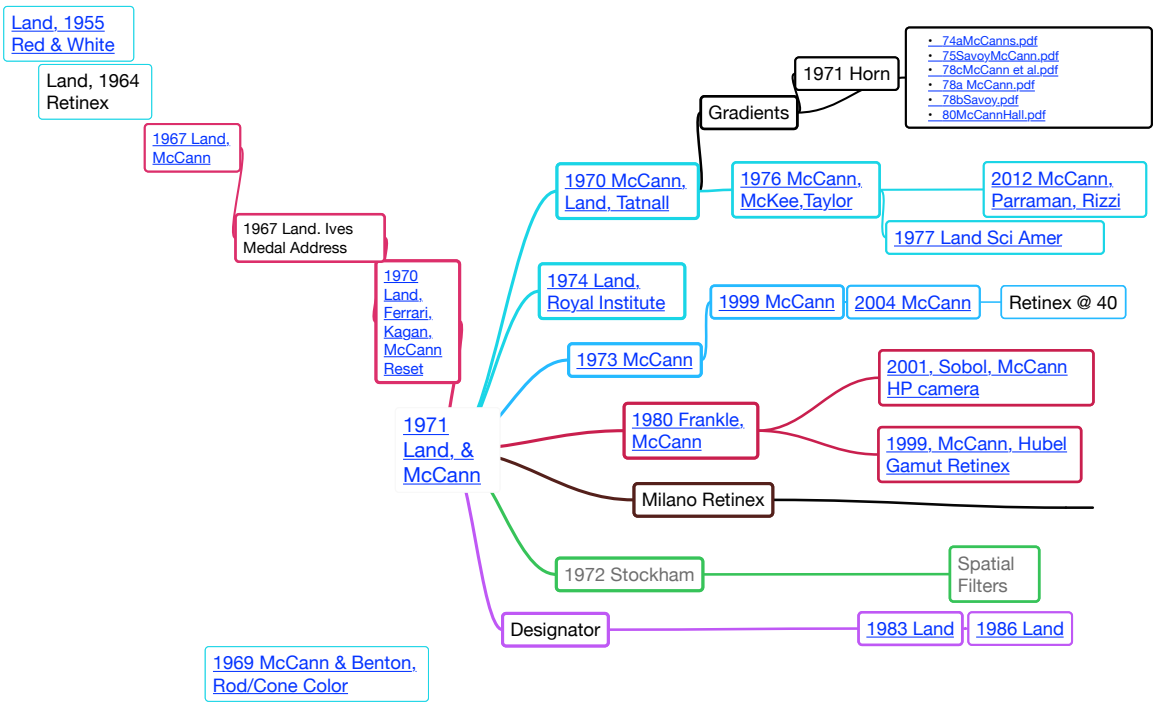


Figure 2.4: Timeline of the history of the Retinex theory and its interpretations [McCann *et al.* 2017].

2.2 Physics of light

2.2.1 Horn's interpretation

In 1974, Horn proposed a novel interpretation of the Retinex theory [Horn *et al.* 1974]. This interpretation can be divided in several contributions. The major one consists of a physical model which claims that the image intensity is the product of the reflectance and the illumination. At the time, his goal was also to model the human eye mechanism to adapt to different lightning conditions. Consequently, Horn defines in his paper an algorithm based on this decomposition model to determine the "lightness" (i.e. color perception) following the Land's Retinex intuition [Land *et al.* 1971] by discarding the illumination to recover the reflectance. Besides, he was still working in the Mondrian-like world where the light only varies smoothly. In this way, shadows and specular reflections are disregarded.

We are mainly interested in this first contribution of a computational model. Horn

stated that an image is a product of two components: illumination and reflectance, where the illumination contains the lighting dependent information and the reflectance the true color of the objects in the scene. To restore an image with this model, approaches in the literature first decompose the input into these two components. Then, the restored image is either considered as the reflectance (i.e. such as the Land's perceptual interpretation of the Retinex theory), or as the product between a gamma-corrected illumination and the estimated reflectance (i.e. following Horn's physical model). The Horn-Retinex model is still widely used by the state-of-the-art in computer vision and signal processing even today (see Chapter 1).

Interestingly, in Horn's paper and in those which quickly followed mimicking the human vision system, the idea of a colored illumination was widespread (i.e. for instance in the work [Kimmel *et al.* 2003]). Indeed, Horn's algorithm processes three color channels separately to estimate the reflectance corresponding to long, middle and short wavelength bands (i.e. the RGB channels). Therefore, it implicitly considers a colored illumination though it is not clearly stated. We could not find this colored property of the illumination in the computer vision literature after that. The idea vanished at some point. This might be due to the difficulty of the decomposition. Considering a grayscale illumination, the problem though still underdetermined is much easier to solve (i.e. we need to estimate $4n$ variables from $3n$ observations instead of $6n$ variables without this prior).

In a first attempt from the computer vision community to extract different components from an image, Barrow *et al.* [Barrow *et al.* 1978] proposed to expand the Horn-Retinex model to extract what they called "intrinsic images". They make a variety of assumptions on the geometry of the scene to recover the surface reflectance, the depth of the scene, its surface normals and the illumination. Thus, they seek three-dimensional information out of a single monochrome intensity image. All the properties of a scene point are thus encoded in a single intensity value at a specific wavelength.

Diverse geometrical computational models have been designed to estimate intrinsic components of an image such as the work of Barron *et al.* [Barron *et al.* 2015b] or in the inverse rendering community [Lombardi *et al.* 2016; Yu *et al.* 2022]. Estimating the geometry of the scene is already challenging in day images. Since we only have access to a single night image as input, we have to make more restrictive assumptions. The

three-dimensional information is discarded. Following Barrow *et al.* [Barrow *et al.* 1978], we consider that the image is formed by central projection onto a planar surface (i.e. the sensor of the camera). The reflectance is assumed to be Lambertian (i.e. the surface at every point in the scene is perfectly diffusely reflecting light rays on the hemisphere). The incident angle of the incident light can be ignored and the Bidirectional reflectance distribution function (BRDF) [Nicodemus *et al.* 1965] is not used. Besides, any specular or ambient component is also neglected which invalidate the use of models such as Phong's model of the reflection of the illumination on a surface [Phong *et al.* 1975]. Local reflection between the objects in the scene is assumed also to be negligible.

As an additional remark, we would like to emphasize that this physical computational model of the product between the reflectance and the illumination is also valid in hyperspectral images. Indeed, it does not depend on the chosen number of bands. We only consider the visible spectrum with RGB images in this thesis.

2.2.2 Reversing the camera image processing pipeline

In all the interpretations of the Retinex theory, the image intensity I is the irradiance at the surface of the camera sensor. It should not be confused with the pixel value of the image. Indeed, each camera applies a specific pipeline to compute the resulting standard RGB image. This pipeline contains non-linear operations (e.g. gamma corrections or tone mappings). To accurately estimate the reflectance and the illumination, they have to be reversed. However, defining a sensor-agnostic method to invert these operations is very challenging in practice. It is still a very active field of research [Liu *et al.* 2020; Ying *et al.* 2017].

In the next section, we want to show the relation between estimating the reflectance (in both Land's and Horn's interpretation) and the problem known as Color constancy.

2.2.3 Color constancy

The color constancy problem consists of finding the best RGB triplet (i.e. the "illuminant") to correct an improperly white-balanced camera image [Barron *et al.* 2017; Afifi *et al.* 2019]. It can be modeled with a decomposition model similar to the one of

Horn-Retinex:

$$I = l * R \tag{2.1}$$

where $I \in \mathbb{R}^{3n}$ is the degraded image, $R \in \mathbb{R}^{3n}$ is the reflectance, $l \in \mathbb{R}^3$ is the illuminant. As stated in [Barron *et al.* 2015a], in the color constancy problem, the absolute scaling of l is not important and is commonly normalized to simplify the computations.

The connection with the Retinex decomposition problem is thus trivial. This illuminant can be regarded as a globally constant colored illumination (i.e. ambient light) on the scene. Extracting this illuminant and thus discarding the illumination, one can recover the reflectance like in the perceptual interpretation of the Retinex theory. This reflectance can be seen as the "true" colors of the scene (i.e. the colors under an ambient white light). Therefore, by doing so, the chrominance of the global illumination is computed but not the luminance (i.e. the intensity of this illumination). In the Retinex decomposition problem, we seek to estimate both the chrominance and luminance of the illumination but it also has to be a local illumination.

2.3 Chapter conclusion

In this chapter, we went back to the first interpretations of the Retinex theory. Land and McCann defined the properties of human vision in a series of experiments. They also proposed algorithms to mimic these results with a computer. However, they have not provided a computational model that we can use in our methods. On the contrary, Horn built a much more suitable model for computer vision algorithms. It considers the principles of the physics of light.

We stressed that in our context, the geometry of the scene can hardly be integrated in the Retinex model. However, this decomposition can only be applied on raw images before the nonlinear operations of the image processing pipeline. This property has been ignored in the recent works of the literature. The Retinex theory is also closely related to the problem of color constancy. It corresponds to Land's perceptual interpretation with a perfect Retinex decomposition integrating a colored illumination.

Having a better understanding of the Retinex theory and its different interpretations,

we summarize our first contribution in the next chapter. We define a new prior combined with a DIP-based generative method to obtain the Horn-Retinex components and restore them in Chapter 3.

In Chapter 4, we question the model and the priors found in the recent literature. We propose some improvements to the Horn-Retinex decomposition model and define a GAN-based architecture to leverage it.

Chapter 3

A new regularization for Retinex decomposition

3.1 Initial problem statement 46
3.2 The proposed prior 47
3.2.1 The trivial solution and the exposure prior 47
3.2.2 New problem formulation 48
3.3 Experiments 48
3.3.1 Restoration of the components 48
3.3.2 Implementation details 49
3.3.3 Quantitative comparison 49
3.3.4 Qualitative results 51
3.3.5 Ablation and hyperparameter study 52
3.4 Chapter conclusion 53

Due to the scarcity of low/normal-light image pairs in some applications like satellite imaging described in the introduction, we choose to begin by studying an unsupervised method that does not require any dataset to be trained on apart from the single image input. We aim to extract the Retinex components of the input image with two coupled deep image priors. Processing the components in these sub-spaces is assumed to be more efficient than restoring the whole image. Since the reflectance and the illumination components have specific characteristics, we propose to restore each of them differently.

In this chapter, we summarize the contributions of our work [Lecert *et al.* 2022]. They are manifold. First, we rewrite the Retinex decomposition as a scale mixture which is a widespread and well-studied model in the image processing literature especially in the wavelet domain with Gaussian priors [Wainwright *et al.* 2000; Wainwright *et al.* 2001; Schwartz *et al.* 2004] in Section 3.1. In the context of low light images, we identify a trivial solution (i.e. when the scaling factor is equal to one), and analyze its properties. Then, we propose a new prior that addresses this problem while still letting the deep image priors explore as many plausible solutions as before in Section 3.2. In Section 3.3, we propose to restore the image with two gamma corrections, one for each component. The fact that component-specific corrections are used, shows the necessity to decompose the low-light image. We demonstrate the effectiveness of our decomposition achieving good performance using simple gamma corrections. Finally, we observe no more halo artifacts on the restored image. For all-but-one metrics, our approach gives results as good as the supervised state-of-the-art indicating the potential of a generative framework for low-light image enhancement.

3.1 Initial problem statement

We follow the classic interpretation of the Retinex theory found in the literature. It is similar to the one in the one in Section 1.2.1 with a smooth grayscale illumination component. In this interpretation, the illumination map should not contain any texture details but still keep the structure of the scene. Thus, we first use the structure-aware illumination smoothness prior in [Wei *et al.* 2018; Zhang *et al.* 2019],

$$\mathcal{L}_{\text{IS}} = \left\| \frac{\nabla L}{\max(\nabla I, \epsilon)} \right\|_1. \tag{3.1}$$

Using deep image priors to generate the components, we can define the following optimization problem,

$$(\hat{\theta}_L, \hat{\theta}_R) = \underset{\theta_L \in \mathbb{R}^p, \theta_R \in \mathbb{R}^{p'}}{\operatorname{argmin}} \|I - T_{\theta_L}(z_L) . * T_{\theta_R}(z_R)\|_2^2. \tag{3.2}$$

$$T_{\theta_R} : z_R \sim \mathcal{U}(0,1)^{3n} \mapsto T_{\theta_R}(z_R) \in \mathbb{R}^{3n} \tag{3.3}$$

$$T_{\theta_L} : z_L \sim \mathcal{U}(0,1)^{n} \mapsto T_{\theta_L}(z_L) \in \mathbb{R}^{n} \tag{3.4}$$

where (z_R, z_L) are the input noises, and $(T_{\theta_R}(z_R), T_{\theta_L}(z_L))$ their respective outputs.

Since the illumination has to only contain low frequencies, high frequencies of the image including noise end up in the reflectance. A core property of the deep image prior is to be robust to noise and to converge faster on naturally looking images [Ulyanov *et al.* 2018]. To further reduce the noise, we add a TV penalty $\rho_{\mathrm{TV}}(R)$ to the problem following the work in [Liu *et al.* 2019].

3.2 The proposed prior

3.2.1 The trivial solution and the exposure prior

$L_i = 1$, and thus $R_i = I_i$, should only be admissible when there is a light source or an overexposed area in I at pixel i. Indeed, the illumination can be smoothed out by the prior, the reconstruction of the low-light image still be correct and yet the problem could occur. In RetinexDIP [Zhao *et al.* 2021], the authors use an illumination consistency prior which ties the component to the maximum of the low-light image over the color channels. This constrains the component to be close to a first plausible guess. The authors then try to find the best decomposition in the direct neighborhood of the approximation. This reduces the possibility to improve the process further than the initial guess. We propose a new regularization in order to only accept the trivial solution when it is feasible. We define the exposure prior as follows,

$$\mathcal{L}_{\mathrm{E}} = \left\| g\Big(\max_{c \in \{R,G,B\}} I_c \Big) - g\Big(T_{\theta_L}(z_L) \Big) \right\|_2^2 \tag{3.5}$$

where g is a threshold function $g(x) = \begin{cases} x, & x > t \\ 0, & \text{otherwise} \end{cases}$. We choose $t = 0.9$ here so that this constraint only affects the high values of the component. Therefore, the solution set includes the ones with $L_i = 1$ when $I_i = 1$ but forbids the illumination to be equal to one if there is no light source or overexposed regions in the input low-light image.

3.2.2 New problem formulation

Instead of minimizing over the parameter space of a DIP, the authors in [Sagel *et al.* 2020] proposed to relax this constraint to improve the performance expanding the solution space. We seek the best compromise between the data fidelity term and the DIPs outputs reaching potentially better solutions. Therefore, we define the additional terms $\|T_{\theta_L}(z_L) - L\|_2^2$ and $\|T_{\theta_R}(z_R) - R\|_2^2$ to keep the estimated components close to the outputs of the DIPs. Each generated component is able to drift from its respective DIP solution. Consequently, in the SUB-DIP formulation [Sagel *et al.* 2020], the structure of the CNN is really considered as a prior and not as a hard constraint unlike DIP [Ulyanov *et al.* 2018], DoubleDIP [Gandelsman *et al.* 2019] or RetinexDIP [Zhao *et al.* 2021]. This is an additional difference between our work and RetinexDIP [Zhao *et al.* 2021]. We now seek to estimate the best parameters of the DIPs $(\hat{\theta}_L, \hat{\theta}_R)$ as well. Hence, the optimization problem becomes

$$\begin{aligned}(\hat{L}, \hat{R}, \hat{\theta}_L, \hat{\theta}_R) = \underset{L,R,\theta_L,\theta_R}{\operatorname{argmin}} \; & \|I - L. * R\|_2^2 + \lambda_{\text{IS}}\mathcal{L}_{\text{IS}} \\ & + \lambda_{\text{DIP}_R}\|T_{\theta_R}(z_R) - R\|_2^2 + \lambda_{\text{DIP}_L}\|T_{\theta_L}(z_L) - L\|_2^2 \\ & + \lambda_{\text{TV}}\rho_{\text{TV}}(R) + \lambda_{\text{E}}\mathcal{L}_{\text{E}}. \end{aligned} \tag{3.6}$$

We initialize the components R and L as proposed in LIME [Guo *et al.* 2017], i.e., as $L = \max_{c \in \{R,G,B\}} I_c$, $R = I./L$. The complete decomposition process is illustrated in Fig. 3.1.

3.3 Experiments

3.3.1 Restoration of the components

To restore the components, we take a random subset of 30 paired images from the training set, decompose them with our method and estimate the gamma values. We found that we get better results using a different value for each component. Thus, we use two unique

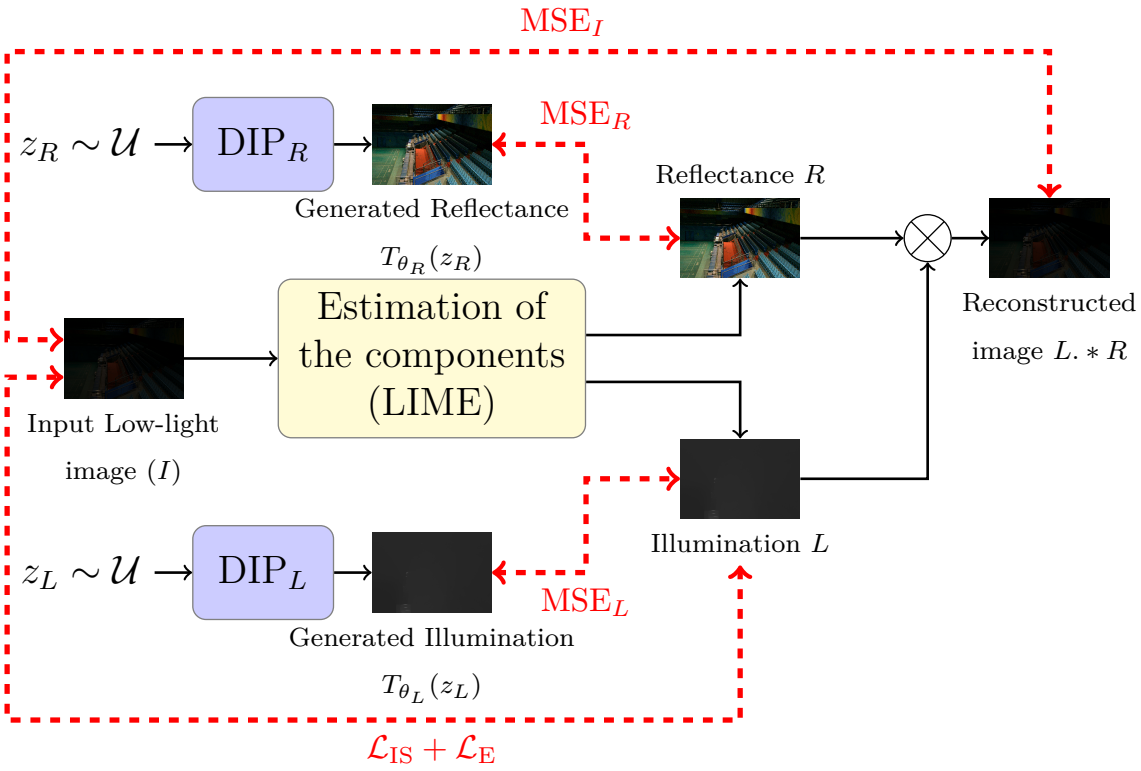


Figure 3.1: Retinex decomposition scheme: The DIPs $(T_{\theta_R}, T_{\theta_L})$, initialized with random noises (z_R, z_L), produce two initial estimate $(T_{\theta_R}(z_R), T_{\theta_L}(z_L))$. They are defined in Eq. (3.2)-(3.4). These are further improved with the general loss ((3.6)) to produce the final components $\hat{R}$ and $\hat{L}$.

values for all images. This demonstrates the necessity of the Retinex decomposition.

3.3.2 Implementation details

We use the ADAM optimizer [Kingma *et al.* 2015] with a fixed learning rate of $1e^{-4}$ and 12000 optimization steps, Pytorch [Paszke *et al.* 2019] as framework and the Kornia library [Riba *et al.* 2019]. We empirically find the coefficients $\lambda_{\text{IS}} = 1e^{-4}, \lambda_{\text{E}} = 1e^{2}, \lambda_{\text{TV}} = 1e^{-10}, \lambda_{\text{DIP}_R} = 1e^{-2}, \lambda_{\text{DIP}_L} = 1e^{-1}, \gamma_R = 0.4, \gamma_L = 0.2$.

3.3.3 Quantitative comparison

We evaluate our method on the test set of the LOL dataset [Wei *et al.* 2018] composed of 500 low/normal-light image pairs taken from real scenes by changing exposure time and ISO.

We adopt the following metrics to evaluate the performance of our approach: PSNR, SSIM [Wang *et al.* 2004], LPIPS [Zhang *et al.* 2018], NIQE [Mittal *et al.* 2013], CPCQI [Gu *et al.* 2018] and NIQMC [Gu *et al.* 2017]. Therefore, we hope to measure the whole phenomenon of the low light degradation thanks to pixel-wise, classic and learned perceptual metrics.

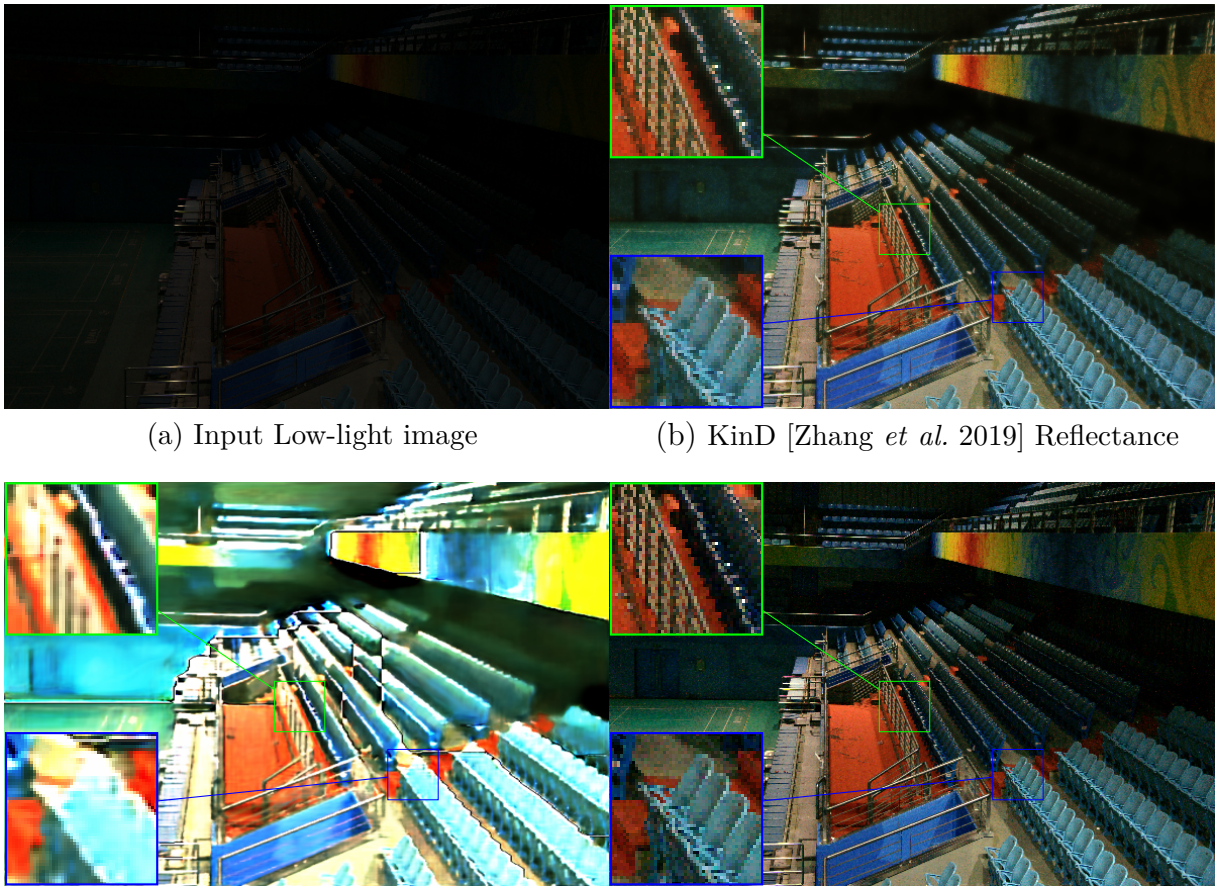
The chosen state-of-the-art competitors are KinD [Zhang *et al.* 2019], LIME [Guo *et al.* 2017], EnlightenGAN [Jiang *et al.* 2021], Zero-DCE [Li *et al.* 2020] and RetinexDIP [Zhao *et al.* 2021]. The first one is a well-known completely supervised network trained on the LOL dataset [Wei *et al.* 2018]. The followings methods are unsupervised, either traditional or trained with unpaired data and the latter uses deep image priors in a similar framework. Table 3.1 summarizes the results. Our method outperforms the unsupervised approaches on most of the metrics while being close to KinD [Zhang *et al.* 2019].

