



UNIVERSITÉ DE MONTPELLIER
FACULTÉ DES SCIENCES

MASTER 2 - IMAGINE

Projet Image, sécurité et deep learning

Sujet 12 - Débruitage ou restauration d'images par CNN

CR2 - Etat de l'art de la restauration d'images (avec CNN)

[HAI918I]

BES Jean-Baptiste
COMBOT Evan

27/10/2024

[Lien GitHub de notre projet image](#)

1 Introduction

Dans ce compte rendu, nous avons réalisé un état de l'art sur la restauration d'images à l'aide de réseaux de neurones convolutifs (CNN). Nous avons identifié de nombreux modèles compatibles avec notre projet. Cependant, nous avons décidé, dans un premier temps de sélectionner trois modèles qui nous semblent les plus adaptés à la restauration d'images anciennes. Le premier modèle, Deep Latent Space Translation, repose sur une architecture d'auto-encodeur variationnel (VAE). Le deuxième, Pix2Pix, utilise une architecture U-Net associée à un réseau antagoniste génératif conditionnel (cGAN). Enfin, le troisième modèle, Deep Image Prior, est un réseau conçu pour résoudre des problèmes inverses tels que la réduction de bruit, la super-résolution et l'inpainting.

2 État de l'art (avec CNN)

De manière générale, l'état de l'art pour la restauration d'images anciennes utilisant des réseaux de neurones convolutifs (CNN) présente des similitudes avec les méthodes classiques.

2.1 Détection des zones endommagées

Cette partie consiste à créer un masque binaire pour identifier les zones à restaurer. On utilise le modèle **EdgeConnect** qui est un CNN capable de détecter les bordures et les zones endommagées à l'aide d'un réseau d'inpainting guidé par les contours. Le modèle de détection génère un masque M où chaque pixel $M_{i,j}$ est soit 1 (endommagé) soit 0 (intact).

$$\mathcal{L}_{\text{inpainting}} = \frac{1}{|M|} \sum_{(i,j) \in M} |I_{\text{output}}(i,j) - I_{\text{ground truth}}(i,j)|$$

Cette fonction de perte est généralement une fonction d'entropie croisée binaire où y_i est le label réel (0 ou 1) pour le pixel i , p_i est la probabilité prédite que le pixel i soit endommagé et N est le nombre total de pixels dans l'image.



FIGURE 1 – Masques binaires

2.2 Inpainting

Comme dans la restauration d'images anciennes classique, l'inpainting permet de restaurer les parties endommagées de l'image. La fonction de perte L_1 mesure la différence entre l'image restaurée et l'image originale uniquement sur les zones indiquées par le masque M .

$$\mathcal{L}_{\text{inpainting}} = \frac{1}{|M|} \sum_{(i,j) \in M} |I_{\text{output}}(i,j) - I_{\text{ground truth}}(i,j)|$$

où $|M|$ représente le nombre de pixels marqués dans le masque M . Pour améliorer le résultat, on peut ajouter une perte de similarité structurelle (SSIM) :

$$\mathcal{L}_{SSIM} = 1 - SSIM(I_{\text{output}}, I_{\text{ground truth}})$$



FIGURE 2 – Inpainting

2.3 Débruitage

Un CNN est utilisé pour apprendre une fonction f_θ paramétrée par les poids θ du réseau, qui mappe une image bruitée $I_{\text{bruitée}}$ à une image débruitée I_{output} . La fonction de perte est généralement basée sur la perte L_2 ,

$$\mathcal{L}_{\text{débruitage}} = \frac{1}{N} \sum_{i=1}^N (I_{\text{output}}(i) - I_{\text{clean}}(i))^2$$

2.4 Super-résolution

Cette partie vise à augmenter la résolution, la qualité et les détails de l'image restaurée. Les techniques comme le SRCNN (Super Resolution Convolutional Neural Network) cherchent à minimiser la perte L2 entre l'image super-résolue et l'image haute résolution originale.

$$\mathcal{L}_{\text{SR}} = \frac{1}{N} \sum_{i=1}^N (I_{\text{SR}}(i) - I_{\text{HR}}(i))^2$$

ou l'utilisation de ESRGAN pour une méthode plus avancée

$$\mathcal{L}_{\text{GAN}} = E_{I_{\text{HR}}} [\log D(I_{\text{HR}})] + E_{I_{\text{SR}}} [\log(1 - D(I_{\text{SR}}))]$$

où D est le discriminateur qui apprend à distinguer les images générées des vraies images.