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	Runtime (in s)
Supervised				
KinD [Zhang *et al.* 2019]	17.26	**0.77**	**0.187**	1.47
KinD [Zhang *et al.* 2019] DecompNet → γCorr	17.91	0.64	0.321	1.06
Unsupervised				
LIME [Guo *et al.* 2017]	10.10	0.38	0.383	0.26
EnlightenGAN [Jiang *et al.* 2021]	17.48	0.65	0.322	0.12
Zero-DCE [Li *et al.* 2020]	14.86	0.56	0.335	**0.0012**
RetinexDIP [Zhao *et al.* 2021]	11.69	0.48	0.351	20.89
Ours → γCorr	**18.11**	0.68	0.306	2220

Table 3.1: Best and second-best results are highlighted in bold, and blue respectively. KinD [Zhang *et al.* 2019] relies on paired degraded/ground-truth images to extract its priors. LIME [Guo *et al.* 2017] is a signal-prior based approach. Zero-DCE [Li *et al.* 2020] and EnlightenGAN [Jiang *et al.* 2021] are unsupervised methods but still needs to be trained on a dataset of unpaired multi-exposure images. On the contrary, RetinexDIP [Zhao *et al.* 2021] and ours are *fully unsupervised*. The latter outperforms the unsupervised competitors while being close to the supervised one. Since we build on the deep image prior [Ulyanov *et al.* 2018], it has the same drawback being the high computation time. The difference between RetinexDIP [Zhao *et al.* 2021] and ours is due to the different number of epochs as discussed in Section 3.3.5.

Since the only available dataset of ground-truth reflectance and illumination [Grosse *et al.* 2009] is only composed of 16 images, we compare our components against those of the Decomposition Net of KinD [Zhang *et al.* 2019] and RetinexDIP [Zhao *et al.* 2021] in Table 3.1

and in Fig. 3.2.



(a) Input Low-light image (b) KinD [Zhang *et al.* 2019] Reflectance

(c) RetinexDIP [Zhao *et al.* 2021] (d) Ours

Figure 3.2: Our *fully unsupervised* approach achieves on par intrinsic components with the KinD [Zhang *et al.* 2019] network trained in an end-to-end fashion on the same dataset. Even with 12000 iterations, RetinexDIP [Zhao *et al.* 2021] reflectance is cartoon-like.

3.3.4 Qualitative results

We visually compare the different components generated by the solutions in Fig. 3.2. We obtain visually pleasing components with our approach close to the supervised competitor, KinD [Zhang *et al.* 2019]. On the contrary, although we use 12000 iterations to get better components with RetinexDIP [Zhao *et al.* 2021], the reflectance is still cartoon-like and contains artifacts.

We add Figs. 3.4 to 3.6 which respectively illustrate additional examples of the extracted components, restored images and restored components. All these examples come

from the LOL dataset. In Fig. 3.4, we can see that the model has trouble extracting the reflectance from the gray objects in the pictures in the second row. The color ends up in the illumination while it boosts the noise in the reflectance. The darker the color is, the lower the signal-noise ratio will be, as expected.

3.3.5 Ablation and hyperparameter study

To compare the efficiency of the priors, we implement the illumination consistency prior of RetinexDIP [Zhao *et al.* 2021] in our framework for a fairer ablation study. We reduce the weight of the illumination smoothness as it can alleviate the problem without solving it. Consequently, the visual quality of the components are subpar. The results are shown in Fig. 3.3 and Table 3.2. With our prior, our approach achieves better scores and leads to a better decomposition without artifacts.

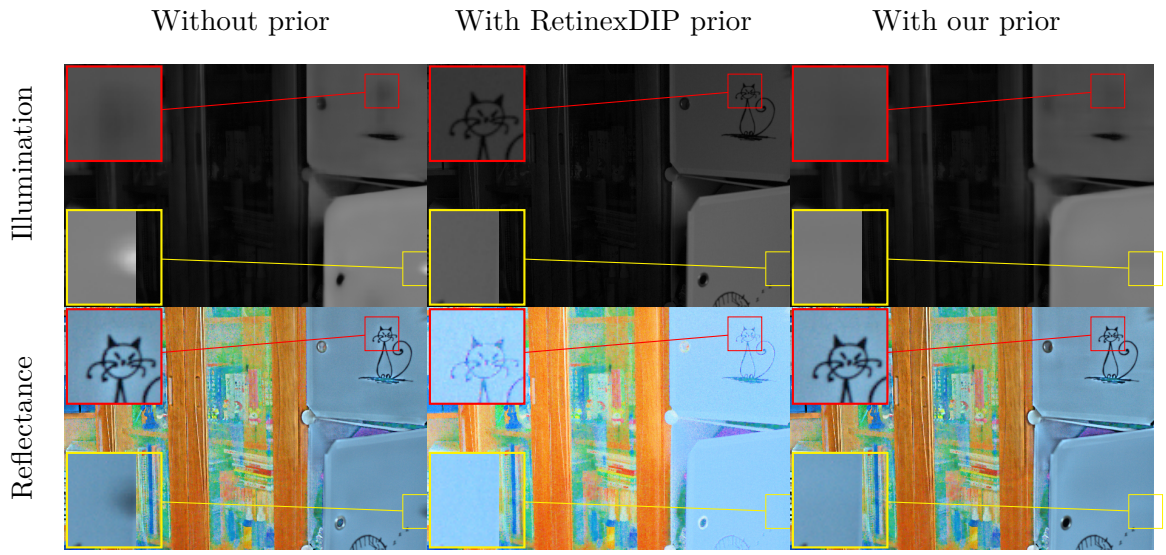


Figure 3.3: Without any prior, the components can contain artifacts (yellow square). Although RetinexDIP prior solves this issue, the textural details are still in the illumination after the decomposition (red square). Our prior gets the best of both worlds.

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	NIQMC(↑)	CPCQI(↑)	NIQE(↓)
Without prior	17.36	**0.68**	0.311	3.781431	0.798175	5.931700
RetinexDIP prior	**17.72**	0.67	0.325	**3.932083**	0.793341	6.173116
Our prior	17.44	**0.68**	**0.306**	3.824624	**0.813172**	**5.836579**

Table 3.2: Our prior achieves the best scores for most of the metrics.

Since RetinexDIP [Zhao *et al.* 2021] uses only 300 iterations as default compared to the 12000 iterations of our method, we analyze both methods in Table 3.3 by changing this parameter. We also include the results when applying gamma corrections on the components of RetinexDIP [Zhao *et al.* 2021]. As shown in Table 3.3, our approach achieves good performance even if we reduce the number of iterations to 300.

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	NIQMC(↑)	CPCQI(↑)	NIQE(↓)
RetinexDIP 300 iter [Zhao *et al.* 2021]	11.69	0.48	0.351	3.718755	1.209930	8.035102
RetinexDIP 12000 iter [Zhao *et al.* 2021]	11.95	0.49	0.354	3.648840	**1.216181**	8.132518
RetinexDIP 12000 iter [Zhao *et al.* 2021] → γCorr	17.46	**0.69**	0.380	3.951571	0.461452	**4.536125**
Ours 300 iter → γCorr	16.14	0.59	0.397	3.332199	0.624579	7.168487
Ours 12000 iter → γCorr	**18.11**	0.68	**0.306**	4.207515	0.915857	5.882059
Ours 12000 iter → γ_LCorr only	14.89	0.55	0.331	**4.331838**	1.147143	5.890018

Table 3.3: Even with the same restoration process and number of iterations, our approach gives the best scores for most of the metrics.

3.4 Chapter conclusion

In this work, we identified a trivial solution problem and proposed a new regularization term to fix it. We have demonstrated its efficiency in an in-depth ablation study. Our framework achieves visually pleasing intrinsic components on par with state-of-the-art supervised methods and outperforms the unsupervised competitors.

To further improve the restoration process, we propose in the next chapter diverse relaxations and improvements to the Retinex model. We highlight widespread limitations of the existing model with real life examples. Indeed, it does not follow the principles of the physics of light.

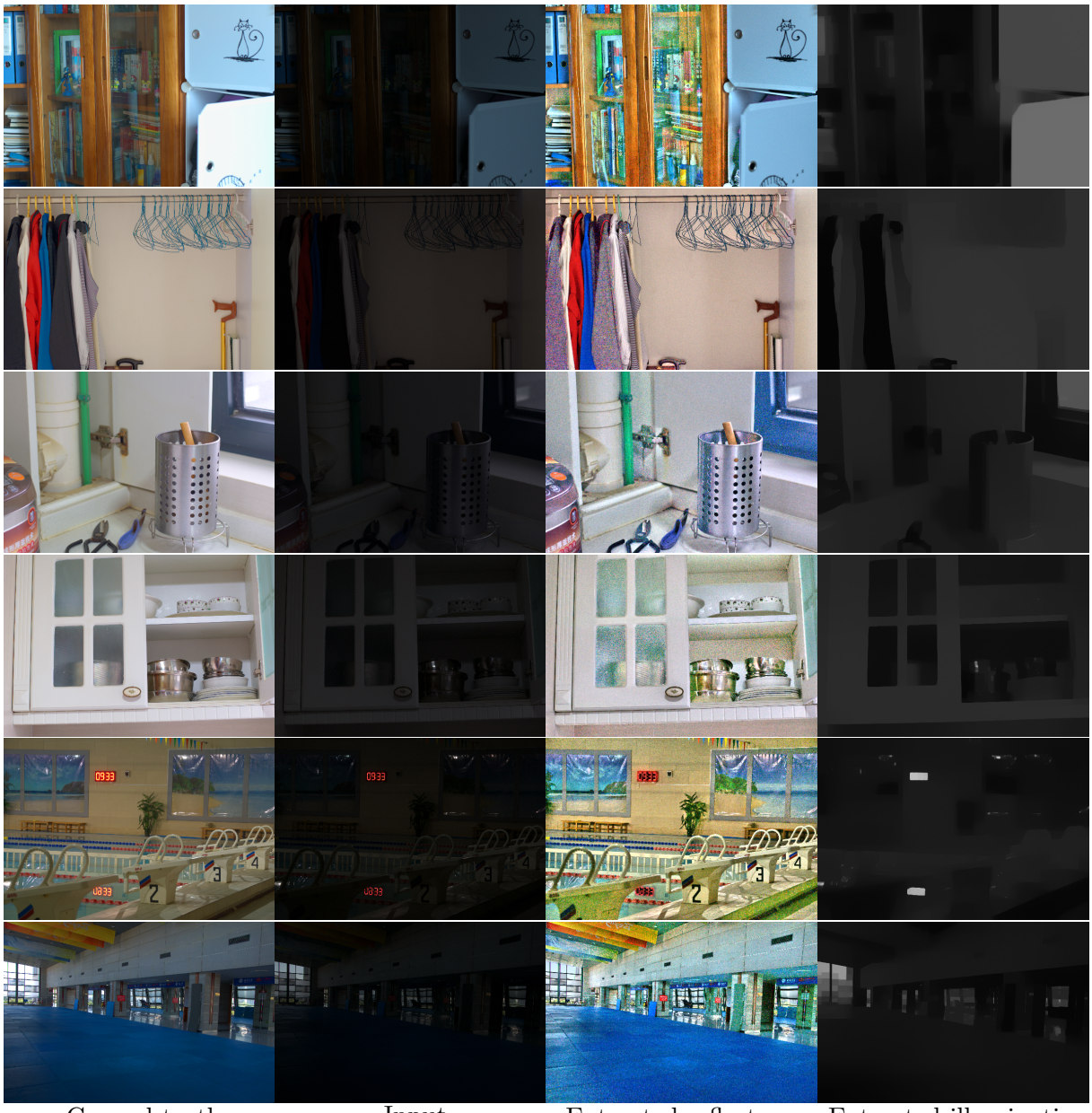


Figure 3.4: Additional examples of the Retinex components we obtained.

Ground-truth Input Restored image

Figure 3.5: Additional examples of restored images.

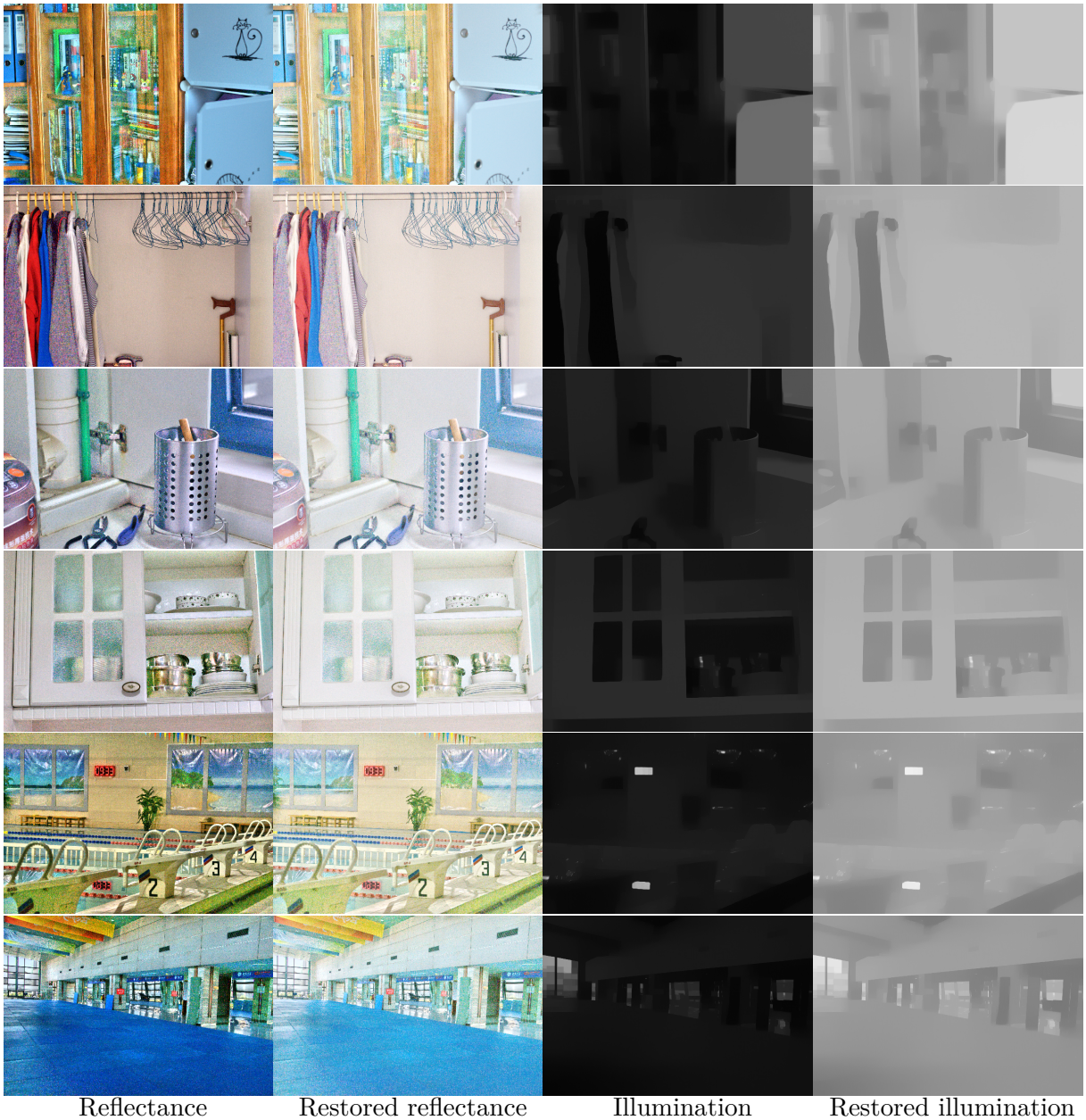


Figure 3.6: Examples of restored components.

Chapter 4

GAN architecture leveraging a Retinex model with colored illumination

4.1 Improvements to the Retinex model 59
4.1.1 A colored illumination 59
4.1.2 New priors 61
4.2 The Architecture 62
4.2.1 Architecture choices 62
4.2.2 The loss terms 64
4.2.3 Visualization and restoration of the components 66
4.3 Results 70
4.3.1 Metrics & Evaluation methodology 70
4.3.2 Implementation details 71
4.3.3 Ablation study 73
4.3.4 Qualitative comparison 73
4.3.5 Guarantees on a restoration without hallucination 81
4.4 Chapter conclusion 83

In this chapter, we argue that the most common Retinex decomposition model with a smooth grayscale illumination found in the low-light literature is too restrictive. We propose to relax this constraint by defining a new decomposition model with a colored illumination. The extraction of the components then becomes an even more difficult problem. To solve it, we propose firstly the idea of extracting common information[1] according to the physics of light, and define a GAN-based architecture allowing, *in fine*, a restoration of low-light images in an unsupervised way. This chapter has been published in a previous work [Lecert *et al.* 2023].

Unsupervised methods are well-suited to our problem because we do not have access to the ground-truth restored images. Deep image priors are slow at test time and there is a need for faster restoration approaches. Therefore, in order to train quicker GAN-based architectures, we turned to a bigger dataset, the Waymo dataset [Sun *et al.* 2020]. Moreover, our new architecture better decomposes images according to our model which leads to higher-quality restored images as shown in the next sections.

According to the physical definition of reflectance [ISO-92882022 *et al.* 2022], this property of a material represents the fraction of the radiance of the light reflected by a surface over the radiance of the light received by this surface. However, a consequence of this definition, is that the value of the reflectance of a material varies with respect to the wavelength of the incident light. In other words, in nighttime images, where the lighting is colored, the estimated reflectance does not contain the true colors of an object but only partial information, and the illumination is colored. Therefore, we propose a novel Retinex model that takes into account low-light characteristics, and decompose an image into the product of the reflectance and illumination, with two main differences. First, the reflectance is *the common information between daytime and nighttime image domains after a specific correction is applied*. Indeed, the low-light reflectance that we can estimate is only a portion of the whole normal-light reflectance. Thus, the low-light reflectance is the common information, the best estimate of the reflectance we can possibly extract from the two domains because of this degradation. By contrast, previous contributions in low-light restoration assumed that the reflectance was equal under daytime and nighttime

[1]We stress that this notion is distinct from the mutual information coming from information theory. The mutual information is a measure of dependence between random variables while we define this common information as the features independent of the two domains.

lighting. Second, the *illumination is colored*, whereas previous contributions considered grayscale illumination.

In summary, the main contributions in this chapter are as follows:

- We formulate different improvements to the original Retinex decomposition model thanks to a colored illumination and define new appropriate priors for the components. The first one is designed to avoid the scale ambiguity problem of the decomposition and the second one deals with the problem of a saturated sensor and its effect on the resulting components.

- We also propose a new architecture with deep neural networks inspired by state-of-the-art source separation and style transfer methods trained in an unsupervised fashion taking a single standard RGB image as input. This deep neural network has two branches, one for each of the component and outputs two colored images: the RGB illumination and the reflectance. It is trained with additional loss terms corresponding to physical priors such as the reflectance being the degraded common information between the night and daylight image distributions.

- We demonstrate the efficiency of our method compared to the competitors in the literature on a real world dataset without any ground-truth [Sun *et al.* 2020]. We then show the first visualization of the Retinex components following the physics of light as well as the original Horn's model [Horn *et al.* 1974] while only coarse approximations can be found in the literature.

4.1 Improvements to the Retinex model

4.1.1 A colored illumination

Since the Retinex decomposition model is only valid if applied to the irradiance of the camera sensor, it cannot be used directly on a standard RGB image. The intensity of the image needs to be linear with respect to this irradiance. Thus, the non-linear camera operations (i.e. mainly gamma corrections and tone mappings) need to be reversed. Complex pipeline can be used to achieve this goal whether by estimating the camera response

model [Kim *et al.* 2012; Ying *et al.* 2017; Jiang *et al.* 2013] or by reversing each step of the image processing pipeline [Liu *et al.* 2020; Huang *et al.* 2020; Brooks *et al.* 2019]. We assume in this chapter that we can reverse the image processing pipeline by only inverting the gamma correction ($\gamma = 2.2$).

The reflectance is officially defined by [ISO-92882022 *et al.* 2022] as the fraction of the radiance reflected by a surface over the radiance received by that surface. In the computer graphics community (e.g. [Cook *et al.* 1982; Boss *et al.* 2021; Yu *et al.* 2022]), the illumination is assumed to be a colored component. Since the spectral reflectance curves of the material present in the scene depends on the wavelength, this property is needed to accurately simulate the reflection of light rays. If the light sources in the scene are colored and not "white", the camera sensor only receives partial information about the whole spectral reflectance curve. Thus, the reflectance cannot be considered as the true color of the scene in this case. It is only a ratio map over the three bands, RGB, in the visible spectrum. Therefore, it can be counter-intuitive and not look like a realistic image. Different authors tried to reconstruct the spectral reflectance curve of the scene in a discrete fashion [Zhu *et al.* 2021; Yan *et al.* 2020; Fubara *et al.* 2020; Borsoi *et al.* 2021] or in a continuous one [Xu *et al.* 2023]. However, identifying each material in low-light images is extremely challenging and ill-posed in practice without using strong priors on the diversity of the present elements because of the metamerism effect (i.e. one RGB color can be the result of different combinations of wavelength). Indeed, in a night image, all colors result from artificial lights (e.g. streetlamps, car headlights, ...) reflecting on the different objects in the scene and then going straight through the camera sensor. Since these artificial lights are colored and the reflectance spectra of the objects in the scene are highly non-linear, we only observe a tiny portion if not none of the "true" colors (i.e. the color under a white light) of the scene. We introduce a different definition of the Retinex decomposition to address these challenges. In the literature, indoor datasets such as LOL [Wei *et al.* 2018] don't fully represent the complexity of the degradation in outdoor images. Reducing the exposure to simulate a low-light image is too simplistic to capture the whole shift of the distribution. In this chapter, we also do not consider the multiple scattering of light rays or the attenuation of the fog during the night in the outdoor scene to simplify the model as opposed to [Narasimhan *et al.* 2003] for instance.

Instead, we define the reflectance as the corrected common information between two

distributions of images (which are assumed to have similar scenes and objects but not paired images). Indeed, the reflectance extracted from low-light images is degraded and thus not equal to the one estimated from normal-light images. In that sense, the reflectance is really independent of the domain while the illumination is the light-dependent component and contains, for instance, the specific noise of low-light images.

With these definitions, we can formulate the Retinex decomposition problem as follows:

$$I^{\gamma} = \left(\frac{L}{\alpha} + \eta\right). * \alpha R \tag{4.1}$$

where $I \in [0,1]^{3n}$ the RGB image, $L \in [0,1]^{3n}$ the illumination, $\alpha \in \mathbb{R}^*$ a scaling factor as the decomposition leads to an infinity of solutions. $R \in [\varepsilon, 1-\varepsilon]^{3n}$ as the only object which absorbs all light is a black hole. On the contrary, perfect mirrors are still not widely commercialized and may not appear frequently in our everyday lives. $\varepsilon = 1e^{-8}$ in our experiments. Thus, we relax the original previous model of the grayscale constraint of the illumination. The component now has a local chrominance in addition to a local luminance value. In Section 4.1.2, we quickly define an additional prior to reduce the solution set with the scaling factor α and extend a previously published prior in [Lecert *et al.* 2022] to address a colored illumination. Since the illumination is supposed to contain the low-light noise and degradations, adding an illumination smoothness prior (e.g. the one in [Wei *et al.* 2018; Zhang *et al.* 2019] would be inefficient).

4.1.2 New priors

Scale ambiguity (High reflectance prior)

The α factor introduced in the model (4.1) highlights the scale ambiguity problem in the Retinex decomposition. Any positive real value can lead to a plausible solution. To reduce even further the solution set we propose a new prior defined as follows:

$$\mathcal{L}_{HR} = \left\| \frac{1}{\alpha} \right\|_1 \tag{4.2}$$

$\alpha = 1$ as an initial value before the optimization process. As we are working with low-light images, we assume that the illumination should have the lowest possible value. On the other hand, minimizing this prior is equivalent to seeking for the highest reflectance.

Intuitively, it can be seen as considering a high V-channel (HSV) for the reflectance, one with a low value for the illumination. This also means that the optimization process is biased against black bodies.

Exposure prior (RGB version)

In this section, we extend the exposure prior [Lecert *et al.* 2022] to RGB images. As long as the camera sensor is not saturated by the light ray (i.e. $I_{c\in\{R,G,B\}} \neq 1$), the illumination cannot be saturated as well (i.e. $L_{c\in\{R,G,B\}} \neq 1$). This prior is defined to prevent the trivial solution where $L = 1$ and thus $I = R$. Only light sources or overexposed regions in the input image should lead to such values in the components. This prior is defined as

$$\mathcal{L}_E = \sum_{c\in\{R,G,B\}} \left\| g\Big(\max_{c\in\{R,G,B\}} I_c \Big) - g(L_c) \right\|_2^2 \tag{4.3}$$

where g is a threshold function $g(x) = \begin{cases} x, & x > 1-\epsilon \\ 0, & \text{otherwise} \end{cases}$ and $\max_{c\in\{R,G,B\}} I_c$ an approximation of the illumination following the work LIME by Guo *et al.* [Guo *et al.* 2017].

4.2 The Architecture

4.2.1 Architecture choices

In this section we describe the architecture shown in Fig. 4.1 that we propose to decompose an image. Since we do not have access to the ground-truth images and want a low execution time, we use a GAN-based architecture. To better separate the input image into the two components, we make use of two discriminators, one to generate each of the component respectively.

The network is composed of two branches to extract each component. We build on the YTMT source separation strategy [Hu *et al.* 2021] which consists of alternating positive ReLu on one branch and negative ReLu on the other to avoid losing information and to better connect the two networks together. We use two UNets [Ronneberger *et al.* 2015]. The illumination branch receives as input an approximation of the illumination of the input image I.

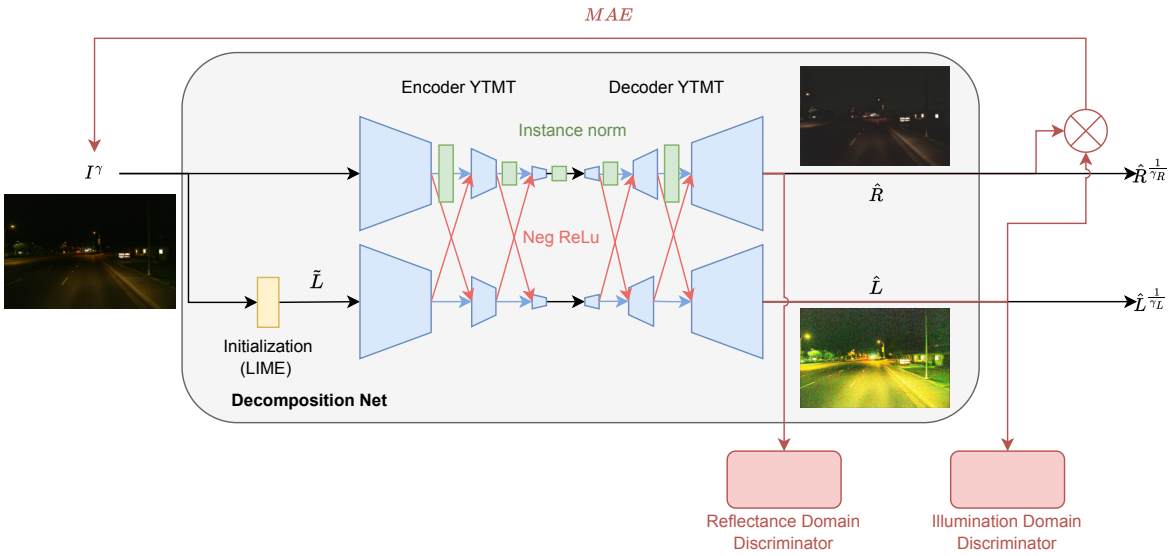


Figure 4.1: An illustration of architecture composed of two branches: the upper branch extracts the reflectance from the image normalizing the style of the features, the lower branch on the contrary keeps the style information to produce the illumination and takes as input an approximation of this component to facilitate the process. The two branches can swap information with the YTMT strategy [Hu *et al.* 2021] for source separation. Each component has its own discriminator to follow the definition of the Retinex components with common information.

In MUNIT [Huang *et al.* 2018], the authors managed to transfer the style of an image such that the resulting image belongs to another domain while preserving the content of the image. To perform that, they improve upon the work of [Gatys *et al.* 2015] and the surprising result of instance normalization [Ulyanov *et al.* 2017]. It consists of aligning the mean and variance of the content features with those of the style features [Huang *et al.* 2017]. By nature, this problem is similar to the Retinex decomposition problem if we consider the reflectance as the content we try to preserve and the illumination as a complex style (i.e. a whole RGB image instead of mean and variance parameters). Therefore, we add instance normalization modules to the reflectance extraction branch. The information of the style of the image flows through the illumination branch.

4.2.2 The loss terms

The reconstruction loss

For ease of notation, we omit the α scaling factor in the following equations to compute the two estimated components $(\hat{L}, \hat{R})$.

$$\tilde{L}_d = \lambda * \max_{c \in \{R,G,B\}} I_{c,d} \tag{4.4}$$

$$G_R : I_d \in \mathbb{R}^{3n} \mapsto \hat{R}_d \in \mathbb{R}^{3n} \tag{4.5}$$

$$G_L : \tilde{L}_d \in \mathbb{R}^{3n} \mapsto \hat{L}_d \in \mathbb{R}^{3n} \tag{4.6}$$

where λ is the mean triplet RGB over the spatial dimensions of the input image I_d, $d \in \{0, 1\}$ the label being equal to 1 for the normal-light domain and 0 for the low-light one. G_R and G_L denote the autoencoders for both R and L components, I_d and $\tilde{L_d}$ are the inputs given by LIME.

To make sure that the two generated components we get can reconstruct the input image according to the Retinex model, we use the mean absolute error for the structure of the image and the angular error to ensure that the color is accurately recovered with the decomposition. This can be summed up as the following terms,

$$\mathcal{L}_{MAE} = \left\| I_d^{\gamma} - \hat{L}_{d.} * \hat{R}_d \right\|_1 \tag{4.7}$$

$$\mathcal{L}_{color} = \arccos\left(\frac{I_d^{\gamma} \cdot \left(\hat{L}_{d.} * \hat{R}_d \right)}{\left\| I_d^{\gamma} \right\| \left\| \hat{L}_{d.} * \hat{R}_d \right\|} \right) \tag{4.8}$$

We don't use the latent reconstruction terms like in [Huang *et al.* 2018] since applying the illumination of one image to the reflectance of another would result in an unrealistic and not plausible image and then would mislead the discriminators during the training

process. The $\mathcal{L}^1$-norm guarantees that no information is lost during the decomposition process. However, we directly extract a noisy illumination instead of a noiseless version since it is easier to do so and then denoise the resulting component.

Domain discriminator adversarial functions

Our architecture relies on adversarial loss terms to find the components in an unsupervised fashion. We define the two discriminators function as follows,

$$D_R : \hat{R}_d \mapsto \hat{d} \tag{4.9}$$

$$D_L : \hat{L}_d \mapsto \hat{d} \tag{4.10}$$

$$\hat{d} \in \{0, 1\}.$$

where $\hat{d}$ is the estimated label resulting from each of the component. We use a multiscale discriminator architecture such as the one in [Pizzati *et al.* 2021]. The discriminators need to be able to identify the domain of the input component they get (i.e. separate each component according to their domain). We empirically find that the training of the generators is more stable with the Least Squares GAN [Mao *et al.* 2017] than the other versions. The parameters of (G_R, G_L) are fixed in this pass.

$$\mathcal{L}_{D_L} = \sum_{d \in \{0,1\}} \mathbb{E}_{L_d}\Big[\Big(D_L(G_L(L_d)) - d\Big)^2\Big] \tag{4.11}$$

$$\mathcal{L}_{D_R} = \sum_{d \in \{0,1\}} \mathbb{E}_{I_d}\Big[\Big(D_R(G_R(I_d)) - d\Big)^2\Big] \tag{4.12}$$

Generator adversarial functions

To train the generators (G_R, G_L), we fix the parameters of (D_R, D_L). The illumination generator should extract the component from the image and the domain should be accurately identified by the corresponding discriminator. On the contrary, we seek to extract information which can't be classified by the reflectance discriminator between the low and normal-light domains. Therefore, we optimize it to align the low-light reflectance to the

normal-light one. This leads to the following equations,

$$\mathcal{L}_{G_L} = \sum_{d\in\{0,1\}} \mathbb{E}_{L_d}\Big[\Big(D_L(G_L(L_d)) - d\Big)^2\Big] \tag{4.13}$$

$$\mathcal{L}_{G_R} = \sum_{d\in\{0,1\}} \mathbb{E}_{I_d}\Big[\Big(D_R(G_R(I_d)) - 1\Big)^2\Big] \tag{4.14}$$

The resulting optimization problem

As a result, we obtain the following problem to train the decomposition network,

$$\begin{aligned}\Big(\hat{G}_R, \hat{G}_L\Big) = \underset{G_R,G_L,\alpha}{\operatorname{argmin}}\ & \lambda_{MAE}\mathcal{L}_{MAE} + \lambda_{color}\mathcal{L}_{color} \\ & + \lambda_{HR}\mathcal{L}_{HR} + \lambda_E\mathcal{L}_E \\ & + \lambda_{adv}(\mathcal{L}_{G_L} + \mathcal{L}_{G_R})\end{aligned} \tag{4.15}$$

and for the Retinex components domain discriminators,

$$\Big(\hat{D}_R, \hat{D}_L\Big) = \underset{D_R,D_L}{\operatorname{argmin}}\ \mathcal{L}_{D_R} + \mathcal{L}_{D_L}. \tag{4.16}$$

4.2.3 Visualization and restoration of the components

One of the key benefits of considering a complex style as the illumination (i.e. a style that has the same dimensions of an image here) is that it can be visualized. Some examples of the obtained Retinex components are illustrated in Fig. 4.2.

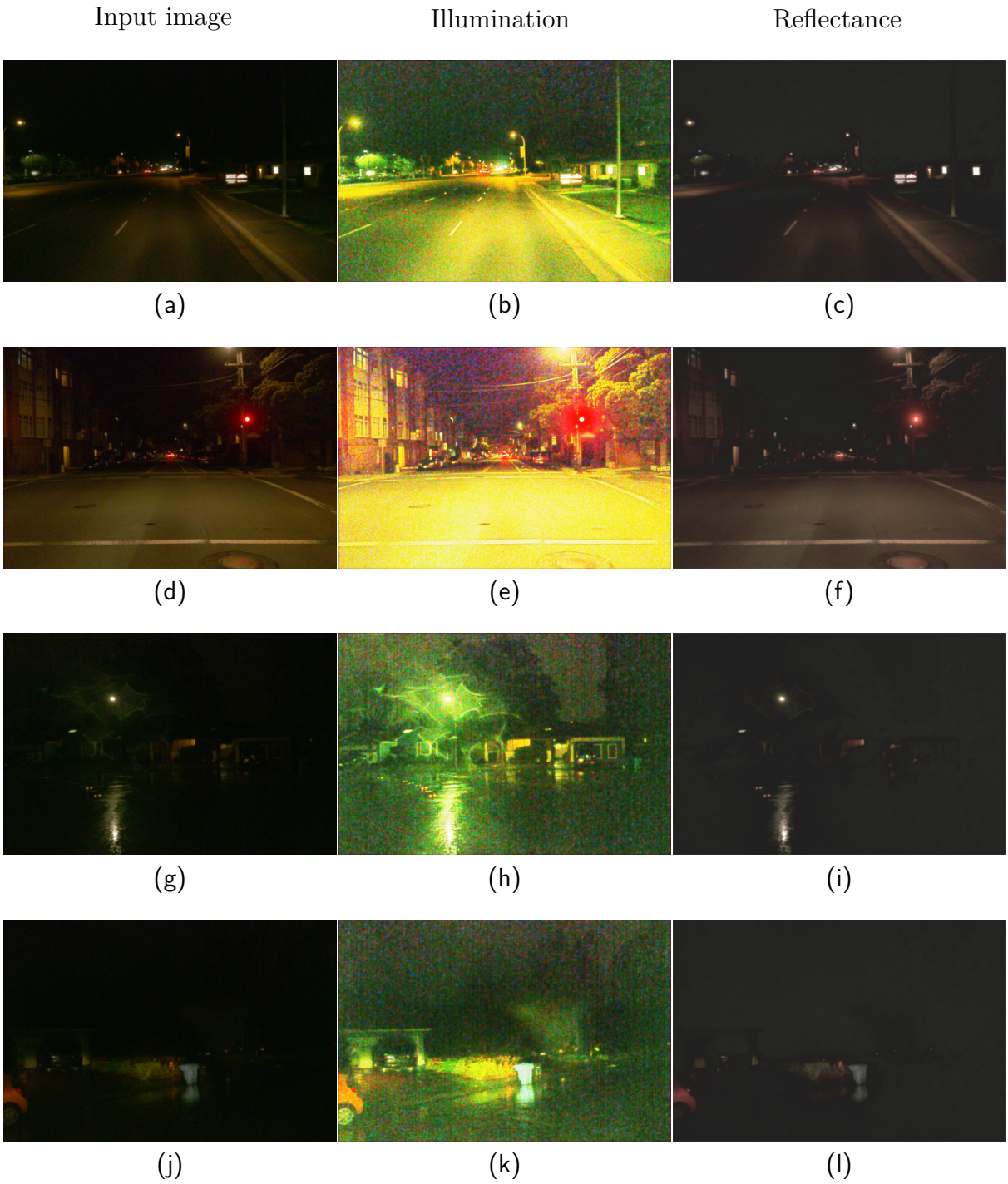


Figure 4.2: From left column to right column: the input image, the extracted raw illumination and reflectance. The illumination contains the specific low-light noise and degradations.

To the best of our knowledge, this is the first time that these components are linked to the style transfer literature and that the common and specific domain information are displayed. The reflectance is sharper than the input image. The dark areas present in it can be intuitively explained as loss or missing information about the scene. It shows that this component also needs a custom restoration. Besides, shadows and light rays coming from the car headlights and streetlamps still end up in the reflectance. However, the glare effect of the light sources are reduced and the colors are less saturated. On the other hand, the illumination contains the low-light noise and degradations. We seek to restore low-light images but not estimate a daylight version. Thus, we denoise the illumination component instead of the reflectance like in the previous works in the literature and avoid any tone mapping to avoid amplifying the noise. We use a weight map to denoise the component according to the structure of the scene in the reflectance in Fig. 4.3. Using the maximum of the reflectance over the color RGB channels gives more information than a simple gradient and guide the denoising network to strengthen the denoising process where the reflectance does not have a lot of information in the dark areas. We try to find the best compromise to keep the maximum of information in the image. Since the reflectance contains the textural details of the different objects of the scene, we seek to amplify this information to highlight and make it easier to distinguish the elements. The different works in the literature do it with a tone mapping function such as a gamma correction [Guo *et al.* 2017] or by the use of a neural network [Li *et al.* 2020; Zhang *et al.* 2021]. Defining which type of function to apply here can be difficult without any ground-truth images or priors to control the exposure of the image (e.g. section 3.3 in [Li *et al.* 2020]). Therefore, we decide to use the simple and efficient gamma correction with a low execution time. We empirically find the best gamma values by maximizing the LPC-SI metric [Hassen *et al.* 2013]. We also found that we can get visually better results with a higher correction on the blue channel to balance the colored noise of the illumination. This effect is dataset-specific though and not mandatory in other cases. It may be due to the sensors that the authors used but we couldn't verify this hypothesis. We also tried a unique gamma for all the channels or in the HSV domain but the results were either too whitened by the process or the colors too saturated. An example of the restoration process of the illumination is shown in Fig. 4.3 and in Fig. 4.4 for the reflectance. Using a gamma correction on the reflectance does not reveal a hidden noise or another low-light degradation. This shows the high quality of the decomposition.

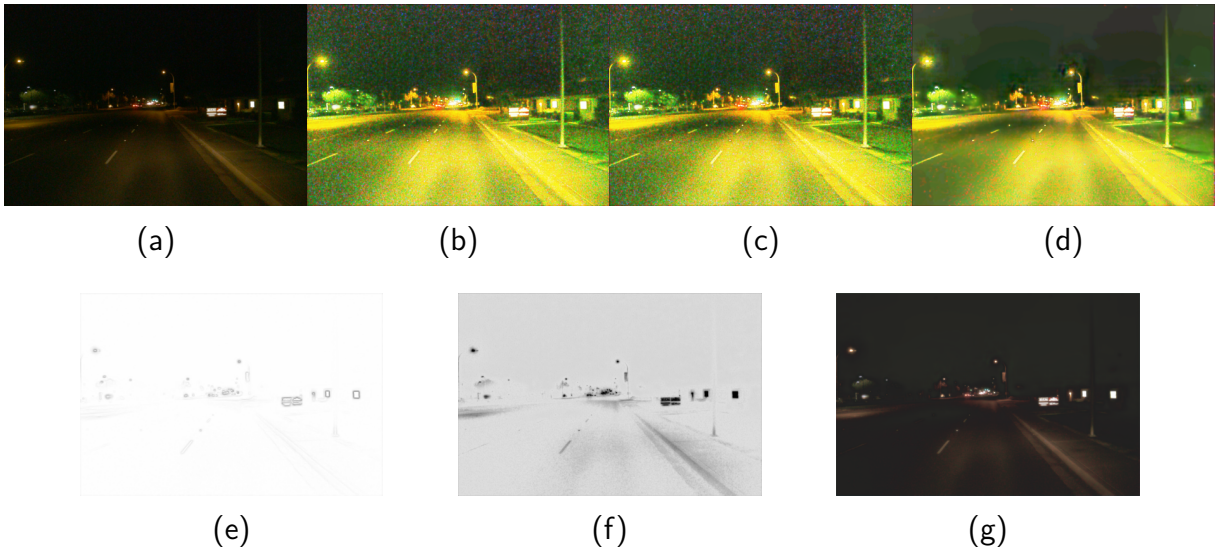


Figure 4.3: From left to right: the input image Fig. 4.3a, the extracted raw illumination Fig. 4.3b, the denoised component Fig. 4.3c, the component if we denoise before the decomposition Fig. 4.3d. To denoise the illumination, the noise map is weighted. Since the gradient of the reflectance Fig. 4.3e has less information about the structure of the scene than the weight map Fig. 4.3f as the negative approximation of the illumination as defined in LIME [Guo *et al.* 2017], we use the latter here. We could not further denoise the image as it would lead to a loss of details. Fig. 4.3d shows that if we denoise the image before the decomposition, it only affects the illumination as the reflectance remains untouched Fig. 4.3g. We decide to denoise the component after the decomposition as the former would lead to some artifacts introduced by the denoising network and amplified by the restoration of the components.

(a) Input image (b) Raw reflectance (c) Gamma corrected reflectance

Figure 4.4: From left to right: the input image, the extracted raw reflectance, the gamma corrected component. The low-light noise is not strengthened after the restoration which confirms the components are correctly separated according to the defined model. We restore with a different γ for each RGB color channel as we empirically find it leads to visually better output images. The higher gamma value for the blue color channel giving the component its non-natural hue is set to balance the green noise of the illumination. See Fig. 4.11 for an example of the components extracted from an image of the BDD100K dataset [Yu *et al.* 2020].

4.3 Results

4.3.1 Metrics & Evaluation methodology

As we are not seeking to approximate the distribution of daylight images with the restored images and we have no ground-truth, neither commonly used metrics like FID [Heusel *et al.* 2017], IS [Salimans *et al.* 2016] or CIS [Huang *et al.* 2018] nor classic reference-based metrics such as PNSR, SSIM [Wang *et al.* 2004] or LPIPS [Zhang *et al.* 2018] can be used here. The LPC-SI metric [Hassen *et al.* 2013] measures the sharpness of an image through local phase coherence of complex wavelet coefficients. Even though it cannot measure the whole low-light degradation, being sharp is one of the properties we desire for the result. We choose the best gamma values to apply for the gamma correction over the RGB color channels with a trade-off between the LPC-SI metric and the visual quality of the images.

As there are no ground-truth images available for our problem, we consider a two-steps process to evaluate the methods: First, we use a visual approach: we observe the level of noise and note if there are hallucinations in the outputs. Then, since there are no

ground-truth in the datasets, we cannot use metrics with reference. Thus, we compare the methods using a non-visual test with the reference-free LPC-SI metric [Hassen *et al.* 2013]. The results are illustrated in Table 4.1. The hallucinating methods can reach higher scores since they invent very sharp objects. On the contrary, ours gives the sharpest images among the methods which cannot hallucinate. The LPC-SI scores of the different methods are shown in Table 4.1.

Methods	LPC-SI(↑)
MUNIT [Huang *et al.* 2018]	0.95428
$\text{EnlightenGAN}_{\text{pretrained}}$ [Jiang *et al.* 2021]	0.97610
$\text{EnlightenGAN}_{\text{waymo}}$ [Jiang *et al.* 2021]	**0.97848**
Retinex DIP color [Lecert *et al.* 2022]	0.94439
Retinex DIP gray [Lecert *et al.* 2022]	0.94686
Zero-DCE [Li *et al.* 2020]	0.96943
$\text{Gamma Correction}_{\text{HSV}}$	0.96746
$\text{Gamma Correction}_{\text{RGB}}$	0.96463
KinD++ [Zhang *et al.* 2021]	0.96572
LIME [Guo *et al.* 2017]	0.96343
Ours	**0.97152**

Table 4.1: LPC-SI scores [Hassen *et al.* 2013] on the Waymo dataset with respectively in blue methods that hallucinate and in black methods which don't. Scores in bold are the highest scores in each of the category. We obtain the best LPC-SI score among the non-hallucinating approaches.

4.3.2 Implementation details

We use the ADAM optimizer [Kingma *et al.* 2015] with a fixed learning rate of $1e^{-4}$ optimized over 200 epochs, Pytorch [Paszke *et al.* 2019] as framework and the Kornia library [Riba *et al.* 2019]. We empirically find the coefficients $\lambda_{\text{recon}} = 5e^1, \lambda_{color} = 1e^1, \lambda_{HR} = 1, \lambda_E = 5e^1, \gamma_R = 2, \gamma_G = 2, \gamma_B = 6, \sigma_R = 15, \sigma_G = 10, \sigma_B = 15$. We crop the images to the size 256x256 and group them by 3 to make a batch. To denoise the component, we use the plug-and-play denoising network trained on spatially varying noise in [Le Pendu *et al.* 2022]. We found out that we get better results using the same noise level map for all color channels as illustrated in Fig. 4.5. Different datasets take part in the following studies. Their properties are shown in Table 4.2.

Dataset	Publicly available	Indoor/Outdoor	Contains paired images	Contains real-world scenes	Size
Waymo [Sun *et al.* 2020]	✓	Outdoor	✗	✓	128 093 normal-light images & 15419 low-light images
BDD [Yu *et al.* 2020]	✓	Outdoor	✗	✓	14772 low-light images
LOL [Wei *et al.* 2018]	✓	Indoor	✓	✓	500 pairs

Table 4.2: Properties of the different datasets we use throughout the chapter.

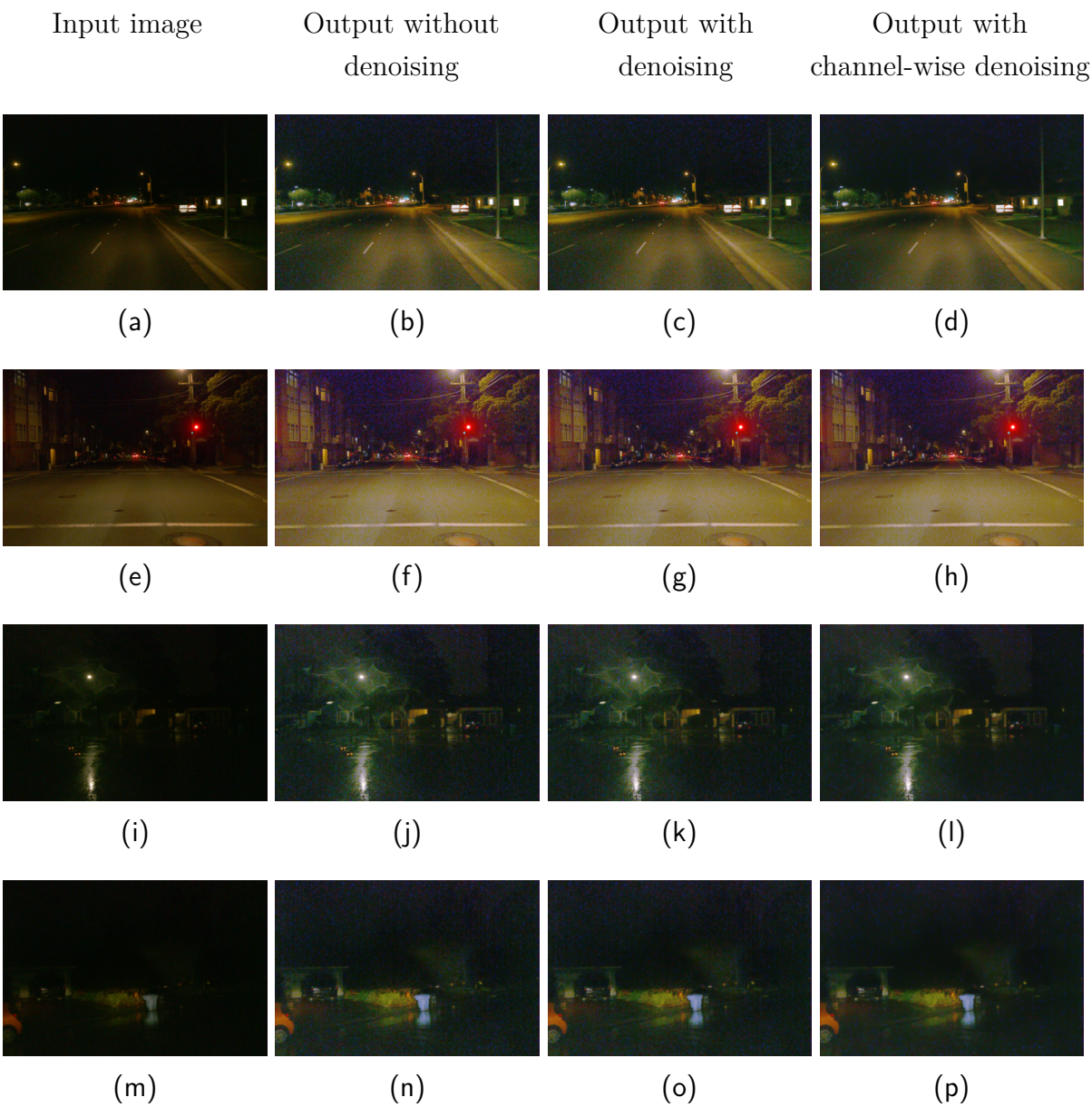
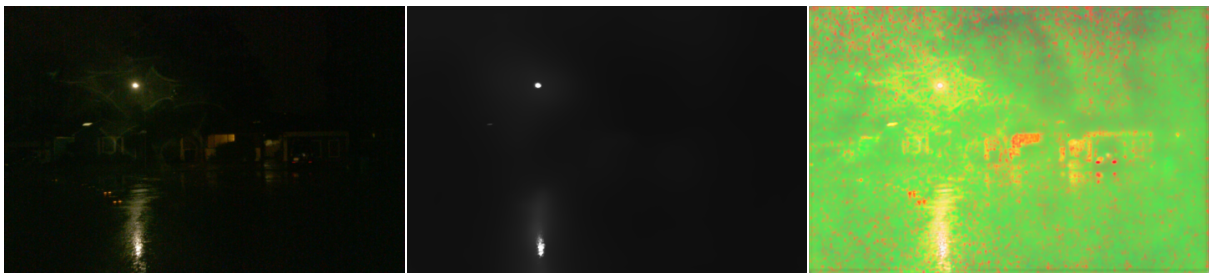


Figure 4.5: From left column to right column: the input image, the output of our method without denoising, with denoising and denoising with a different noise level for each RGB color channel.

4.3.3 Ablation study

If we consider the original Retinex model with a grayscale smooth illumination, we get the results shown in Fig. 4.6. Then, even if we denoise the reflectance, we can't obtain visually pleasing outputs here.



(a) Input image (b) Illumination (c) Reflectance

Figure 4.6: From left to right: the input image, the extracted raw illumination and reflectance if we consider a grayscale smooth illumination as previously used in the literature.

In Fig. 4.7, Retinex components of a daylight image are illustrated. Textural details such as the frontage of the buildings end up in the reflectance.

(a) Input image (b) Illumination (c) Reflectance

Figure 4.7: From left column to right column: the input normal-light image, the extracted raw illumination and the reflectance. Textural details such as the frontage of the buildings end up in the reflectance which demonstrates the quality of the decomposition model.

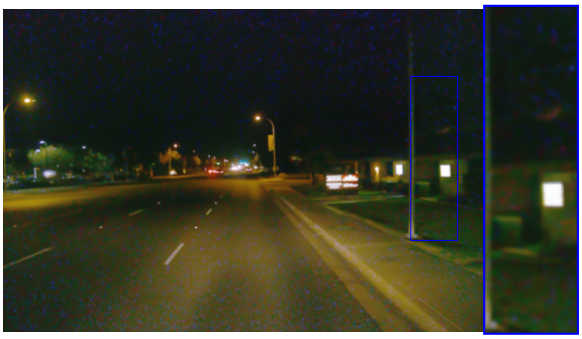
4.3.4 Qualitative comparison

On the Waymo dataset

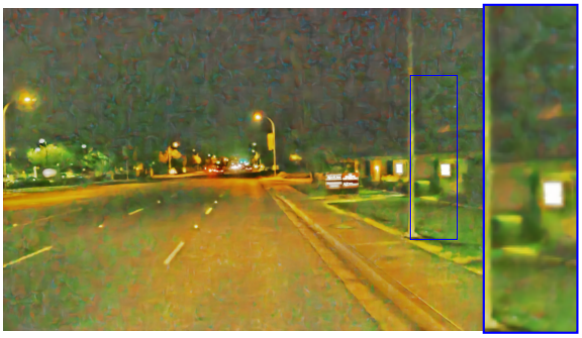
The state-of-the-art results are illustrated in Figs. 4.8 and 4.9. Fig. 4.9 contains the outputs of several style transfer methods. These works are mainly aiming at augmenting

data to enhance datasets with the goal of training networks which will be robust to these modifications. CoMoGAN [Pizzati *et al.* 2021] simulates night images with daylight images and can't do the reverse process as seen in Fig. 4.9b. In Figs. 4.9c and 4.9d, ManiFest [Pizzati *et al.* 2022] and MUNIT [Huang *et al.* 2018] completely modify parts of the image and do not preserve the integrity of the scene which is undesirable in our case.

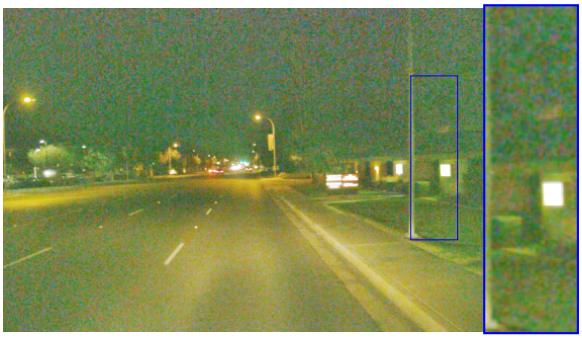
(a) Input image



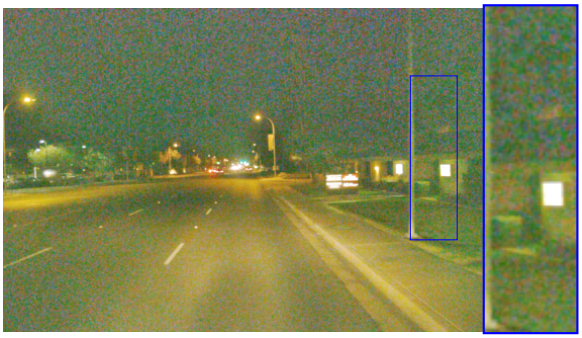
(b) Our method



(c) KinD++



(d) RetinexDIP color

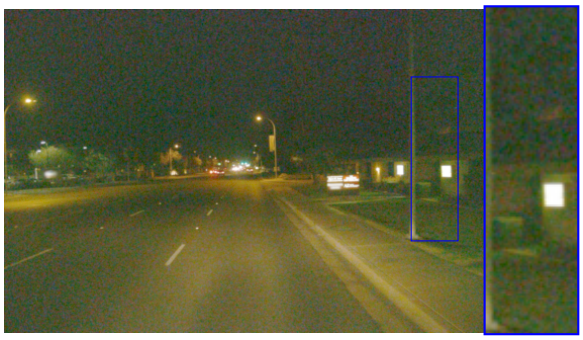


(e) RetinexDIP gray

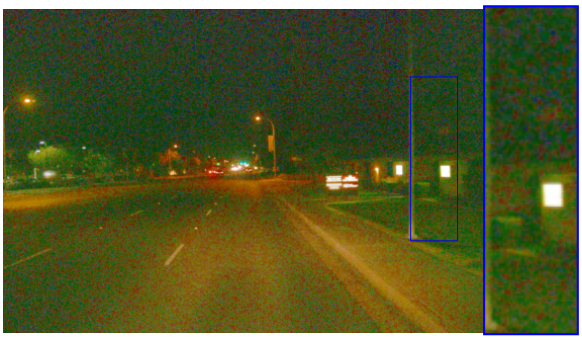
Figure 4.8: First part of the qualitative comparison between the restoration methods applied to the input image Fig. 4.8a and the state-of-the-art approaches. Our method Fig. 4.8b leads to a visually better result than the competitors with respectively Fig. 4.8c KinD++ [Zhang *et al.* 2021], Fig. 4.8d Retinex DIP color [Lecert *et al.* 2022], Fig. 4.8e Retinex DIP gray [Lecert *et al.* 2022].

(f) EnlightenGAN

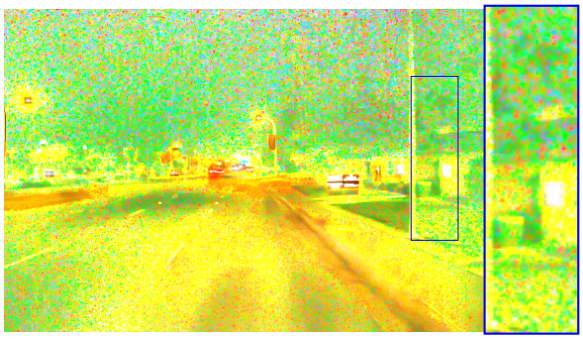
(g) EnlightenGAN fine-tuned on Waymo



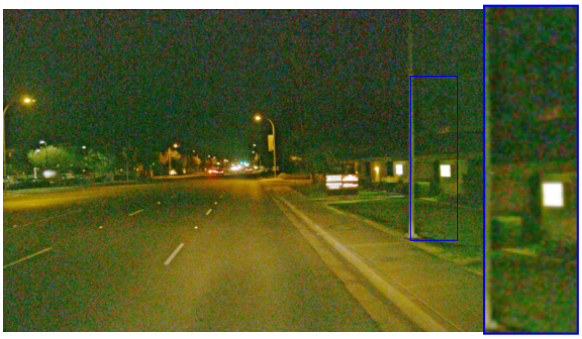
(h) Gamma correction in RGB



(i) Gamma correction in HSV



(j) LIME



(k) Zero-DCE

Figure 4.8: Second part of the qualitative comparison between the restoration methods applied to the input image Fig. 4.8a and the state-of-the-art approaches. Fig. 4.8f EnlightenGAN [Jiang *et al.* 2021] and Fig. 4.8g EnlightenGAN fine-tuned on the Waymo dataset [Jiang *et al.* 2021] hallucinates trees at the top of the image as shown in the red squares. Applying gamma corrections in the RGB and HSV spaces leads to the outputs illustrated in Figs. 4.8h and 4.8i with an undesirable "fog" effect whitening the image. Fig. 4.8j LIME [Guo *et al.* 2017] and Fig. 4.8k Zero-DCE [Li *et al.* 2020] amplify the noise in the darkest parts of the image.

(a) (b) (c) (d)

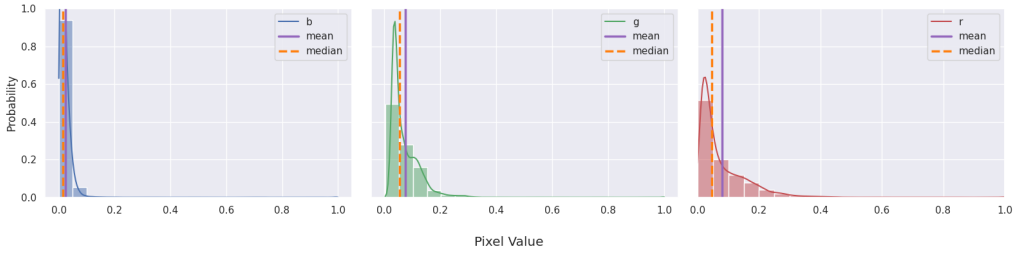
Figure 4.9: Style transfer and data augmentation methods with from left to right: Fig. 4.9a The input image, Fig. 4.9b CoMoGAN [Pizzati *et al.* 2021], Fig. 4.9c ManiFest [Pizzati *et al.* 2022], Fig. 4.9d MUNIT [Huang *et al.* 2018] trained on the Waymo dataset. These style transfer methods do not preserve the integrity of the scene in the input image and add hallucinations.

In Fig. 4.8f, EnlightenGAN [Jiang *et al.* 2021] hallucinates trees in the background as shown in the red squares which is obviously not desirable in our case. However, the image is sharper. We fine-tune it on the same dataset to see if we could improve the results and still obtained hallucinations in Fig. 4.8g.

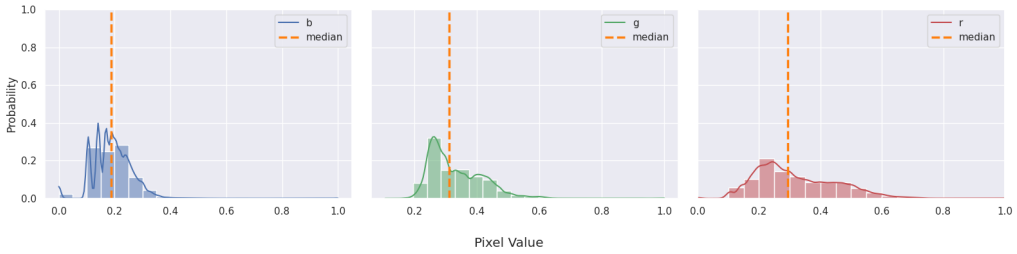
For the gamma correction, we apply the same gamma values as with our method (i.e. with a higher gamma for the blue channel) in the RGB color space and the result is shown in Fig. 4.8h. The "fog" effect results from the gamma correction applied on all the color channels. The histogram of the blue channel is shifted to the right as shown in Fig. 4.10b. This effect is not present if we restore the V channel in the HSV color space instead, see Fig. 4.8i and its corresponding histogram Fig. 4.10c. With our method, it's not the case even if we restore the reflectance with a gamma correction on all RGB channels as shown in Figs. 4.10d and 4.10e. There is little to no difference applying a gamma correction on all RGB channels or V channel in HSV with respect to the histograms. Moreover, the histograms are flatter than the ones with gamma correction only which can be seen as histogram equalizations or contrast enhancement. We get visually more appealing results with our method than the gamma correction.

Looking at the histograms of the input images, the blue channel pixels have really low intensity and there are more information about the scene in the red and green channels.

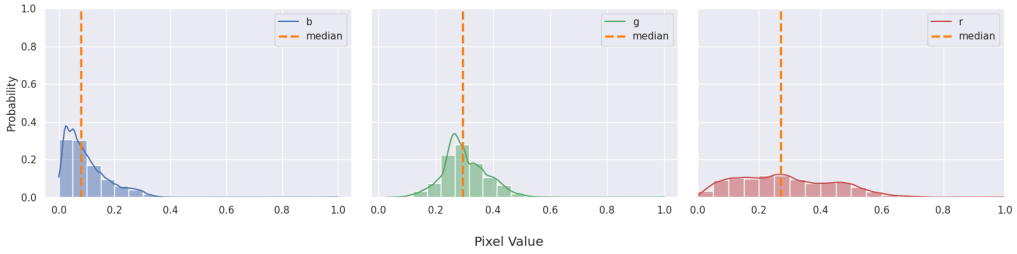
Our method Fig. 4.8b leads to a visually better result than the competitors such as LIME, Zero-DCE or KinD++ Figs. 4.8c, 4.8j and 4.8k. Applying our network trained on



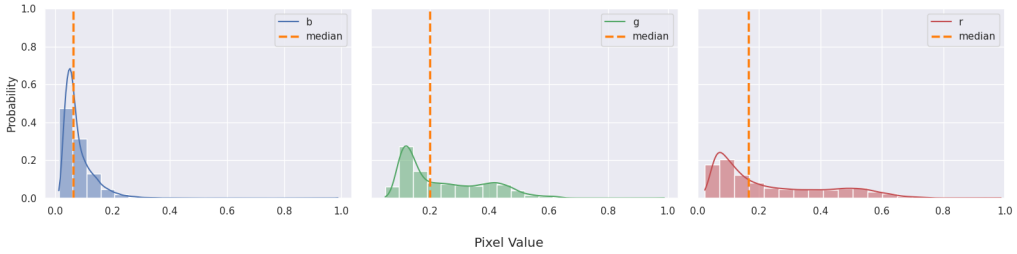
(a) Input image



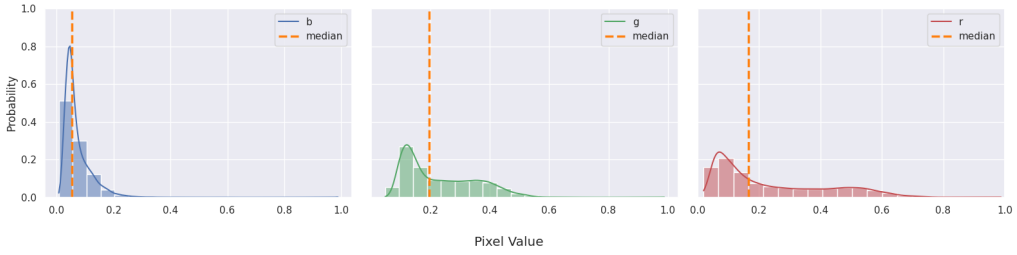
(b) Gamma correction in RGB



(c) Gamma correction in HSV



(d) Our method with a gamma correction in RGB



(e) Our method with a gamma correction in HSV

Figure 4.10: Comparison of histograms between the input image Fig. 4.10a, the image after applying a simple gamma correction on all color channels in RGB Fig. 4.10b or the V channel in HSV Fig. 4.10c, the image restored with a gamma correction according to our method in RGB Fig. 4.10d or HSV Fig. 4.10e.

the Waymo dataset [Sun *et al.* 2020] to the BDD100k dataset [Yu *et al.* 2020] leads to the results shown in Fig. 4.11. We only reduce the $\gamma_B = 2$ to the same value as the other color channels as this dataset does not suffer from the green hue. We obtain similar results with a network pretrained on another dataset which really highlights the generalization ability of the restoration. Moreover, the visual quality of the decomposition is on par with the one on the Waymo dataset. Nonetheless, we emphasize that this dataset is composed of images with similar scenes (same objects and backgrounds).

(a) Input image (b) Restored image with our method

(c) Raw illumination (d) Restored illumination

(e) Raw reflectance (f) Restored reflectance

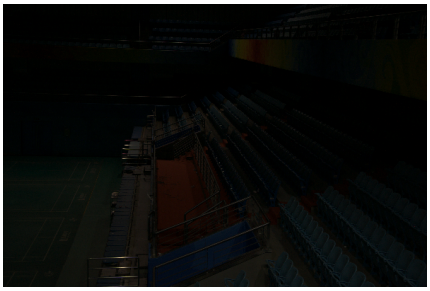
Figure 4.11: An example of the obtained Retinex components and the restoration output on an image coming from the BDD100k dataset [Yu *et al.* 2020]. We apply the network trained on the Waymo dataset [Sun *et al.* 2020]. From left column to right column: the input image Fig. 4.11a, the restored image Fig. 4.11b, the extracted raw illumination Fig. 4.11c and its restored version Fig. 4.11d, the extracted raw reflectance and its restored version Figs. 4.11e and 4.11f. The illumination still contains the low-light noise and degradations which strengthen the conclusion on the quality of the decomposition.

Failure cases

Fig. 4.12 illustrates the outputs of our method if the network is trained on the commonly known LOL dataset. Figs. 4.12a and 4.12b are respectively the input low-light image and its corresponding ground-truth normal-light image. Figs. 4.12c and 4.12d show the illumination and reflectance components we can extract if a colored illumination is considered. The dataset size being relatively small with around 500 paired images, a GAN-based architecture has trouble to learn in an unsupervised fashion. Specifically, the common information of widely diverse scenes is really challenging to estimate. If we consider a grayscale smooth illumination like in previous works, the constraints are strong enough to guide the decomposition but the low-light noise still ends up in the reflectance and makes it even more difficult to get rid of it. If we apply the network already trained on the Waymo dataset to the LOL dataset, we obtain results as shown in Figs. 4.12g and 4.12h. Since there is an enormous gap between the type of scene and degradation between the two datasets, the network experiences difficulty in extracting the correct information.

4.3.5 Guarantees on a restoration without hallucination

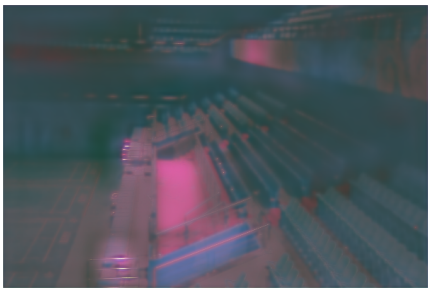
The main goal driving the design of our approach is to prevent adding fake details in the darkest parts of the input image. Here, we insist on the guarantees of our method regarding this aspect. First, it is a two-steps method: a decomposition phase and the restoration of the obtained components. The GAN only decomposes the image and is restricted by the reconstruction term to follow the Retinex model. The product of the two components is close to the input image (i.e. but not equal to take into account the noise). To restore them, simple gamma corrections are used which can not hallucinate. One could argue that if a component is equal to zero, then we lose the information in the output image. To address this point, $R \in [\varepsilon, 1-\varepsilon]^{3n}$ which prevents this issue. If we get a null illumination, it would mean that the input image is also null and therefore there is no information to restore from it. Our method neither loses nor adds information during the decomposition process. Besides, in practice, we do not observe any issue compared to the input image after a gamma correction for reference.



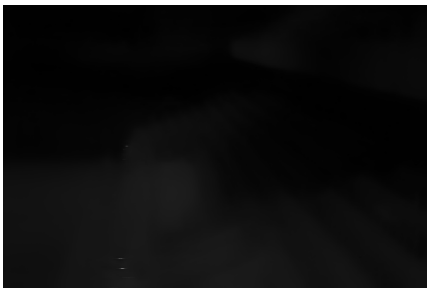
(a) Input image

(b) Ground-truth restored image

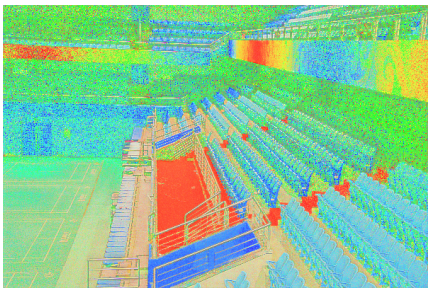
(c) Colored illumination



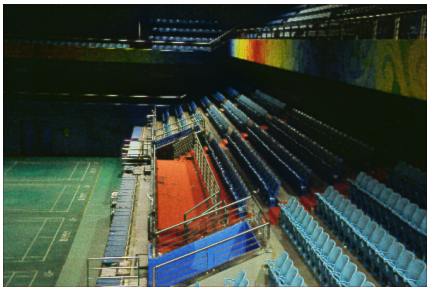
(d) Reflectance



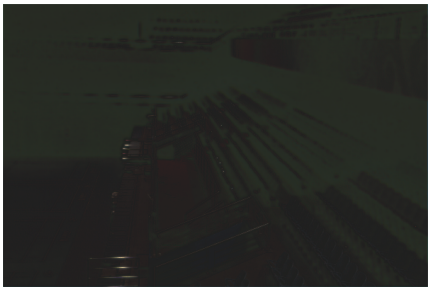
(e) Gray illumination



(f) Reflectance



(g) Gray illumination after pre-training



(h) Reflectance

Figure 4.12: Illustration of failure cases of the decomposition if applied on the LOL dataset on the input image Fig. 4.12a compared to the ground-truth image Fig. 4.12b. The model can't extract the two components if trained with a colored illumination Figs. 4.12c and 4.12d because the dataset is too small and contains a high diversity of scenes. The same problem occurs if pretrained on the Waymo dataset [Sun *et al.* 2020] Figs. 4.12g and 4.12h. Using a grayscale illumination for training Figs. 4.12e and 4.12f, it recovers a smooth illumination which was also a previous prior in the literature.

4.4 Chapter conclusion

In this chapter, we proposed a new approach based on state-of-the-art source separation and style transfer methods to decompose in an unsupervised fashion outdoor nighttime images. We improved the original Retinex model by extracting common information between the low and normal-light domain thanks to a colored illumination. Moreover, we also defined a new architecture with deep neural networks building on this physical model. To the best of our knowledge, this is the first time this definition of the Retinex components is put into practice. It makes it feasible to visualize the complex style known as the illumination and the reflectance in an image. Applied to the Waymo dataset, our method is more stable and produces visually pleasing images and more importantly without hallucinating parts of the image compared to the state-of-the-art methods.

However, different aspects of the method could be improved in a future work. Indeed, each non-linear operation applied by the camera pipeline like a gamma correction or a specific tone mapping makes it more difficult to decompose the image. Reversing this pipeline from a single input image is already an active research field in the literature. See [Liu *et al.* 2020] for instance. It remains an interesting research direction to improve the accuracy of the estimation of the Retinex components. The restoration of the components could also be improved. The gamma correction and denoising network could be replaced by a method inferring the missing information in the reflectance component and a more accurate noise model as long as the information of the image processing pipeline is provided.

We investigated the idea of preserving as much information as possible from the input scene with the Retinex model. Doing so, we aimed at learning to reverse the specific degradations of each component. This is essentially related to finding the optimal map to transport the image from the low-light distribution to the normal-light one. In the next chapter, we explore this topic and the link between the Retinex theory and the problem of the Schrödinger bridge. We tackle this problem building on the literature. Existing works introduce diffusion models to solve the dynamic form of the Schrödinger Bridge. As the empirical distributions may not be as faithful as necessary to the true latent image distributions, we seek to add physical priors from our context to regularize this optimization problem.

Chapter 5

Retinex theory and solving Schrödinger Bridges with Diffusion models

5.1 Introduction to the Monge-Kantorovich problem . . . 87
5.2 Link between the Retinex theory and the Schrödinger Bridge problem . . . 88
5.2.1 Entropy-regularized optimal transport and the static Schrödinger Bridge . . . 88
5.2.2 Link to the Retinex theory . . . 89
5.2.3 The dynamic form of the Schrödinger bridge . . . 90
5.3 The Schrödinger Bridge with score-based generative models . . . 91
5.3.1 The formulation of the problem . . . 91
5.3.2 Procedures to reduce the bias in the solutions . . . 95
5.3.3 Regularizations for the Schrödinger Bridge . . . 97
Contrastive regularization . . . 97
Adversarial term . . . 99
Retinex priors . . . 99
5.4 Results . . . 102
5.4.1 IMF . . . 102
5.4.2 CUT Scheme . . . 104
5.4.3 Adding the Retinex priors . . . 104
5.5 Chapter conclusion . . . 108

In the context of the Horn-Retinex decomposition, in the perfect case, the restoration of a night image consists of extracting the complete reflectance and correcting the illumination such that the resulting image belongs to the day domain. We seek to preserve this reflectance in the process while transporting the image from the low-light distribution to the normal-light one. This only changes the domain-dependent component, i.e. the illumination. This is essentially the intrinsic idea of optimal transport. If the empirical distributions of the two considered domains match the respective ground-truth distributions, finding the least distance mapping from one to another would enable us to perfectly simulate the degradations or on the contrary restore the images. In practice, since the reflectance we can estimate is degraded, both the illumination and the reflectance have to be restored in their own way. We explore in this chapter the link between high-dimensional optimal transport techniques and the Horn-Retinex model through the Schrödinger Bridge problem.

A first introduction to the classical Monge-Kantorovich and entropy-regularized optimal transport problems is given in the first part. Then, we highlight the intrinsic link between the static Schrödinger Bridge and the Retinex theory in the definition of the reflectance. Afterward, we present the dynamic form of the Schrödinger Bridge, the main optimization problem we seek to solve. This formulation is built on the recent score-based generative models. Therefore, a quick summary of how they work is made with an emphasis on their relation with the simulation of stochastic differential equations. The Schrödinger Bridge framework generalizes the diffusion models in the sense that, the diffusion models make a bridge matching the standard normal distribution and a data distribution, whereas the Schrödinger Bridge framework allows matching any two distributions. Multiple sources of bias are identified in the process. Recently published procedures such as Iterative Proportional Fitting (IPF) with diffusion models, or Iterative Markov Fitting (IMF) can mitigate a part of the bias and are presented. Additional priors coming from the Contrastive Unpaired Translation (CUT) scheme are also introduced since they enable the learned bridge to generalize beyond observed samples to some extent. Finally, we introduce priors which are new in this context and coherent with the physics of light to further guide the simulation of the bridges.

5.1 Introduction to the Monge-Kantorovich problem

Gaspard Monge was the first to describe the problem of optimal transport in his work [Monge *et al.* 1781]. He sought to find the most efficient way for a worker in terms of effort (i.e. distance to cross or time spent) to dig the smallest pile of dirt required to match an arbitrary shape. Mathematicians now formulate this problem to compare two probability distributions, for instance.

Here we start by considering the balanced version of a probabilistic transport with mass splitting: the so-called Monge-Kantorovich problem [Kantorovitch *et al.* 1958]. Consider two data distributions with positive densities π_0 (the night domain in our case) and π_T (the day domain) w.r.t the Lebesgue measure both with support on $\mathbb{R}^d$. Then, the problem can be described as follows:

$$\Pi^*_{0,T} = \underset{(I_0,I_T)}{\operatorname{argmin}} \left\{ \mathbb{E}_{(I_0,I_T)}(c(I_0, I_T)) : I_0 \sim \pi_0, I_T \sim \pi_T \right\} \tag{5.1}$$

where $c(x, y)$ is the cost function, $\Pi^*_{0,T}$ the optimal transport plan, (I_0, I_T) the coupling of random variables over the product space $\mathbb{R}^d \times \mathbb{R}^d$. This plan can be seen as a joint distribution between π_0 and π_T over $\mathbb{R}^d \times \mathbb{R}^d$ minimizing an arbitrary cost.

We use the squared $L2$ Euclidean distance as ground cost function $c(x, y) = \|x - y\|_2^2$ which enables us to define the squared 2-Wasserstein distance (or Earth mover distance) between two measures as

$$W_2^2(\pi_0, \pi_T) = \inf_{\gamma \in \mathcal{U}(\pi_0,\pi_T)} \left\{ \mathbb{E}_{(X,Y)\sim\gamma} \|X - Y\|_2^2 \right\}. \tag{5.2}$$

where $\mathcal{U}(\pi_0, \pi_T)$ is the set of couplings between measures π_0 and π_T. This metric leverages the Euclidean distance between two samples to a distance between two distributions. Other metrics can be used instead but we choose a widely common and robust metric in these types of problems. An instance of a possible coupling is depicted in Fig. 5.1.

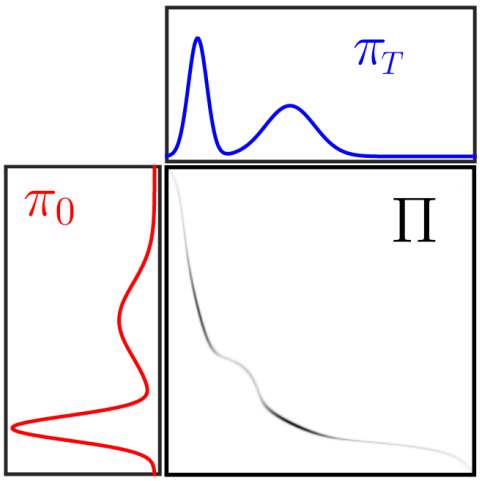


Figure 5.1: Modified illustration of a coupling between continuous measures. Original source [Peyre *et al.* 2019].

5.2 Link between the Retinex theory and the Schrödinger Bridge problem

5.2.1 Entropy-regularized optimal transport and the static Schrödinger Bridge

The transport plan from the previous section is assumed to be computed in the ideal case with infinite resources. However, in practice, solving (5.1) is challenging because the solution has a high computational complexity. Thus, a smoother solution is usually calculated instead thanks to the entropy of the probabilistic coupling [Cuturi *et al.* 2013]. This prior regularizes the solution by slightly increasing the entropy of the plan and lowering its sparsity which makes it easier to compute. We obtain the following problem:

$$\Pi^*_{0,T} = \underset{\Pi \in \mathcal{U}(\pi_0, \pi_T)}{\operatorname{argmin}} \Big\{ \mathbb{E}_{(I_0, I_T)}(\|I_0 - I_T\|_2^2) - \varepsilon \, \mathrm{H}(\Pi) \Big\}, \tag{5.3}$$

$\mathrm{H}(\Pi)$ being the entropy of the transport plan weighted by ε. Increasing ε accelerates computational algorithms and leads to faster convergence [Peyre *et al.* 2019]. Letting $\varepsilon \to 0$, we recover the previous problem (5.1).

Introducing the Gibbs distribution with a squared $L2$ cost

$$\mathrm{d}\mathcal{K}(x,y) = e^{\frac{-\|x-y\|_2^2}{\varepsilon}}\,\mathrm{d}\pi_0(x)\,\mathrm{d}\pi_T(y) \tag{5.4}$$

((π_0, π_T) are represented as probability measures here), we can reformulate (5.3) as

$$\Pi_{0,T}^{\mathrm{SB}} = \underset{\Pi\in\mathcal{U}(\pi_0,\pi_T)}{\operatorname{argmin}}\ \mathrm{KL}(\Pi|\mathcal{K}). \tag{5.5}$$

This is the formulation of the static Schrödinger problem [Léonard *et al.* 2014a; Léonard *et al.* 2012]. Schrödinger [Schrödinger *et al.* 1931] first defined its equivalent dynamic form (5.15) that we detail below.

5.2.2 Link to the Retinex theory

In the last chapter, we showed we can only estimate the degraded reflectance R_0 from the day and night domains dubbed the *common information*. In this way, the reflectance also needs to be restored with a specific correction.

Rephrasing the conclusion with a more formal demonstration, we consider Λ_{scene} the set of wavelengths of the rays projected onto the scene (i.e. the rays forming the illumination), reflected by the materials and captured by the sensor, Λ_{visible} the set of the wavelengths of the visible spectrum, $\hat{R}_{\text{scene}}$ the reflectance estimated from an image, R_{visible} the "true" reflectance of the scene if illuminated by a white light (i.e. Λ_{visible}). We have

$$\hat{R}_{\text{scene}} = \begin{cases} R_{\text{visible}} = R_T, & \text{if } \Lambda_{\text{scene}} = \Lambda_{\text{visible}} \\ R_0, & \text{otherwise } (\Lambda_{\text{scene}} \subset \Lambda_{\text{visible}}) \end{cases} \tag{5.6}$$

hence R_0 is completely determined by R_T. It implies that

$$\Lambda_{\text{scene}} \subset \Lambda_{\text{visible}} \Rightarrow \Omega_{R_0} \subset \Omega_{R_T} \Rightarrow \mathrm{H}(R_0|R_T) = 0 \tag{5.7}$$

where Ω_{R_0} is the set of all possible values of R_0 and Ω_{R_T} the set of all possible values of R_T. The mutual information between the variable of day images and the one of night

images thus becomes

$$\begin{aligned} \mathrm{I}(I_0; I_T) &:= \mathrm{I}(R_0; R_T) && (5.8)\\ &= \mathrm{H}(R_0) - \mathrm{H}(R_0|R_T) \\ &= \mathrm{H}(R_0). && (5.9) \end{aligned}$$

Therefore, (5.3) can be rephrased as

$$\begin{aligned} \Pi_{0,T}^{\mathrm{SB}} &= \underset{\Pi \in \mathcal{U}(\pi_0, \pi_T)}{\operatorname{argmin}} \Big\{ \mathbb{E}_{(I_0, I_T)}(\|I_0 - I_T\|_2^2) + \varepsilon \, \mathrm{KL}(\Pi | \pi_0 \otimes \pi_T) \Big\} && (5.10)\\ &= \underset{\Pi \in \mathcal{U}(\pi_0, \pi_T)}{\operatorname{argmin}} \Big\{ \mathbb{E}_{(I_0, I_T)}(\|I_0 - I_T\|_2^2) + \varepsilon \, \mathrm{I}(I_0; I_T) \Big\} && (5.11)\\ &= \underset{\Pi \in \mathcal{U}(\pi_0, \pi_T)}{\operatorname{argmin}} \Big\{ \mathbb{E}_{(I_0, I_T)}(\|I_0 - I_T\|_2^2) + \varepsilon \, \mathrm{H}(R_0) \Big\}, && (5.12) \end{aligned}$$

where $\pi_0 \otimes \pi_T$ is an independent coupling. We can conclude that the entropy prior for the static Schrödinger problem in the context of the Retinex decomposition boils down to

$$-\mathrm{H}(\Pi_{0,T}^{\mathrm{SB}}) = \mathrm{KL}(\Pi_{0,T}^{\mathrm{SB}} | \pi_0 \otimes \pi_T) = \mathrm{I}(I_0; I_T) = \mathrm{H}(R_0). \tag{5.13}$$

5.2.3 The dynamic form of the Schrödinger bridge

As the obtained plans from (5.1) and (5.3) are joint distributions in the product space of the measures, they are inherently one-step plans.

Instead, if we consider gradually morphing over time the initial measure π_0 into the final measure π_T, then the plans have a dynamic form. We can introduce the time variable $t \in [0, T]$ to model that property. In our case, this dynamic version corresponds to the effect of a vector field v on the initial measure which will follow the minimal length path to reach the final distribution at the last instant joining pairs of points from the two distributions. Finding the optimal plan amounts to finding this path in the space of all possible paths $\tau : [0, T] \mapsto \mathbb{R}^d$. This framework was first introduced by Benamou *et al.* [Benamou *et al.* 2000]. Let $T = 1$ be the final time step and $\mathbb{P}_t$ a path probability measure. The velocity field is said to generate this probability path if it obeys

the continuity equation (5.14) [Villani *et al.* 2009]:

$$\frac{\partial \mathbb{P}_t(x)}{\partial t} + \boldsymbol{\nabla} \cdot \mathbb{P}_t(x) v_t(x) = 0 \tag{5.14}$$
$$\mathbb{P}_0 = \pi_0, \quad \mathbb{P}_1 = \pi_1$$

where $\boldsymbol{\nabla}\cdot$ is the divergence operator defined with respect to a spatial variable $x \in \mathbb{R}^n$ (i.e. $\boldsymbol{\nabla}\cdot := \sum_{i=1}^{n} \frac{\partial}{\partial x_i}$). Intuitively, if we move away from the solutions of this equation $\boldsymbol{\nabla} \cdot \mathbb{P}_t(x) v_t(x) > 0 \Rightarrow \frac{\partial \mathbb{P}_t}{\partial t} < 0 \Rightarrow \mathbb{P}_t(x)$, the mass decreases. On the contrary, if we move closer to these solutions $\boldsymbol{\nabla} \cdot \mathbb{P}_t(x) v_t(x) < 0 \Rightarrow \frac{\partial \mathbb{P}_t}{\partial t} > 0 \Rightarrow \mathbb{P}_t(x)$, the mass increases. This is the essential law to conserve the mass between the two marginals at $t \in \{0, 1\}$.

In this framework, Leonard [Léonard *et al.* 2014a; Léonard *et al.* 2012] showed that (5.5) is the equivalent of the original dynamic problem by Schrödinger [Schrödinger *et al.* 1931]:

$$\mathbb{P}^{\mathrm{SB}} = \underset{\mathbb{P}}{\operatorname{argmin}} \{\mathrm{KL}(\mathbb{P}|\mathbb{Q}) : \mathbb{P}_0 = \pi_0, \mathbb{P}_1 = \pi_1\} \tag{5.15}$$

Schrödinger sought to find the most likely evolution of a particle between two distributions with respect to a reference path measure $\mathbb{Q}$ associated to a stochastic process. In the rest of this chapter, we consider as stochastic process, a Markovian diffusion process modelled by score-based generative models (i.e. diffusion models).

5.3 The Schrödinger Bridge with score-based generative models

5.3.1 The formulation of the problem

Score-based generative models (i.e. diffusion models) have recently emerged to be a better alternative to the state-of-the-art GANs in inverse problems [Song *et al.* 2019; Song *et al.* 2021]. They are more stable to train, more explainable, achieve higher quality samples and better fit multimodal data distributions at a higher computational expense.

Yuan *et al.* [Yuan *et al.* 2022] have already applied them to low-light restoration. They seek in their work to restore the sky to better identify the stars. In their framework, the denoising process is conditioned on the augmented low-light image. The authors randomly

apply Gaussian noise, Gaussian blur, and cutout. The ground-truths of the images are provided in their dataset. Apart from this attempt, very few papers have treated this problem with these models.

Diffusion models had the same goal as GANs at first: to generate an image from an input random noise. Deep neural networks were trained to reverse a diffusion process. This diffusion process consisted of degrading an image by iteratively adding a Gaussian noise until the resulting noise follows the standard normal distribution. The forward diffusion process was thus defined as follows

$$\mathrm{d}I_t = g_t(I_t) + \sigma_t \, \mathrm{d}B_t \tag{5.16}$$

$g_t(\cdot)$ being the drift coefficient, σ_t the diffusion coefficient and B_t a multivariate standard Brownian motion. The first chosen parameter values in the literature were $g_t(I_t) = -\frac{1}{2}I_t$ and $\sigma_t = 1$ giving an Ornstein-Uhlenbeck process. Reversing this diffusion process then leads to gradually denoising the image at each step and generating data from the desired distribution at the end of the Markov process. Hence, both the forward diffusion process and the backward one correspond to bridges between the standard normal distribution and the data distribution.

There are two main ways to reverse such SDE (5.16). *Time-reversal* is the first and original way to do so [Anderson *et al.* 1982]. The reverse SDE is obtained by a strict reversal of every intermediate drift at each instant. *Bridge Matching* is the second approach. It is a relaxed version of time-reversal where the reverse SDE only has to match the marginals at each instant of the forward process [Liu *et al.* 2022]. The resulting SDE thus mimics the real backward process. In practice, bridge matching makes the problem easier to solve and even leads to higher quality generated samples. Therefore, we consider this method in the rest of the chapter.

We first focus on learning how to simulate the forward process. Regarding our problem (5.15), $\mathbb{Q}_{|0,1}$ is defined as the reference path Markov measure corresponding to the diffusion process below (5.17). The Doob's h-transform can be used to condition a stochastic process to hit a particular value at a specific time [Øksendal *et al.* 2003]. Therefore, the diffusion process is conditioned at initial *and* terminal points (I_0, I_1) (hence the notation

$\mathbb{Q}_{|0,1}$).

$$\begin{aligned} \mathrm{d}I_t^{0,1} &= \{f_t(I_t^{0,1}) + \sigma_t^2 \nabla \log \mathbb{Q}_{1|t}(I_1|I_t^{0,1})\}\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t \\ I_0^{0,1} &= I_0, \quad I_1^{0,1} = I_1. \end{aligned} \tag{5.17}$$

where $\nabla \log \mathbb{Q}_{1|t}(I_1|I_t^{0,1})$ is the score function (i.e. the *gradient of the log density of the distribution*) we learn with a neural network. Integrating $\mathbb{Q}_{|0,1}$ over the independent coupling $\pi_0 \otimes \pi_1$ such that $\mathbb{P} = \int \mathbb{Q}_{|0,1}\,\mathrm{d}\pi_0 \otimes \mathrm{d}\pi_1$ gives us $\mathbb{P}$, a non-Markov mixture of bridges. The intuition behind the idea of solving (5.15) with the diffusion process (5.17) is that it amounts to finding the best Markov process approximation which matches the marginals $\mathbb{P}_t$ for each t.

To define the loss function and learn the score function, we have to specify which type of bridge we want to simulate. We choose a Brownian bridge following the methodology in [Shi *et al.* 2023] with a time-homogeneous stochasticity. It is the simplest continuous-time stochastic process such that there is no noise at time $t \in \{0, T\}$. To the best of our knowledge, more complex bridges have not been used yet to this purpose. It is not intuitive if they could guarantee a better generative process though. Hence, $f_t(I_t^{0,1}) = 0, \sigma_t = \sigma$ in (5.17). We can now reformulate this equation as

$$\mathrm{d}I_t^{0,1} = \sigma_t^2 \nabla \log \mathbb{Q}_{1|t}(I_1|I_t^{0,1})\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t \tag{5.18}$$

$$= v(t, I_t^{0,1})\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t \tag{5.19}$$

$$= \frac{I_1 - I_t^{0,1}}{1-t}\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t \tag{5.20}$$

$$I_t^{0,1} = tI_1 + (1-t)I_0 + \sigma\sqrt{t(1-t)}Z, \quad Z \sim \mathcal{N}(0, \mathrm{Id}). \tag{5.21}$$

(5.21) enables us to sample from the bridge. Intuitively, a Gaussian noise is added to the linear interpolation between one sample from the low-light domain and one from the normal-light domain sampling t from $t \sim \mathcal{U}(0,1)$. We emphasize that like the regular Brownian bridges, the uncertainty on the intermediate representation $I_t^{0,1}$ is maximal at $t = 0.5$.

To learn the drift term of (5.20), a regular convolutional neural network (U-Net with

skip attentions and time embedding) with parameters θ is fitted with the loss

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left\{ \int_0^1 \omega(\sigma_t, t)\mathcal{L}(\theta)/\sigma_t^2 \, \mathrm{d}t \right\} \tag{5.22}$$

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbb{P}_{t,1}} \left[\|\sigma_t^2 \nabla \log \mathbb{Q}_{1|t}(I_1|I_t^{0,1}) - v_\theta(t, I_t^{0,1})\|_2^2 \right] \tag{5.23}$$

$$= \mathbb{E}_{(I_0,I_1)\sim\mathbb{P}_{0,1}, Z\sim\mathcal{N}(0,\mathrm{Id})} \left[\|I_1 - I_0 - \sigma_t \sqrt{t/(1-t)} Z - v_\theta(t, I_t^{0,1})\|_2^2 \right] \tag{5.24}$$

$$\omega(\sigma, t) = \frac{1}{1 + \sigma^2 t/(1-t)} \tag{5.25}$$

where ω is a weight scaling the loss preventing the Gaussian noise to dominate the other terms as $t \to 1$. Even though the two samples are not a true pair degraded/ground-truth image, the conditioned score function converges in expectation over a big enough number of samples towards the true unconditioned score function [Hyvärinen *et al.* 2005].

In practice, however, the computations of the score function have multiple sources of bias. First, the estimation of the score function is not flawless and the loss (5.23) is not null. We explain in Section 5.3.2 one procedure which reduces this bias. Second, in high dimensions, the samples are scattered which makes the resulting empirical distribution too sparse. Hence, the estimations of the image manifolds are not accurate. The empirical distributions have an intrinsic bias which skews the computations of the bridges. The obtained bridges only join observed samples but do not generalize beyond the datasets. This is a well-known problem in optimal transport and machine learning in general. Manifold estimation is still an active research field. We study in Section 5.3.3 diverse regularization terms which can mitigate this problem.

Alternative methods in the framework of optimal transport are currently being explored. These approaches are based on simulating ODEs instead of SDEs. They are called Flow Matching methods [Lipman *et al.* 2023; Liu *et al.* 2023]. These straight (or rectified) flows [Liu *et al.* 2023] are a special case of the SDEs when $\sigma_t \to 0$ in the previous equations. They are a generalization of the Normalizing Flows approaches in essence as they do not impose constraints on the latent distribution nor the architecture of the neural networks involved. There is no definitive consensus on whether SDEs or ODEs are strictly better than the other in the problem of generative modelling. We chose to follow the work of Shi *et al.* [Shi *et al.* 2023] for the following reasons.

According to Leonard [Léonard *et al.* 2014a], the uniqueness property of the solution can only hold if $\sigma > 0$. That implies that the solutions obtained with ODEs are not optimal and do not solve the Schrödinger Bridge problem. In the empirical studies they made, Shi *et al.* [Shi *et al.* 2023] show that there is a trade-off between the number of discretization steps (i.e. 1 in the case of ODEs) and the accuracy of the sampling procedure. The authors also find out that as $\sigma \to 0$, the KL divergence of (5.15) increases which in turn alters the number of iterations required to solve the problem.

Some counterarguments exist to favor ODEs to SDEs. Indeed, as stated by Liu *et al.* [Liu *et al.* 2023], learning to simulate an ODE with a neural network leads to faster inference and avoid the expensive computations of the neural drift at each time step. Indeed, there are practical first-order numerical procedure for solving ODEs such as the Euler method. A single step with this method leads to the solution

$$I_1 = I_0 + v_\theta(0, I_0). \tag{5.26}$$

5.3.2 Procedures to reduce the bias in the solutions

To simulate the learned Markov process, we have to sample from a neural network several times. However, the score function that it learned is only approximated and an error still remains. Therefore, the accumulated errors through the whole Markov process creates a heavy bias that needs to be diminished.

In the literature, the traditional approach to solve the Schrödinger Bridge problem (5.15) is the Iterative Proportional Fitting algorithm (IPF) [Kruithof *et al.* 1937; Fortet *et al.* 1940]. De Bortoli *et al.* [De Bortoli *et al.* 2021] improve it at high dimensions introducing diffusion models. Their method consists of learning two co-dependent processes (i.e. one forward and one backward) with two respective neural networks thanks to the time-reversal of the forward SDE. They define their IPF algorithm with the recursion

$$\mathbb{P}^{2n+1} = \underset{\mathbb{P}}{\operatorname{argmin}}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n}) : \mathbb{P}_1 = \pi_1\} \tag{5.27}$$

$$\mathbb{P}^{2n+2} = \underset{\mathbb{P}}{\operatorname{argmin}}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n+1}) : \mathbb{P}_0 = \pi_0\} \tag{5.28}$$

Since the bias accumulates during the simulation of the SDE, it is lower in the first steps, therefore learning co-dependent processes effectively manage to reduce it.

Shi *et al.* build on the previous method to propose the Iterative Markov Fitting procedure (IMF) [Shi *et al.* 2023]. They chose the bridge matching way of reversing the SDE (5.15). Relaxing the strict constraint resulting from time-reversal, IMF preserves the initial and final distributions unlike IPF.

The IMF algorithm starts by computing a first mixture of bridges $\mathbb{P}^0 = \int \mathbb{Q}_{|0,1}\, \mathrm{d}\pi_0 \otimes \mathrm{d}\pi_1$ with samples from the original initial and final distributions thanks to (5.21). Multiple Brownian bridges are constructed between batches of observed samples, the mixture of these bridges thus link the two distributions. The closest backward Markov process to this mixture is then learned by minimizing the corresponding loss (i.e. (5.23) adapted to the backward process). Let the Markov measure associated to this learned process be $\mathbb{M}^1$. A new coupling $\mathbb{M}^1_{0,1}$ is obtained by simulation of the SDE with the learned backward drift v_ϕ. Thus, a new mixture of bridges can be created by sampling from $\mathbb{P}^1 = \int \mathbb{Q}_{|0,1}\, \mathrm{d}\mathbb{M}^1_{0,1}$. Since this mixture of bridges depends on the backward network, a forward network v_θ is fitted with (5.23) which gives a new Markov measure $\mathbb{M}^2$ with a reduced bias. The loop ends with the processing of a new coupling $\mathbb{M}^2_{0,1}$ from which a mixture of bridges can be calculated. These operations are iterated for N steps.

In theory, the IMF boils down to solving the following problems iteratively,

$$\mathbb{P}^{2n+1} = \underset{\mathcal{M}}{\mathrm{proj}}(\mathbb{P}^{2n}) = \underset{\mathbb{P}}{\mathrm{argmin}}\{\mathrm{KL}(\mathbb{P}^{2n}|\mathbb{P}) : \mathbb{P} \in \mathcal{M}\} \tag{5.29}$$

$$\mathbb{P}^{2n+2} = \underset{\mathcal{R}(\mathbb{Q})}{\mathrm{proj}}(\mathbb{P}^{n+1}) = \underset{\mathbb{P}}{\mathrm{argmin}}\{\mathrm{KL}(\mathbb{P}^{2n+1}|\mathbb{P}) : \mathbb{P} \in \mathcal{R}(\mathbb{Q})\}. \tag{5.30}$$

(5.29) gives the closest Markov measure (in the set of all Markov measures $\mathcal{M}$) to the mixture of bridges $\mathbb{P}^{2n}$ by training a diffusion model. The second problem (5.30) is equivalent to finding the closest mixture of bridges to the Markov measure $\mathbb{P}^{2n+1}$ in the reciprocal class of $\mathbb{Q}$ (i.e. a path measure whose bridge starts and ends in the same marginal distributions as $\mathbb{Q}$ [Léonard *et al.* 2014b]).

In practice, to get a new coupling from an SDE with a learned drift, the SDE is

discretized with the Euler–Maruyama approximation for an arbitrary number of diffusion steps (i.e. empirically 30 steps leads to high quality results [Shi *et al.* 2023]). After applying this SDE to several batches, the resulting generated images are cached. They are now considered as the new temporary dataset for the next iteration.

Shi *et al.* [Shi *et al.* 2023] also provide the results of their experiments concerning the study of the properties of the IMF algorithm with diffusion models. Introducing a bit of stochasticity in the SDE accelerates the convergence of the algorithm making the marginals smoother. Besides, they hypothesize that the optimal amount of noise is dependent on the task. Finally, their experiments tend to show that the error of the IMF algorithm with diffusion models is linearly proportional to the number of dimensions of the distributions.

5.3.3 Regularizations for the Schrödinger Bridge

To tackle the second source of bias and improve the generalization of the solutions beyond observed samples, diverse priors can be added to the computations of the loss. Kim *et al.* were among the first to do so in their work [Kim *et al.* 2023]. They use the Contrastive Unpaired Translation (CUT) [Park *et al.* 2020] method well-known in the style transfer literature but only applied to GANs before. It consists of two priors: a contrastive learning-based prior and an adversarial term. However, they do not use the IMF algorithm and only learn one network in their approach.

Contrastive regularization

As usual in the style transfer literature, Park *et al.* [Park *et al.* 2020] seek in their CUT scheme to preserve the content of an image and align its style to a target. Their idea is to maximize mutual information between the input image and the generated output thanks to a patch-wise contrastive matching approach. As depicted in Fig. 5.2, two corresponding patches should be closely encoded in the latent space to obtain style-independent features while repelling any other patch. These repelled "negative" patches are extracted from the same image as it led to better empirical results. The InfoNCE [Oord *et al.* 2019] loss is commonly used to achieve this. The InfoNCE loss is equivalent to a classification of a patch from the generated image such that the corresponding ground-truth input patch is the associated class to the contrary of the "negative" patches. It boils down to minimizing

the following cross-entropy

$$\mathcal{L}_{\text{NCE}}(x, y^+, y^-) = -\log\left[\frac{\exp(x \cdot y^+/\rho)}{\exp(x \cdot y^+/\rho) + \sum_{n=1}^{N} \exp(x \cdot y_n^-/\rho)}\right] \tag{5.31}$$

where ρ is a temperature parameter, (x, y^+, y^-) are respectively the input patch, the "positive" ground-truth patch and the negative patches as vectors. Patches are normalized and then encoded to representations in the latent space thanks to a two-layer multilayer perceptron. Then, the contrastive loss is applied. The authors of the CUT method use this loss at several scales to maximize the mutual information independently of the scale. Interestingly, Oord *et al.* proved in their work [Oord *et al.* 2019] the inequality regarding two arbitrary input variables (X, Y)

$$\text{I}(X; Y) \geq \log(N) - \mathcal{L}_{\text{NCE}}(x, y^+, y^-), \tag{5.32}$$

N being the number of samples. Thus, minimizing this loss effectively maximizes the mutual information between the provided variables. Moreover, an increase in the number of samples also results in the rise of this information.

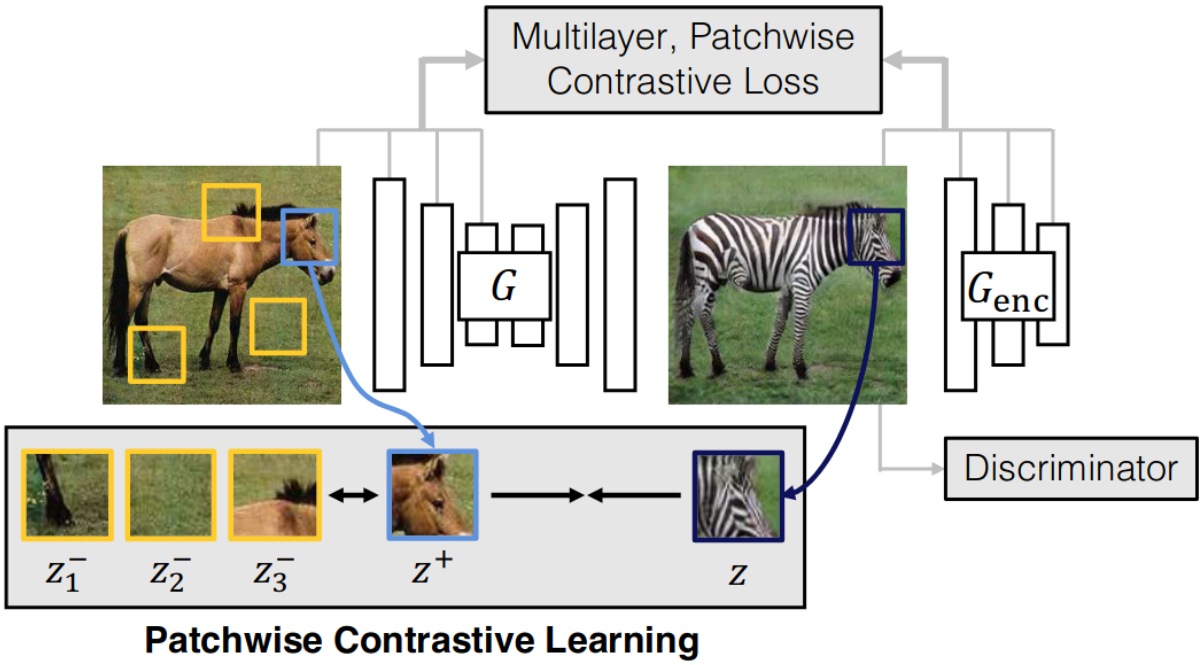


Figure 5.2: Illustration of the patch-based contrastive learning method in Contrastive Unpaired Translation (CUT) [Park *et al.* 2020].

In the context of the Schrödinger Bridge, Kim *et al.* [Kim *et al.* 2023] maximizes the mutual information between I_0 and $I_{t+1}^{0,1}$, the intermediate representation at time $t+1$ in the sampling of the Markov process with a neural network. It can be seen as the dynamic form of the negative prior $-\mathrm{I}(I_0; I_1)$ added to (5.3). It boils down to the following problem in the forward process

$$\min_{(I_0, I_t^{0,1})\sim\mathbb{M}_{0,1}} \int_0^1 \mathcal{L}_{\mathrm{NCE}}(x_{I_t^{0,1}}, y_{I_0}^+, y_{I_0}^-)\,\mathrm{d}t \Leftrightarrow \max_{(I_0, I_t^{0,1})\sim\mathbb{M}_{0,1}} \int_0^1 \mathrm{I}(I_0; I_t^{0,1})\,\mathrm{d}t \tag{5.33}$$

$$\Leftrightarrow \max_{I_0\sim\pi_0, I_1\sim\pi_1} \mathrm{I}(I_0; I_1). \tag{5.34}$$

They stop the calculations of the gradient at time t to reduce the required memory, preventing the computations between $[0, t]$. This way, they only learn the drift at only one time step t.

Adversarial term

To ensure that the output is visually similar to images of the target domain, an additional discriminator D_ψ is used in CUT. The associated loss term is based on the least squares generative adversarial networks [Mao *et al.* 2017]. In their work, Mao *et al.* found that a L^2-norm makes the training of the GANs more stable while improving the quality of the generated outputs. Let a be the label of the fake data and b the one of the true samples. In the work of Kim *et al.* [Kim *et al.* 2023], it leads to the loss term

$$\mathcal{L}_{\mathrm{Adv}}(D_\psi) = \frac{1}{2}\mathbb{E}_{I_1\sim\pi_1}\left[\|D_\psi(I_1) - b\|_2^2\right] + \frac{1}{2}\mathbb{E}_{I_t\sim\mathbb{M}_{0,1}}\left[\|D_\psi(v_\theta(t, I_t) + I_t) - a\|_2^2\right] \tag{5.35}$$

$$\mathcal{L}_{\mathrm{Adv}}(v_\theta) = \frac{1}{2}\mathbb{E}_{I_t\sim\mathbb{M}_{0,1}}\left[\|D_\psi(v_\theta(t, I_t) + I_t) - b\|_2^2\right] \tag{5.36}$$

where $\mathbb{M}_{0,1}$ is the coupling obtained with the simulation of the SDE with the drift term given by a neural network. The ψ parameters are fixed in the optimization of (5.36) and θ are fixed when computing (5.35) like in the regular GAN framework.

Retinex priors

Concerning our problem, we seek to introduce physical priors to regularize the computations of the Schrödinger Bridge in IMF. This should ease the approximation of the bridge with true explicable and consistent priors with the physics of light. It could enable the algorithm to potentially reach more accurate solutions even out of the observed

distributions. Recall (5.9), $\mathrm{I}(I_0; I_1) = \mathrm{H}(R_0)$, the entropy of the low-light reflectance.

We maximize the "dynamic" prior $\mathrm{I}(I_0^{0,1}; I_{t+1}^{0,1})$ like in (5.34) as it is equivalent to maximizing the average of the entropy of the reflectance along the path

$$\min_{(I_0, I_t^{0,1}) \sim \mathbb{M}_{0,1}} \int_0^1 \mathcal{L}_{\mathrm{NCE}}(x_{I_t^{0,1}}, y_{I_0}^+, y_{I_0}^-) \, \mathrm{d}t \Leftrightarrow \max_{(I_0, I_t^{0,1}) \sim \mathbb{M}_{0,1}} \int_0^1 \mathrm{H}(R_t) \, \mathrm{d}t \tag{5.37}$$

thus following the methodology of Kim *et al.* [Kim *et al.* 2023] and the CUT framework. Adding this prior can seem contradictory with regard to (5.11). Our intuition behind these two opposing priors is that the estimated reflectance $\mathrm{H}(R_0)$ in (5.11) is biased from the observed samples. This is mitigated by increasing the entropy of the variable (5.37). This follows Jaynes' principle of maximum entropy in the Bayesian interpretation of the prior [Jaynes *et al.* 1988]. We do not make as few assumptions as possible on the distribution of the variable. We also seek to generalize beyond the empirical distributions to reach the true restored versions of the input images rather than the perfect transport plan between the observed samples. Estimating image manifolds is known in the literature to be made difficult by the curse of dimensionality. One needs exponentially more samples to estimate them at high dimensions and regularization of the optimization problem is seen as one of the solutions for that.

Our first idea was to learn a Retinex decomposition network as well as diffusion models to simulate one Schrödinger Bridge for each component for the forward pass and likewise for the backward pass. This can be done simultaneously in practice. However, since both initial and final distributions are not stationary over the optimization, the networks could not converge to stable solutions. We therefore return to the original single full-image bridge combined with physical regularizations terms. One interesting future direction could be to model this additional drift term in the decomposition.

We decided to focus on the dynamical form of optimal transport as the existing numerical algorithms give more stable solutions and are computationally efficient. In our problem, we use the IMF algorithm and the CUT scheme as well as the architecture in Chapter 4 (without the discriminators) to decompose each representation along the bridge and apply Retinex priors. We highlight the fact that no neural network could be trained beforehand to extract the reflectance as the mutual information from the two

static domains would only extract the degraded reflectance. Moreover, extracting the mutual information from two images of the same domain, $\mathrm{I}(I_0; I_0)$ would only estimate the entropy $\mathrm{H}(I_0)$ (or $\mathrm{H}(I_1)$ respectively). $\mathrm{H}(R_1)$ can not be estimated this way. Hence the impossibility to use it in the regularization of the transport of daylight images to nighttime images.

We therefore seek to fit three networks for both forward and backward processes. Let f_ϕ, v_θ, D_ψ be the networks respectively, the Retinex decomposition model, the diffusion model learning the forward bridge and the CUT discriminator. We name in the next sections the process that goes from nighttime to daylight images, the forward process, and the opposite, the backward process, to remain consistent with the notation so far. To train $f_\phi : I \in [0,1]^{3n} \mapsto (R \in [\varepsilon, 1-\varepsilon]^{3n}, L \in [0,1]^{3n})$, we optimize the PatchNCE loss (5.31) as a substitute of the previous discriminators in the architecture as well as the regular Retinex priors.

We first recall the Retinex priors we used in Chapter 4 in our current settings

$$\mathcal{L}_{HR} = \left\| \frac{1}{\alpha} \right\|_1 \tag{5.38}$$

$$\mathcal{L}_{E} = \sum_{c \in \{R,G,B\}} \left\| g(\tilde{L}_t) - g(L_{c,t}) \right\|_2^2 \tag{5.39}$$

$$\mathcal{L}_{MAE} = \left\| I_t^\gamma - L_t . * R_t \right\|_1 \tag{5.40}$$

$$\mathcal{L}_{color} = \frac{L_t \cdot R_t}{\|L_t\| \|R_t\|} \tag{5.41}$$

where g is a threshold function $g(x) = \begin{cases} x, & x > 1-\epsilon \\ 0, & \text{otherwise} \end{cases}$ and $\tilde{L}_t = \max_{c \in \{R,G,B\}} I_{c,t}$, the grayscale approximation of the illumination. We train the Retinex decomposition network with the following problem

$$\begin{aligned} \phi_f^* = \underset{\phi}{\operatorname{argmin}} \int_0^1 \Big[& \lambda_{\mathrm{NCE}} \mathcal{L}_{\mathrm{NCE}}(x_{R_t}, y^+_{R_0}, y^-_{R_0}) + \lambda_{MAE} \mathcal{L}_{MAE} + \lambda_{color} \mathcal{L}_{color} \\ & + \lambda_E \mathcal{L}_E + \lambda_{HR} \mathcal{L}_{HR} \Big] \, \mathrm{d}t \end{aligned} \tag{5.42}$$

R_1 is used in the backward pass.

While we train the components extractor, we simultaneously fit the score-based generative network to learn the forward Brownian bridge adding the adversarial term (5.36). It results in the following additional loss:

$$\theta_v^* = \underset{\theta}{\operatorname{argmin}} \left\{ \int_0^1 \left[\omega(\sigma_t, t)\mathcal{L}(\theta_v)/\sigma_t^2 + \lambda_{\text{Adv}}\mathcal{L}_{\text{Adv}}(v_\theta) \right] \mathrm{d}t \right\} \tag{5.43}$$

$$\mathcal{L}(\theta_v) = \mathbb{E}_{(I_0,I_1)\sim\mathbb{P}_{0,1},Z\sim\mathcal{N}(0,\text{Id})} \left[\| I_1 - I_0 - \sigma_t\sqrt{t/(1-t)}Z - v_\theta(t, I_t^{0,1}) \|_2^2 \right] \tag{5.44}$$

$$\omega(\sigma, t) = \frac{1}{1 + \sigma^2 t/(1-t)}. \tag{5.45}$$

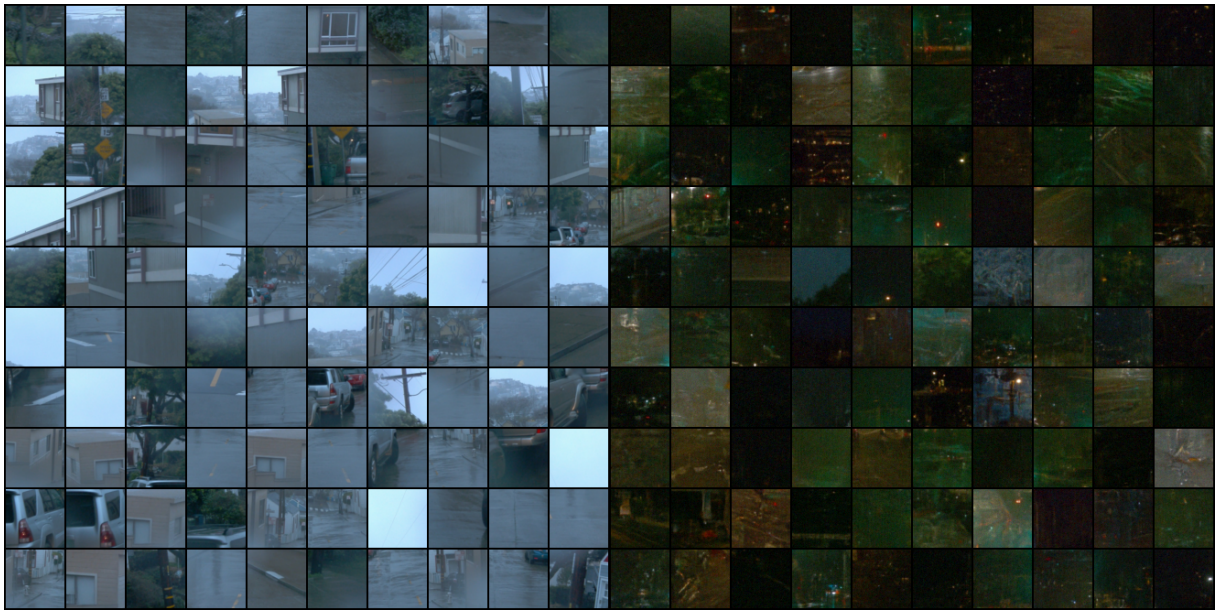
To learn the backward process, three similar networks are trained reversing the endpoints. Finally, the discriminator is fit with (5.35). It helps to approximate the target distribution at higher dimensions. We tried with a lower number of regularization terms removing the adversarial term but it led to lower quality samples in the end.

5.4 Results

5.4.1 IMF

Fig. 5.3 illustrates the output images we get with the IMF algorithm without any regularization term. It produces some visually interesting results especially with simple patches such as road lines. However, we can still see some hallucinations in the details of the darkest or overexposed parts of the images. Some objects also disappear in the outputs. To reduce the complexity of the problem, the method is checked at smaller scales than entire images. Patches are transported instead.

Input images Output images



Input images Output images

Figure 5.3: Depiction of the images we obtain with the IMF algorithm without additional regularization. The top row represents the process from nighttime images to daylight images while the bottom row is the opposite. On the left column, the images are the true samples in the sampling chain while the right column depicts the output images.

5.4.2 CUT Scheme

In Fig. 5.4, we represent the obtained outputs applying the CUT framework to regularize a forward Schrödinger Bridge. Like in the previous section, notable hallucinations are still present. Our scope is to reduce them introducing our physical priors. Instead of training their method on smaller patches, Kim *et al.* scale images to smaller sizes.

Input image Output image

Figure 5.4: Illustration of the output image with the CUT scheme and a forward Schrödinger Bridge like in the work of Kim *et al.* [Kim *et al.* 2023].

5.4.3 Adding the Retinex priors

In this section, we follow the procedure described in Section 5.3.3. We first seek to empirically validate (5.34) and thus (5.37) (i.e. casting our problem in the optimal transport framework) with respect to the intuition of the *common information* that we extract in Chapter 4.

We select samples at the start of the sampling chain ($t = 0$) in both processes. Fig. 5.5 and Fig. 5.6 depict the Retinex components we learn to extract without the adversarial term and adding it, respectively. An observation we can make is that without this term we recover similar components to the ones of Chapter 4. This confirms that training a

decomposition network by maximizing the mutual information between the input sample and the intermediate representation along the sampling chain is equivalent to learning the *common information* between the two domains. However and to our surprise, when we add the discriminators, the components swap places. A further investigation is required to determine the cause of this effect.

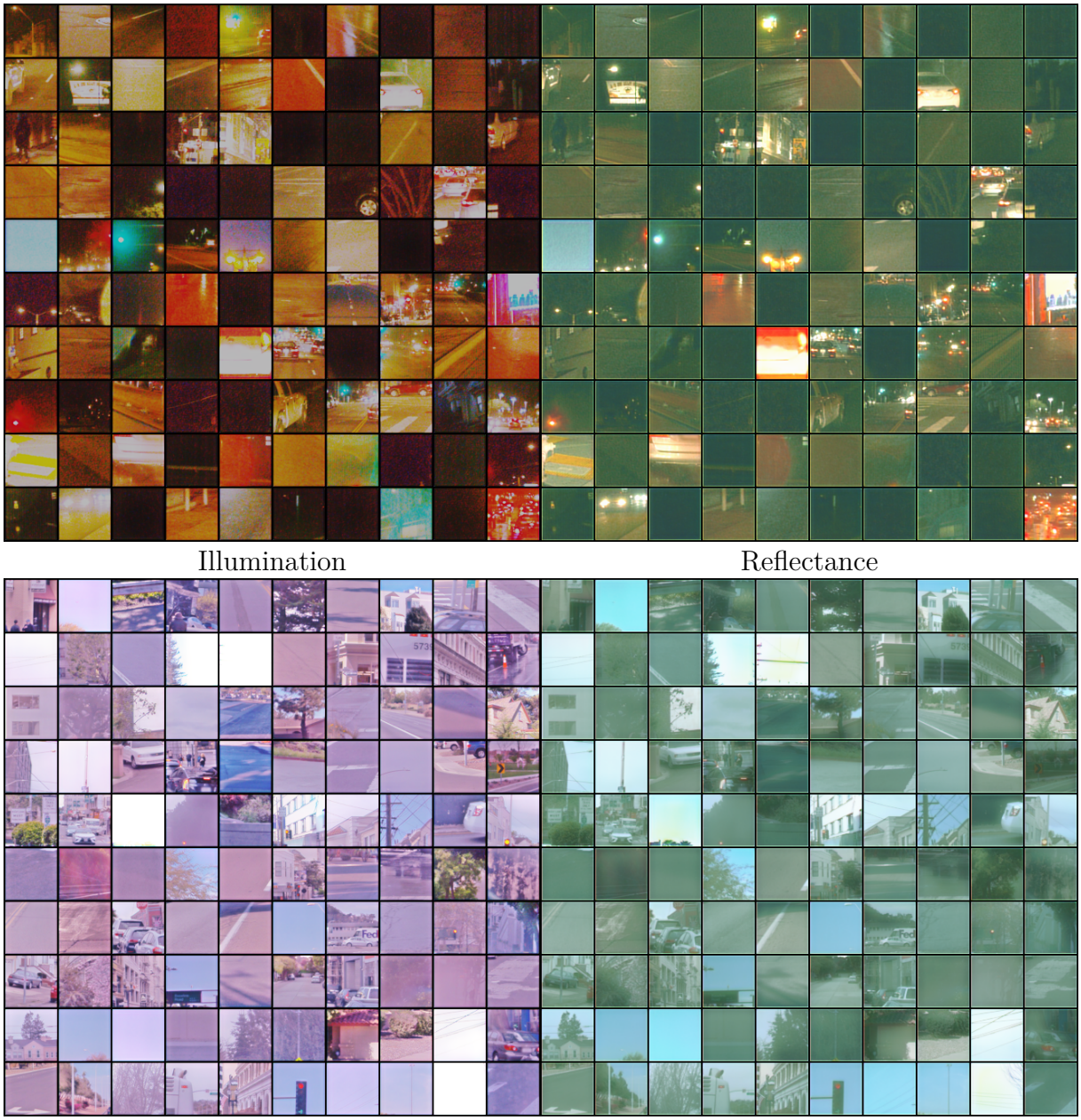


Figure 5.5: Illustration of the Retinex components ($t = 0$) of the sampling chain without the adversarial term going from low-light images to normal-light images (top row) and the reverse process (bottom row).

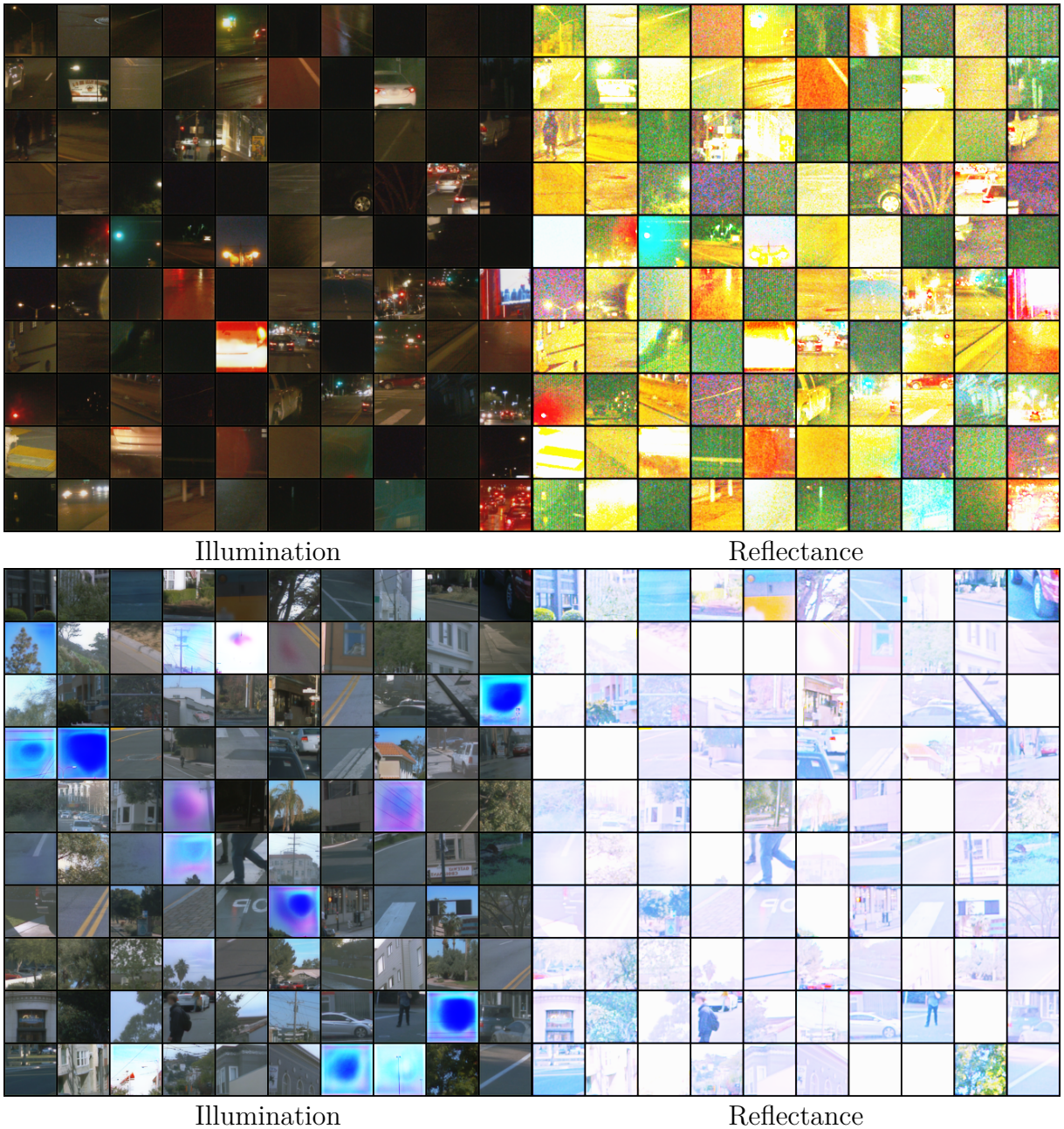


Figure 5.6: Illustration of the Retinex components ($t = 0$) of the sampling chain with the adversarial term going from low-light images to normal-light images (top row) and the reverse process (bottom row).

We now study the output images we obtain with such priors including the adversarial regularization. Fig. 5.7 represents our results. We can see that the two co-dependent

processes diverge in the optimization. This happens with or without the adversarial term. We plan to study this phenomenon as detailed in the perspectives.

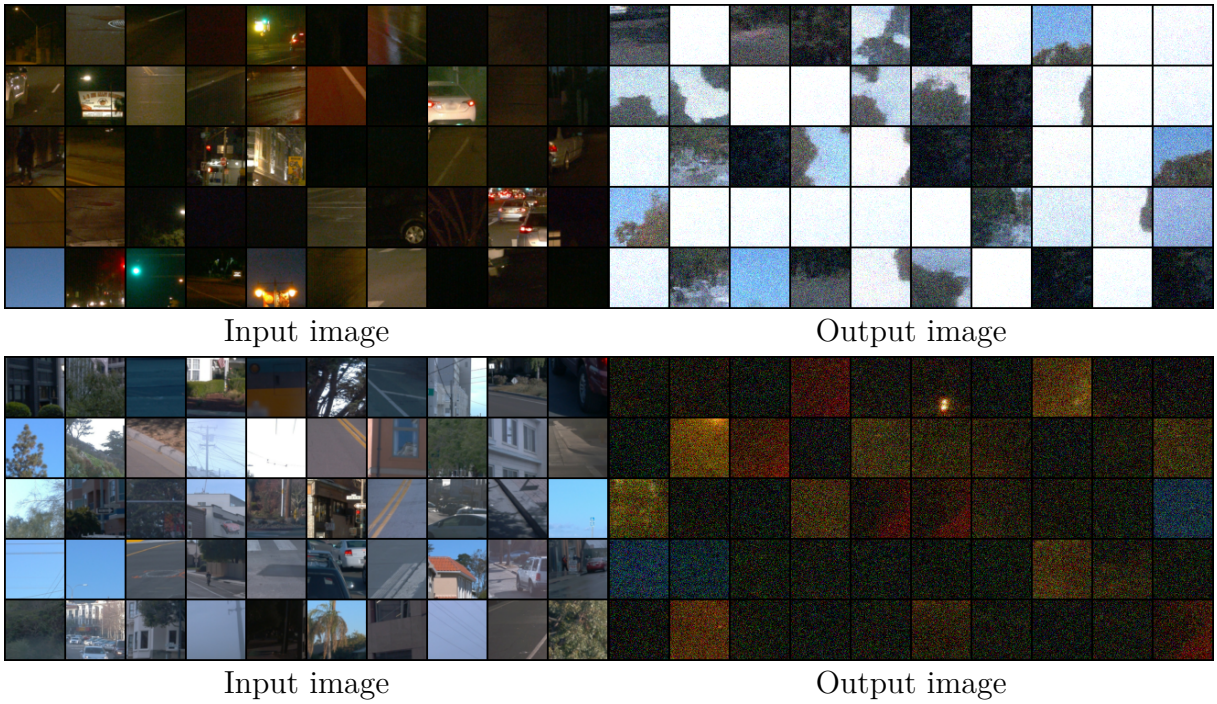


Figure 5.7: Illustration of the generated images after the process going from low-light images to normal-light images (top row) and the reverse process (bottom row).

5.5 Chapter conclusion

In this chapter, we first reviewed the classic Monge-Kantorovich problem and its entropy-regularized version. Then, we showed the link between the Retinex theory and the problem of the static Schrödinger Bridge through the definition of the reflectance. Afterward, we recalled the definition of the dynamic form of the Schrödinger Bridge. We thus proceeded by presenting the recently known score-based generative models. Two state-of-the-art procedures in these types of problems are explained: Iterative Proportional Fitting and Iterative Markov Fitting, both combined with diffusion models. They reduce the intrinsic bias of the learned score function. To further mitigate the bias in the obtained solutions, different regularizations terms coming from the Contrastive Unpaired Translation scheme are detailed. They consist of an adversarial and a contrastive regularization term. We also analyze the state-of-the-art approaches in the context of our problem. The bias still

present in the computations of the bridges result in hallucinations in the output images. We then proposed to add priors based on the Retinex theory to further improve the restoration. They are consistent with the physics of light. We first observed that we can validate the idea of common information in Chapter 4 in the framework of optimal transport by studying one diffusion step at $t = 0$. Our results also raised some questions about the convergence of the algorithm under the Retinex priors.

Conclusion

Summary

In this thesis, we tackled the problem of restoring low-light images. First, we defined the parameters we considered. We identified the degradations of these images and isolated the issues related to the color deviations and the noise. These distortions come from the low number of photons and the presence of colored light sources. These settings are prominent in nighttime images. Besides, the lack of ground-truths in outdoor scenes drove us to explore unsupervised approaches. We also decided to focus on methods which can guarantee that the information we extract in the images is preserved during the restoration: we do not want to only simulate the day scene but seek to avoid adding hallucinations in the content.

In Chapter 3, we considered the commonly found Retinex decomposition model in the literature. A trivial solution in the decomposition process is identified. It may cause it to fail. The difficulty lies in the fact that it is a valid solution in some cases. Indeed, the illumination can be saturated if the input image is saturated as well (i.e. when the image contains a light source or an overexposed area). We proposed a new prior to avoid absurd solutions from the solution set during the optimization process. In an ablation study, we demonstrated the efficiency of our regularization term. We built on deep image priors to propose a method which does not need any dataset to be trained. It achieves on par results with the state-of-the-art supervised methods without halo artifacts.

We investigated the problem of restoring outdoors images without ground-truth in Chapter 4. So far, the Retinex theory was left aside in unsupervised methods. We

went back to the original definitions of its components to propose a new decomposition model with colored illumination consistent with the physics of light. Thus, we relaxed the restrictive prior of a smooth grayscale illumination. The reflectance, on the other hand, is defined as the mutual information between daylight and nighttime images after a specific correction is applied. We proposed a new architecture based on this new model building on style transfer and source separation approaches. It also enables us to visualize the complex style (i.e. the illumination) and the content (i.e. the reflectance). Our method does not add hallucination nor artifact and does not amplify the noise as much as the other state-of-the-arts works. In this chapter, we addressed a difficult problem combining correlated sources to separate while being underdetermined.

Finally, in Chapter 5, we highlighted the link between our objective and the problem of the Schrödinger bridge through mutual information and the definition of the reflectance. Then, we recalled the dynamic form of the Schrödinger Bridge. Multiple approaches to compute this entropy-regularized optimal transport map thanks to diffusion models are presented. We analyzed these methods as well as state-of-the-art priors maximizing the mutual information along the sampling chain in the context of the restoration of low-light images. Since they produce a biased transport map in practice, we also introduced dynamic priors consistent with the physics of light. We validated our intuition of extracting the *common information* from Chapter 4 in our experiments.

Perspectives

We believe that the work carried out in this thesis paves the way to novel interesting research directions.

First, as stated in the introduction, the high level of humidity at night results in blurry and foggy images. Therefore, the first lead would be to integrate these effects combining the Retinex decomposition model with the works [Narasimhan *et al.* 2003; Li *et al.* 2015]. Addressing the scattering effect of the light and consequently the glowing effect of the light sources at night constitutes an exciting direction. This would greatly contribute to improving computer vision algorithms at night in edge cases especially on high-level tasks such as object detection.

Second, we explained in Chapter 2 and Chapter 4 that the Retinex theory holds on linearized data with respect to the intensity of the signal captured by the camera sensor. Thus, reversing the non-linear steps of the camera image processing pipeline is necessary to collect accurate Retinex components. We made a simple (maybe even simplistic) assumption to reverse this process but a lot of work can be done here. Indeed, blindly estimating the whole process from a single input image is extremely challenging but would help to be consistent with the physics of light. Since the majority of the accessible cameras are commercialized with the intent to be sold to the public, the images are heavily modified to please the customers. Besides, most of the proprietary pipelines are not accessible. More specific or even custom sensors could be used instead, of course.

The next step in making the Retinex model even more realistic could also be to combine different modalities of signal such as LIDARs to better grasp the geometry of the scene. Introducing this information could lead to greatly extend the knowledge we could use to restore the images. Besides, more powerful priors could also be defined to consider specularities with existing and more complex model of the reflectance like the BRDF of Phong's model.

Another interesting direction would also be the evaluation of the quality of low-light images. Even though some metrics already measure the differences between two distributions of samples, they are still too coarse. Finer and more specific metrics would greatly contribute to the understanding of the way images are affected by the low number of photons.

Regarding Chapter 5, the main improvements that could be done remain on the different sources of bias. Indeed, a better numerical estimation of the score function would lead to reach the final target image manifold more accurately. The best samples to generate are possibly out of the empirical distribution. Different types of bridges could also be explored instead of the simple Brownian one. Moreover, contributions to the Flow Matching methods would also participate in the development of these approaches. Finally, if one wanted to learn two bridges for each Retinex component, the drift *of the distributions* coming from the extraction of the components could be modeled with an additional term added to the drift *of the SDEs*. This might lead to promising results in the future. Besides, the question of the convergence of the algorithm with the Retinex

priors is still open.

Finding an effective way to gather pairs of degraded/ground-truth outdoor images with the same scene would make it possible to leverage this additional information, significantly reducing the difficulty of the task.

Lastly, in real-time image processing pipelines, the execution time is an important aspect. We studied three types of approaches in this thesis: one based on the deep image prior, the second on generative adversarial networks and the last one on diffusion models. The high calculation time of the deep image prior constitutes its biggest weakness, but it is hard to overcome since it is inherent to the method. GANs are the fastest generative models nowadays and the current state-of-the-art on that aspect. Accelerating the sampling chain of the diffusion models is a highly active research field right now. Some promising directions include using early stopping [Lyu *et al.* 2022] or trajectory prediction [Mao *et al.* 2023]. Instead, other recent works also propose to reduce the size of the diffusion model to accelerate the computations through distillation [Meng *et al.* 2023] or quantization [Shang *et al.* 2023].

BIBLIOGRAPHY

[Afifi *et al.* 2019] Mahmoud Afifi et al., "When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images", *in*: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA: IEEE, June 2019, pp. 1535–1544, ISBN: 978-1-72813-293-8, DOI: 10.1109/CVPR.2019.00163, URL: https://ieeexplore.ieee.org/document/8953936/ (visited on 08/27/2021).

[Afifi *et al.* 2021] Mahmoud Afifi et al., "Learning Multi-Scale Photo Exposure Correction", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9157–9167.

[Anderson *et al.* 1982] Brian D.O. Anderson, "Reverse-Time Diffusion Equation Models", *in*: *Stochastic Processes and their Applications* 12.*3* (May 1982), pp. 313–326, ISSN: 03044149, DOI: 10.1016/0304-4149(82)90051-5, URL: https://linkinghub.elsevier.com/retrieve/pii/0304414982900515 (visited on 11/09/2022).

[Barron *et al.* 2017] Jonathan T Barron and Yun-Ta Tsai, "Fast Fourier Color Constancy", *in*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 886–894.

[Barron *et al.* 2015a] Jonathan T. Barron, "Convolutional Color Constancy", *in*: *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015 IEEE International Conference on Computer

Vision (ICCV), Santiago, Chile: IEEE, Dec. 2015, pp. 379–387, ISBN: 978-1-4673-8391-2, DOI: 10.1109/ICCV.2015.51, URL: http://ieeexplore.ieee.org/document/7410408/ (visited on 08/30/2021).

[Barron *et al.* 2015b] Jonathan T. Barron and Jitendra Malik, "Shape, Illumination, and Reflectance from Shading", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.*8* (Aug. 2015), pp. 1670–1687, ISSN: 1939-3539, DOI: 10.1109/TPAMI.2014.2377712.

[Barrow *et al.* 1978] Harry G. Barrow and J. Martin Tenenbaum, "Recovering Intrinsic Scene Characteristics from Images", *in*: *Computer Vision Systems* (1978).

[Benamou *et al.* 2000] Jean-David Benamou and Yann Brenier, "A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem", *in*: *Numerische Mathematik* 84.*3* (Jan. 1, 2000), pp. 375–393, ISSN: 0029-599X, 0945-3245, DOI: 10.1007/s002110050002, URL: http://link.springer.com/10.1007/s002110050002 (visited on 04/06/2023).

[Borsoi *et al.* 2021] Ricardo Augusto Borsoi et al., "Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review", *in*: *IEEE Geoscience and Remote Sensing Magazine* 9.*4* (Dec. 2021), pp. 223–270, ISSN: 2168-6831, 2473-2397, 2373-7468, DOI: 10.1109/MGRS.2021.3071158, arXiv: 2001.07307 [eess], URL: http://arxiv.org/abs/2001.07307 (visited on 07/06/2022).

[Boss *et al.* 2021] Mark Boss et al., "NeRD: Neural Reflectance Decomposition from Image Collections", *in*: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, Oct. 2021, pp. 12664–12674, ISBN: 978-1-66542-812-5, DOI: 10.1109/ICCV48922.2021.01245, URL: https://ieeexplore.ieee.org/document/9710856/ (visited on 03/08/2023).

[Brooks *et al.* 2019] Tim Brooks et al., "Unprocessing Images for Learned Raw Denoising", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11036–11045.

[Bychkovsky *et al.* 2011] Vladimir Bychkovsky et al., "Learning Photographic Global Tonal Adjustment with a Database of Input/Output Image Pairs", *in*: *CVPR 2011*, IEEE, 2011, pp. 97–104.

[Cook *et al.* 1982] Robert L Cook and Kenneth E. Torrance, "A reflectance model for computer graphics", *in*: *ACM Transactions on Graphics (ToG)* 1.*1* (1982), pp. 7–24.

[Cuturi *et al.* 2013] Marco Cuturi, "Sinkhorn Distances: Lightspeed Computation of Optimal Transport", *in*: *Advances in neural information processing systems* 26 (2013).

[Dabov *et al.* 2007] Kostadin Dabov et al., "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering", *in*: *IEEE Transactions on Image Processing* 16.*8* (2007), pp. 2080–2095, DOI: 10.1109/TIP.2007.901238.

[De Bortoli *et al.* 2021] Valentin De Bortoli et al., "Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling", *in*: *Advances in Neural Information Processing Systems* 34 (2021), pp. 17695–17709.

[Deng *et al.* 2009] Jia Deng et al., "Imagenet: A Large-Scale Hierarchical Image Database", *in*: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Ieee, 2009, pp. 248–255.

[Dowson *et al.* 1982] DC Dowson and BV666017 Landau, "The Fréchet Distance between Multivariate Normal Distributions", *in*: *Journal of multivariate analysis* 12.*3* (1982), pp. 450–455.

[Fortet *et al.* 1940] Robert Fortet, "Résolution d'un Systeme d'équations de M. Schrödinger", *in*: *J. Math. Pure Appl. IX* 1 (1940), pp. 83–105.

[Fubara *et al.* 2020] Biebele Joslyn Fubara, Mohamed Sedky, and Dave Dyke, "RGB to Spectral Reconstruction via Learned Basis Functions and Weights", *in*: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*,

2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA: IEEE, June 2020, pp. 1984–1993, ISBN: 978-1-72819-360-1, DOI: 10.1109/CVPRW50498.2020.00248, URL: https://ieeexplore.ieee.org/document/9150656/ (visited on 07/01/2022).

[Galdran *et al.* 2018] Adrian Galdran et al., "On the Duality Between Retinex and Image Dehazing", *in*: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA: IEEE, June 2018, pp. 8212–8221, ISBN: 978-1-5386-6420-9, DOI: 10.1109/CVPR.2018.00857, URL: https://ieeexplore.ieee.org/document/8578955/ (visited on 05/25/2021).

[Gandelsman *et al.* 2019] Yossi Gandelsman, Assaf Shocher, and Michal Irani, ""Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors", *in*: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019), arXiv: 1812.00467, URL: http://arxiv.org/abs/1812.00467 (visited on 10/09/2020).

[Gatys *et al.* 2015] Leon Gatys, Alexander S Ecker, and Matthias Bethge, "Texture Synthesis Using Convolutional Neural Networks", *in*: *Advances in neural information processing systems* 28 (2015).

[Goodfellow *et al.* 2014] Ian Goodfellow et al., "Generative Adversarial Nets", *in*: *Advances in neural information processing systems* 27 (2014).

[Grosse *et al.* 2009] Roger Grosse et al., "Ground Truth Dataset and Baseline Evaluations for Intrinsic Image Algorithms", *in*: *International Conference on Computer Vision* (2009), DOI: 10.1109/ICCV.2009.5459428, URL: http://ieeexplore.ieee.org/document/5459428/ (visited on 10/14/2020).

[Gu *et al.* 2020] Jinjin Gu et al., "PIPAL: A Large-Scale Image Quality Assessment Dataset for Perceptual Image Restoration", *in*: *European Conference on Computer Vision (ECCV) 2020*, Cham: Springer International Publishing, 2020, pp. 633–651.

[Gu *et al.* 2022] Jinjin Gu et al., "NTIRE 2022 Challenge on Perceptual Image Quality Assessment", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 951–967.

[Gu *et al.* 2017] Ke Gu et al., "No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization", *in*: *IEEE Transactions on Cybernetics* 47.*12* (Dec. 2017), pp. 4559–4565, ISSN: 2168-2267, 2168-2275, DOI: 10.1109/TCYB.2016.2575544, URL: http://ieeexplore.ieee.org/document/7492198/ (visited on 02/07/2022).

[Gu *et al.* 2018] Ke Gu et al., "Learning a No-Reference Quality Assessment Model of Enhanced Images With Big Data", *in*: *IEEE Transactions on Neural Networks and Learning Systems* 29.*4* (Apr. 2018), pp. 1301–1313, ISSN: 2162-2388, DOI: 10.1109/TNNLS.2017.2649101.

[Guo *et al.* 2017] Xiaojie Guo, Yu Li, and Haibin Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation", *in*: *IEEE Transactions on Image Processing* (Feb. 2017), ISSN: 1941-0042, DOI: 10.1109/TIP.2016.2639450.

[Hassen *et al.* 2013] R. Hassen, Zhou Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence", *in*: *IEEE Transactions on Image Processing* (July 2013), ISSN: 1057-7149, 1941-0042, DOI: 10.1109/TIP.2013.2251643, URL: http://ieeexplore.ieee.org/document/6476013/ (visited on 08/10/2021).

[Heusel *et al.* 2017] Martin Heusel et al., "Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", *in*: *Advances in neural information processing systems* 30 (2017).

[Horn *et al.* 1974] Berthold K.P. Horn, "Determining Lightness from an Image", *in*: *Computer Graphics and Image Processing* 3.*4* (Dec. 1974), pp. 277–299, ISSN: 0146664X, DOI: 10.1016/0146-664X(74)90022-7, URL: https://linkinghub.elsevier.com/retrieve/pii/0146664X74900227 (visited on 10/08/2021).

[Hu *et al.* 2021] Qiming Hu and Xiaojie Guo, "Trash or Treasure? An Interactive Dual-Stream Strategy for Single Image Reflection Separation", *in*: *Advances in Neural Information Processing Systems* 34 (2021), pp. 24683–24694.

[Huang *et al.* 2020] Haofeng Huang et al., "Raw-Guided Enhancing Reprocess of Low-Light Image via Deep Exposure Adjustment", *in*: *Proceedings of the Asian Conference on Computer Vision*, 2020.

[Huang *et al.* 2017] Xun Huang and Serge Belongie, "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization", *in*: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1501–1510.

[Huang *et al.* 2018] Xun Huang et al., "Multimodal Unsupervised Image-to-Image Translation", *in*: *Computer Vision – ECCV 2018*, ed. by Vittorio Ferrari et al., vol. 11207, Cham: Springer International Publishing, 2018, pp. 179–196, DOI: 10.1007/978-3-030-01219-9_11, URL: http://link.springer.com/10.1007/978-3-030-01219-9_11 (visited on 06/02/2022).

[Huber *et al.* 1964] Peter J Huber, "Robust Estimation of a Location Parameter: Annals Mathematics Statistics, 35", *in*: *Ji, S., Xue, Y. and Carin, L.(2008),'Bayesian compressive sensing', IEEE Transactions on signal processing* 56.*6* (1964), pp. 2346–2356.

[Hummel *et al.* 1977] Robert Hummel, "Image Enhancement by Histogram Transformation", *in*: *Computer Graphics and Image Processing* 6.*2* (Apr. 1, 1977), pp. 184–195, ISSN: 0146-664X, DOI: 10.1016/S0146-664X(77)80011-7, URL: https://www.sciencedirect.com/science/article/pii/S0146664X77800117 (visited on 05/15/2021).

[Hyvärinen *et al.* 2005] Aapo Hyvärinen and Peter Dayan, "Estimation of Non-Normalized Statistical Models by Score Matching.", *in*: *Journal of Machine Learning Research* 6.*4* (2005).

[ISO-92882022 *et al.* 2022] ISO-9288:2022, *Thermal Insulation — Heat Transfer by Radiation — Vocabulary*, Geneva, CH: ISO, Aug. 2022, p. 23,

URL: https://www.iso.org/standard/82088.html (visited on 02/28/2023).

[Jaynes *et al.* 1988] ET Jaynes, "The Relation of Bayesian and Maximum Entropy Methods", *in*: *Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations*, Springer, 1988, pp. 25–29.

[Jiang *et al.* 2013] Jun Jiang et al., "What Is the Space of Spectral Sensitivity Functions for Digital Color Cameras?", *in*: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, 2013 IEEE Workshop on Applications of Computer Vision (WACV), Jan. 2013, pp. 168–179, DOI: 10.1109/WACV.2013.6475015.

[Jiang *et al.* 2021] Yifan Jiang et al., "EnlightenGAN: Deep Light Enhancement without Paired Supervision", version 1, *in*: *IEEE Transactions on Image Processing* (2021), arXiv: 1906.06972, URL: http://arxiv.org/abs/1906.06972 (visited on 11/17/2020).

[Kantorovitch *et al.* 1958] Leonid Kantorovitch, "On the Translocation of Masses", *in*: *Management science* 5.*1* (1958), pp. 1–4.

[Ketcham *et al.* 1974] David J. Ketcham, Roger W. Lowe, and J. W. Weber, *Image Enhancement Techniques for Cockpit Displays:* Fort Belvoir, VA: Defense Technical Information Center, Dec. 1, 1974, DOI: 10.21236/ADA014928, URL: http://www.dtic.mil/docs/citations/ADA014928 (visited on 05/15/2021).

[Kim *et al.* 2023] Beomsu Kim et al., *Unpaired Image-to-Image Translation via Neural Schr\"odinger Bridge*, May 24, 2023, arXiv: 2305.15086 [cs, stat], URL: http://arxiv.org/abs/2305.15086 (visited on 06/19/2023), preprint.

[Kim *et al.* 2012] Seon Joo Kim et al., "A New In-Camera Imaging Model for Color Computer Vision and Its Application", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.*12* (Dec. 2012), pp. 2289–2302, ISSN: 1939-3539, DOI: 10.1109/TPAMI.2012.58.

[Kimmel *et al.* 2003] Ron Kimmel et al., "A Variational Framework for Retinex", *in*: (2003), p. 17.

[Kingma *et al.* 2015] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization", *in*: *International Conference on Learning Representations (ICLR)* (2015), arXiv: 1412.6980, URL: http://arxiv.org/abs/1412.6980 (visited on 05/11/2020).

[Kodak *et al.* 1999] Kodak, *True Color Kodak Images*, 1999, URL: https://www.r0k.us/graphics/kodak/ (visited on 04/18/2023).

[Krizhevsky *et al.* 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", *in*: *Advances in neural information processing systems* 25 (2012).

[Kruithof *et al.* 1937] J Kruithof, "Telefoonverkeersrekening", *in*: *De Ingenieur* 52 (1937), pp. 15–25.

[Land *et al.* 1964] Edwin H Land, "The Retinex", *in*: *American Scientist* 52.*2* (1964), pp. 247–264.

[Land *et al.* 1977] Edwin H Land, "The Retinex Theory of Color Vision", *in*: *Scientific American* (Dec. 1977).

[Land *et al.* 1971] Edwin H. Land and John J. McCann, "Lightness and Retinex Theory", *in*: *Journal of the Optical Society of America* 61.*1* (Jan. 1, 1971), p. 1, ISSN: 0030-3941, DOI: 10.1364/JOSA.61.000001, URL: https://www.osapublishing.org/abstract.cfm?URI=josa-61-1-1 (visited on 10/13/2020).

[Le Pendu *et al.* 2022] Mikael Le Pendu and Christine Guillemot, "Preconditioned Plug-and-Play ADMM with Locally Adjustable Denoiser for Image Restoration", *in*: *SIAM Journal on Imaging Sciences* (2022), pp. 1–30.

[Lecert *et al.* 2022] Arthur Lecert et al., "A New Regularization for Retinex Decomposition of Low-Light Images", *in*: *2022 IEEE International Conference on Image Processing (ICIP)*, Oct. 2022, pp. 906–910, DOI: 10.1109/ICIP46576.2022.9897893.

[Lecert *et al.* 2023] Arthur Lecert et al., "GAN Architecture Leveraging a Retinex Model with Colored Illumination for Low-Light Image Restoration", *in*: *IEEE access* 11 (2023), pp. 84574–84588, DOI: 10.1109/ACCESS.2023.3301614.

[LeCun *et al.* 1989] Y. LeCun et al., "Backpropagation Applied to Handwritten Zip Code Recognition", *in*: *Neural Computation* 1.*4* (Dec. 1989), pp. 541–551, ISSN: 0899-7667, DOI: 10.1162/neco.1989.1.4.541, eprint: https://direct.mit.edu/neco/article-pdf/1/4/541/811941/neco.1989.1.4.541.pdf, URL: https://doi.org/10.1162/neco.1989.1.4.541.

[Léonard *et al.* 2012] Christian Léonard, "From the Schrödinger Problem to the Monge–Kantorovich Problem", *in*: *Journal of Functional Analysis* 262.*4* (Feb. 2012), pp. 1879–1920, ISSN: 00221236, DOI: 10.1016/j.jfa.2011.11.026, URL: https://linkinghub.elsevier.com/retrieve/pii/S0022123611004253 (visited on 05/10/2023).

[Léonard *et al.* 2014a] Christian Léonard, "A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport", *in*: *Discrete and Continuous Dynamical Systems-Series A* 34.*4* (2014), pp. 1533–1574.

[Léonard *et al.* 2014b] Christian Léonard, Sylvie Rœlly, and Jean-Claude Zambrini, "Reciprocal Processes. A Measure-Theoretical Point of View", *in*: *Probability Surveys* 11 (2014), pp. 237–269.

[Li *et al.* 2020] Chongyi Li, Chunle Guo, and Chen Change Loy, "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (2020), arXiv: 2103.00860, URL: http://arxiv.org/abs/2103.00860 (visited on 03/08/2021).

[Li *et al.* 2015] Yu Li, Robby T. Tan, and Michael S. Brown, "Nighttime Haze Removal with Glow and Multiple Light Colors", *in*: *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE, Dec. 2015, pp. 226–234, ISBN: 978-1-4673-8391-2, DOI: 10.1109/ICCV.2015.34, URL: http://ieeexplore.ieee.org/document/7410391/ (visited on 12/01/2021).

[Lipman *et al.* 2023] Yaron Lipman et al., "Flow Matching for Generative Modeling", *in*: *The Eleventh International Conference on Learning Representations*, 2023, URL: https://openreview.net/forum?id=PqvMRDCJT9t.

[Liu *et al.* 2019] Jiaming Liu et al., "Image Restoration Using Total Variation Regularized Deep Image Prior", *in*: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (May 2019), DOI: 10.1109/ICASSP.2019.8682856, URL: https://ieeexplore.ieee.org/document/8682856/ (visited on 09/27/2021).

[Liu *et al.* 2020] Yu-Lun Liu et al., "Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline", *in*: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, June 2020, pp. 1648–1657, ISBN: 978-1-72817-168-5, DOI: 10.1109/CVPR42600.2020.00172, URL: https://ieeexplore.ieee.org/document/9157653/ (visited on 12/22/2021).

[Liu *et al.* 2023] Xingchao Liu, Chengyue Gong, and qiang liu, "Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow", *in*: *The Eleventh International Conference on Learning Representations*, 2023, URL: https://openreview.net/forum?id=XVjTT1nw5z.

[Liu *et al.* 2022] Xingchao Liu, Lemeng Wu, Mao Ye, et al., "Let Us Build Bridges: Understanding and Extending Diffusion Generative Models", *in*: *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.

[Lombardi *et al.* 2016] Stephen Lombardi and Ko Nishino, "Reflectance and Illumination Recovery in the Wild", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.*1* (Jan. 1, 2016), pp. 129–141, ISSN: 0162-8828, 2160-9292, DOI: 10.1109/TPAMI.2015.2430318, URL: http://ieeexplore.ieee.org/document/7102772/ (visited on 10/27/2022).

[Lyu *et al.* 2022] Zhaoyang Lyu et al., "Accelerating diffusion models via early stop of the diffusion process", *in*: *arXiv preprint arXiv:2205.12524* (2022).

[Ma *et al.* 2015] Kede Ma, Kai Zeng, and Zhou Wang, "Perceptual Quality Assessment for Multi-Exposure Image Fusion", *in*: *IEEE Transactions on Image Processing* 24.*11* (Nov. 2015), pp. 3345–3356, ISSN: 1941-0042, DOI: 10.1109/TIP.2015.2442920.

[Ma *et al.* 2021] Tian Ma et al., "RetinexGAN:Unsupervised Low-Light Enhancement With Two-Layer Convolutional Decomposition Networks", *in*: *IEEE Access* 9 (2021), pp. 56539–56550, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2021.3072331.

[Mao *et al.* 2023] Weibo Mao et al., "Leapfrog Diffusion Model for Stochastic Trajectory Prediction", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5517–5526.

[Mao *et al.* 2017] Xudong Mao et al., "Least Squares Generative Adversarial Networks", *in*: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2794–2802.

[McCann *et al.* 2017] John J. McCann, *History of the Retinex Theory*, 2017, URL: https://www.retinex2.net/Publications/ewExternalFiles/1971%20Land%2C%20McCann.pdf (visited on 06/15/2023).

[Meng *et al.* 2023] Chenlin Meng et al., "On distillation of guided diffusion models", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14297–14306.

[Mittal *et al.* 2013] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer", *in*: *IEEE Signal Processing Letters* 20.*3* (Mar. 2013), pp. 209–212, ISSN: 1070-9908, 1558-2361, DOI: 10.1109/LSP.2012.2227726, URL: http://ieeexplore.ieee.org/document/6353522/ (visited on 05/06/2021).

[Monge *et al.* 1781] Gaspard Monge, "Mémoire Sur La Théorie Des Déblais et Des Remblais", *in*: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704.

[Narasimhan *et al.* 2003] S.G. Narasimhan and S.K. Nayar, "Shedding Light on the Weather", *in*: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* CVPR 2003: Computer Vision and Pattern Recognition Conference, Madison, WI, USA: IEEE Comput. Soc, 2003, pp. I-665-I–672, ISBN: 978-0-7695-1900-5, DOI: 10.1109/CVPR.2003.1211417, URL: http://ieeexplore.ieee.org/document/1211417/ (visited on 12/01/2021).

[Nicodemus *et al.* 1965] Fred E. Nicodemus, "Directional Reflectance and Emissivity of an Opaque Surface", *in*: *Applied Optics* 4.7 (July 1965), pp. 767–775, DOI: 10.1364/AO.4.000767, URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-4-7-767.

[Øksendal *et al.* 2003] Bernt Øksendal, *Stochastic Differential Equations*, Universitext, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, DOI: 10.1007/978-3-642-14394-6, URL: http://link.springer.com/10.1007/978-3-642-14394-6 (visited on 06/09/2023).

[Oord *et al.* 2019] Aaron van den Oord, Yazhe Li, and Oriol Vinyals, *Representation Learning with Contrastive Predictive Coding*, Jan. 22, 2019, DOI: 10.48550/arXiv.1807.03748, arXiv: 1807.03748 [cs, stat], URL: http://arxiv.org/abs/1807.03748 (visited on 09/12/2022), preprint.

[Park *et al.* 2020] Taesung Park et al., "Contrastive Learning for Unpaired Image-to-Image Translation", *in*: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, Springer, 2020, pp. 319–345.

[Paszke *et al.* 2019] Adam Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library", *in*: *Advances in Neural Information Processing Systems* (Dec. 3, 2019), arXiv: 1912.01703, URL: http://arxiv.org/abs/1912.01703 (visited on 12/06/2021).

[Peyre *et al.* 2019] Gabriel Peyre and Marco Cuturi, "Computational Optimal Transport", *in*: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607.

[Phong *et al.* 1975] Bui Tuong Phong, "Illumination for Computer Generated Pictures", *in*: *Communications of the ACM* 18.*6* (1975), pp. 311–317.

[Pizer *et al.* 1990] S.M. Pizer et al., "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness", *in*: *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, [1990] First Conference on Visualization in Biomedical Computing, Atlanta, GA, USA: IEEE Comput. Soc. Press, 1990, pp. 337–345, ISBN: 978-0-8186-2039-3, DOI: 10.1109/VBC.1990.109340, URL: http://ieeexplore.ieee.org/document/109340/ (visited on 05/16/2021).

[Pizzati *et al.* 2021] Fabio Pizzati, Pietro Cerri, and Raoul de Charette, "CoMoGAN: Continuous Model-Guided Image-to-Image Translation", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14288–14298.

[Pizzati *et al.* 2022] Fabio Pizzati, Jean-François Lalonde, and Raoul de Charette, "Manifest: Manifold Deformation for Few-Shot Image Translation", *in*: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, Springer, 2022, pp. 440–456.

[Ponomarenko *et al.* 2015] Nikolay Ponomarenko et al., "Image Database TID2013: Peculiarities, Results and Perspectives", *in*: *Signal processing: Image communication* 30 (2015), pp. 57–77.

[Provenzi *et al.* 2017] Edoardo Provenzi, "Formalizations of the Retinex Model and Its Variants with Variational Principles and Partial Differential Equations", *in*: *Journal of Electronic Imaging* (Dec. 2017), URL: https://hal.archives-ouvertes.fr/hal-02480258 (visited on 10/07/2021).

[Riba *et al.* 2019] Edgar Riba et al., "Kornia: An Open Source Differentiable Computer Vision Library for PyTorch", *in*: *Winter Conference on Applications of Computer Vision* (Oct. 9, 2019), arXiv: 1910.02190, URL: http://arxiv.org/abs/1910.02190 (visited on 12/06/2021).

[Ronneberger *et al.* 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *in*: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.

[Sagel *et al.* 2020] Alexander Sagel, Aline Roumy, and Christine Guillemot, "SUB-DIP: Optimization on a Subspace with Deep Image Prior Regularization and Application to Superresolution", *in*: *IEEE International Conference on Acoustics, Speech, and Signal Processing* (May 2020), URL: `https://hal.archives-ouvertes.fr/hal-02483083` (visited on 10/08/2020).

[Salimans *et al.* 2016] Tim Salimans et al., "Improved Techniques for Training GANs", *in*: *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, Red Hook, NY, USA: Curran Associates Inc., June 10, 2016, pp. 2234–2242, ISBN: 978-1-5108-3881-9.

[Schrödinger *et al.* 1931] Erwin Schrödinger, *Über Die Umkehrung Der Naturgesetze*, Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u ..., 1931.

[Schwartz *et al.* 2004] Odelia Schwartz, Terrence J Sejnowski, and Peter Dayan, "Assignment of Multiplicative Mixtures in Natural Images", *in*: *Advances in Neural Information Processing Systems* (2004), p. 8.

[Sen *et al.* 2012] Pradeep Sen et al., "Robust Patch-Based Hdr Reconstruction of Dynamic Scenes", *in*: *ACM Transactions on Graphics* 31.*6* (Nov. 2012), pp. 1–11, ISSN: 0730-0301, 1557-7368, DOI: `10.1145/2366145.2366222`, URL: `https://dl.acm.org/doi/10.1145/2366145.2366222` (visited on 04/18/2023).

[Shang *et al.* 2023] Yuzhang Shang et al., "Post-training quantization on diffusion models", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1972–1981.

[Shi *et al.* 2023] Yuyang Shi et al., *Diffusion Schr\"odinger Bridge Matching*, Mar. 29, 2023, arXiv: 2303.16852 [cs, stat], URL: http://arxiv.org/abs/2303.16852 (visited on 04/03/2023), preprint.

[Simonyan *et al.* 2015] K Simonyan and A Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *in*: *3rd International Conference on Learning Representations (ICLR 2015)*, Computational and Biological Learning Society, 2015.

[Song *et al.* 2019] Yang Song and Stefano Ermon, "Generative Modeling by Estimating Gradients of the Data Distribution", *in*: *Advances in neural information processing systems* 32 (2019).

[Song *et al.* 2021] Yang Song et al., "Score-Based Generative Modeling through Stochastic Differential Equations", *in*: *International Conference on Learning Representations*, 2021, URL: https://openreview.net/forum?id=PxTIG12RRHS.

[Sun *et al.* 2020] Pei Sun et al., "Scalability in Perception for Autonomous Driving: Waymo Open Dataset", *in*: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, June 2020, pp. 2443–2451, ISBN: 978-1-72817-168-5, DOI: 10.1109/CVPR42600.2020.00252, URL: https://ieeexplore.ieee.org/document/9156973/ (visited on 07/08/2022).

[Szegedy *et al.* 2016] Christian Szegedy et al., "Rethinking the Inception Architecture for Computer Vision", *in*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[Thompson *et al.* 2002] William B. Thompson, Peter Shirley, and James A. Ferwerda, "A Spatial Post-Processing Algorithm for Images of Night Scenes", *in*: *Journal of Graphics Tools* 7.1 (Jan. 2002), pp. 1–12, ISSN: 1086-7651, DOI: 10.1080/10867651.2002.10487550, URL: http://www.tandfonline.com/doi/abs/10.1080/10867651.2002.10487550 (visited on 01/24/2023).

[Ulyanov *et al.* 2017] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Improved Texture Networks: Maximizing Quality and Diversity in Feed-Forward Stylization and Texture Synthesis", *in*: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, July 2017, pp. 4105–4113, ISBN: 978-1-5386-0457-1, DOI: 10.1109/CVPR.2017.437, URL: http://ieeexplore.ieee.org/document/8099920/ (visited on 03/13/2023).

[Ulyanov *et al.* 2018] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep Image Prior", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (Apr. 5, 2018), arXiv: 1711.10925.

[Villani *et al.* 2009] Cédric Villani et al., *Optimal Transport: Old and New*, vol. 338, Springer, 2009.

[Vonikakis *et al.* 2018] Vasileios Vonikakis, *Vasileios Vonikakis - Datasets*, 2018, URL: https://sites.google.com/site/vonikakis/datasets (visited on 04/18/2023).

[Wainwright *et al.* 2000] Martin J Wainwright and Eero Simoncelli, "Scale Mixtures of Gaussians and the Statistics of Natural Images", *in*: *Advances in Neural Information Processing Systems*, vol. 12, MIT Press, 2000.

[Wainwright *et al.* 2001] Martin J. Wainwright, Eero P. Simoncelli, and Alan S. Willsky, "Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images", *in*: *Applied and Computational Harmonic Analysis* 11.*1* (July 2001), pp. 89–123, ISSN: 10635203, DOI: 10.1006/acha.2000.0350, URL: https://linkinghub.elsevier.com/retrieve/pii/S1063520300903506 (visited on 01/26/2022).

[Wang *et al.* 2013] Shuhang Wang et al., "Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images", *in*: *IEEE Transactions on Image Processing* 22.*9* (Sept. 2013), pp. 3538–3548, ISSN: 1057-7149, 1941-0042, DOI: 10.1109/TIP.2013.

2261309, URL: https://ieeexplore.ieee.org/document/6512558/ (visited on 04/11/2023).

[Wang *et al.* 2004] Zhou Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity", *in*: *IEEE Transactions on Image Processing* (2004).

[Weber *et al.* 2018] Allan G Weber, "USC-SIPI Report #432 The USC-SIPI Image Database: Version 6", *in*: (2018).

[Wei *et al.* 2018] Chen Wei et al., "Deep Retinex Decomposition for Low-Light Enhancement", *in*: *British Machine Vision Conference* (Aug. 2018), arXiv: 1808.04560.

[Wei *et al.* 2021] Kaixuan Wei et al., "Physics-Based Noise Modeling for Extreme Low-Light Photography", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.*11* (2021), pp. 8520–8537.

[Xu *et al.* 2023] Ruikang Xu et al., "Continuous Spectral Reconstruction from RGB Images via Implicit Neural Representation", *in*: *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, Springer, 2023, pp. 78–94.

[Yan *et al.* 2020] Longbin Yan et al., "Reconstruction of Hyperspectral Data From RGB Images With Prior Category Information", *in*: *IEEE Transactions on Computational Imaging* 6 (2020), pp. 1070–1081, ISSN: 2333-9403, DOI: 10.1109/TCI.2020.3000320.

[Ying *et al.* 2017] Zhenqiang Ying et al., "A New Low-Light Image Enhancement Algorithm Using Camera Response Model", *in*: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017 IEEE International Conference on Computer Vision Workshop (ICCVW), Venice: IEEE, Oct. 2017, pp. 3015–3022, ISBN: 978-1-5386-1034-3, DOI: 10.1109/ICCVW.2017.356, URL: http://ieeexplore.ieee.org/document/8265567/ (visited on 02/28/2023).

[Yu *et al.* 2020] Fisher Yu et al., "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning", *in*: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,

2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, June 2020, pp. 2633–2642, ISBN: 978-1-72817-168-5, DOI: 10.1109/CVPR42600.2020.00271, URL: https://ieeexplore.ieee.org/document/9156329/ (visited on 04/25/2023).

[Yu *et al.* 2022] Ye Yu and William A. P. Smith, "Outdoor Inverse Rendering from a Single Image Using Multiview Self-Supervision", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2022), pp. 3659–3675, DOI: 10.1109/TPAMI.2021.3058105.

[Yuan *et al.* 2022] Yu Yuan et al., *Learning to Kindle the Starlight*, Nov. 16, 2022, arXiv: 2211.09206 [cs, eess], URL: http://arxiv.org/abs/2211.09206 (visited on 11/22/2022), preprint.

[Zhang *et al.* 2018] Richard Zhang et al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (Apr. 10, 2018), arXiv: 1801.03924, URL: http://arxiv.org/abs/1801.03924 (visited on 07/06/2021).

[Zhang *et al.* 2019] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo, "Kindling the Darkness: A Practical Low-light Image Enhancer", *in*: *Proceedings of the 27th ACM International Conference on Multimedia* (May 4, 2019), arXiv: 1905.04161.

[Zhang *et al.* 2021] Yonghua Zhang et al., "Beyond Brightening Low-light Images", *in*: *International Journal of Computer Vision* 129.4 (Apr. 2021), pp. 1013–1037, ISSN: 0920-5691, 1573-1405, DOI: 10.1007/s11263-020-01407-x, URL: http://link.springer.com/10.1007/s11263-020-01407-x (visited on 02/28/2023).

[Zhao *et al.* 2021] Zunjin Zhao et al., "RetinexDIP: A Unified Deep Framework for Low-light Image Enhancement", *in*: *IEEE Transactions on Circuits and Systems for Video Technology* (2021), pp. 1–1, ISSN: 1558-2205, DOI: 10.1109/TCSVT.2021.3073371.

[Zhu *et al.* 2021] Zhiyu Zhu et al., "Semantic-Embedded Unsupervised Spectral Reconstruction from Single RGB Images in the Wild",

in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, Oct. 2021, pp. 2259–2268, ISBN: 978-1-66542-812-5, DOI: 10.1109/ICCV48922.2021.00228, URL: https://ieeexplore.ieee.org/document/9710095/ (visited on 07/01/2022).

COLLEGE DOCTORAL BRETAGNE | MATHS, TELECOMS INFORMATIQUE, SIGNAL SYSTEMES, ELECTRONIQUE

Université de Rennes

Inria

Titre : Restauration d'images à faible luminosité à l'aide de méthodes d'apprentissage profond.

Mot clés : Amélioration de la luminosité, Décomposition d'image, Restauration d'image, Problème inverse, Modèle Retinex, Apprentissage profond

Résumé : Aujourd'hui, de nombreux domaines évoluent pour inclure des algorithmes de vision par ordinateur. Or, ceux-ci n'ont pas été conçus pour fonctionner sur des scènes nocturnes. Leurs performances s'en trouvent fortement dégradées ce qui limite leurs applications. Cela est dû aux fortes dégradations lors de la capture d'images de nuit. Elles prennent la forme d'un faible rapport signal à bruit ainsi que de déviations de couleur. Dans cette thèse, nous répondons à cette problématique en cherchant à les restaurer à l'aide de méthodes d'apprentissage profond. Notre contexte nous force à nous concentrer sur des méthodes non supervisées qui n'hallucinent pas. Dans un premier temps, nous identifions une solution triviale au niveau de l'illumination ignorée jusqu'à maintenant. Nous proposons un a priori pour corriger ce problème ainsi qu'une méthode de restauration qui ne nécessite pas de jeu de données d'apprentissage. Nous obtenons des résultats proches des méthodes de l'état de l'art supervisées. Dans un deuxième temps, nous revenons sur les définitions des composantes du modèle Retinex et proposons plusieurs améliorations afin de suivre la physique de la lumière. Une architecture basée GAN est ensuite définie. Notre méthode garantit qu'aucune hallucination n'est ajoutée en sortie. Enfin, dans un dernier temps, nous dévoilons le lien entre notre objectif et le problème du pont de Schrödinger. Nous intégrons des a priori à un algorithme de transport optimal à base de modèles de diffusion afin d'inverser les dégradations.

Title: Low-light image restoration with deep learning techniques

Keywords: Low light enhancement, Image decomposition, Image restoration, Inverse problems, Retinex model, Deep Learning

Abstract: Nowadays, many fields are evolving to include computer vision algorithms. However, these algorithms have not been designed to operate in night scenes. This severely degrades their performance, which limits their applications. This is due to the severe degradations that occurs when night images are captured. These mainly take the form of a low signal-to-noise ratio and color deviations. In this thesis, we address this problem by seeking to restore them using deep learning methods. Our context forces us to focus on unsupervised methods that do not hallucinate. First, we identify a trivial solution in the decomposition that has been ignored until now. We propose a prior to correct this problem, as well as an restoration method that does not require a dataset. We obtain results close to the state-of-the-art supervised methods. Secondly, we review the definitions of the Retinex model components and propose several improvements to keep them coherent with the physics of light. A GAN-based architecture is then defined. Our method ensures that no hallucination is added to the output. Finally, we unveil the link between our goal and the Schrödinger bridge problem. We integrate physical priors to an optimal transport algorithm based on diffusion models to reverse the degradations.