2.5 Colorisation

La colorisation consiste à générer les canaux de couleur a et b dans l'espace de couleur Lab, en prenant l'image en niveaux de gris L comme entrée. La perte utilisée est la perte L_2 .

$$L_{\text{colourisation}} = \frac{1}{N} \sum_{i=1}^N ((a_{\text{output}}(i) - a_{\text{true}}(i))^2 + (b_{\text{output}}(i) - b_{\text{true}}(i))^2)$$

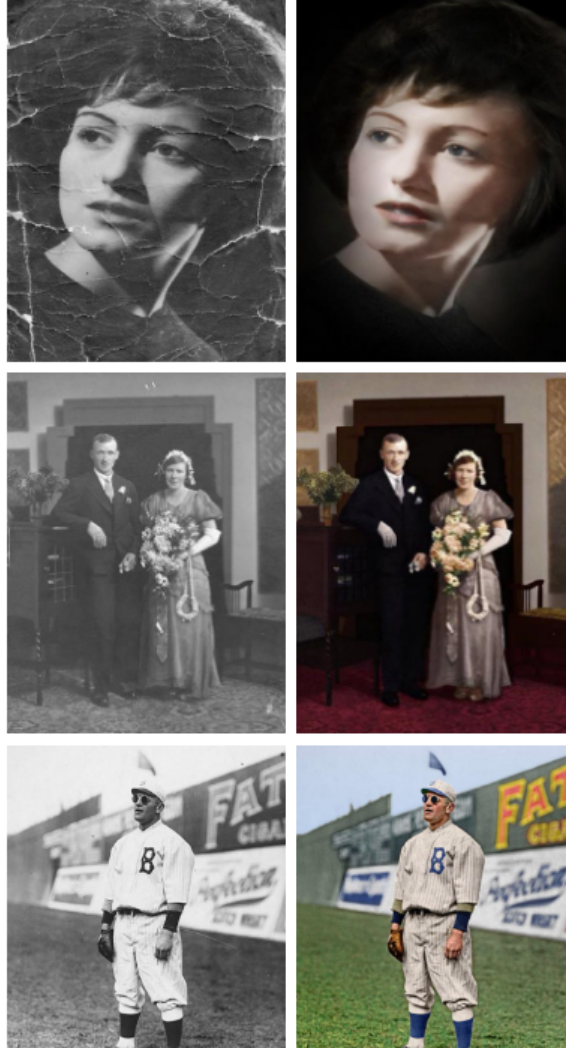


FIGURE 3 – Images anciennes restaurées

2.6 Résumé

Méthode	Architecture	Avantages	Inconvénients
Deep Latent Space Translation	VAE (Auto-encodeur Variationnel)	Capacité à modéliser des distributions complexes	Peut générer des artefacts
Pix2Pix	U-Net + PatchGAN	Bonne qualité de reconstruction avec des paires d'images correspondantes	Nécessite un grand nombre d'exemples d'entraînement
Deep Image Prior	CNN avec optimisation directe des poids	Ne nécessite pas de données d'entraînement	Plus lent en raison de l'optimisation itérative
NoGAN	GAN entraîné avec préformation et ajustement fin	Stable et rapide à converger	Complexité accrue dans la mise en œuvre
SRCNN	CNN pour la super-résolution	Facile à implémenter et à entraîner	Performances limitées pour les augmentations de résolution importantes

TABLE 1 – Comparaison des méthodes de restauration d'images anciennes

3 Les méthodes les plus viables

D'autres méthodes existent pour la restauration d'images anciennes que celles mentionnées ci-dessous mais leur efficacité est moindre car la restauration d'images n'est pas leur spécialité.

3.1 Deep Latent Space Translation

La Deep Latent Space Translation est une technique qui exploite les représentations latentes dans les réseaux de neurones profonds pour effectuer diverses tâches de transformation d'images y compris la restauration. L'idée est de projeter les images dans un espace latent où elles peuvent être manipulées et transformées pour accomplir des tâches telles que l'inpainting, la super-résolution ou la suppression de bruit.

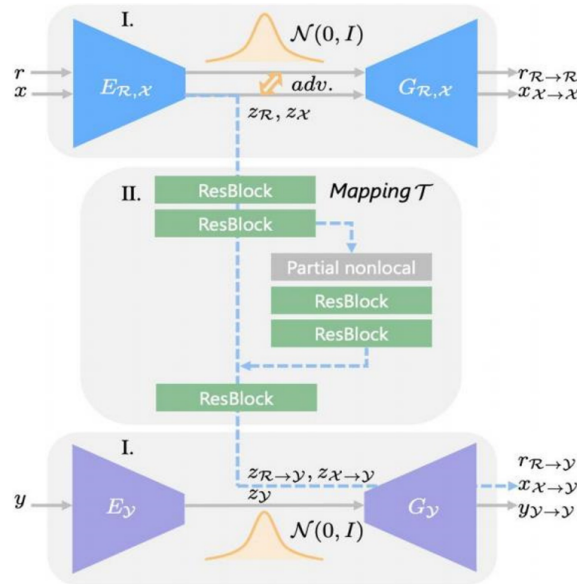


FIGURE 4 – Architecture Deep Latent Space Translation

Les méthodes qui exploitent l'espace latent comme les Variational Autoencoders (VAEs) permettent de régénérer des images en modifiant les représentations latentes pour corriger des imperfections ou remplir les zones manquantes.

Dans l'article de Microsoft **Bringing Old Photos Back to Life** [5] utilisant des VAEs, un réseau de neurones est d'abord déployé pour détecter les parties endommagées de l'image,

c'est-à-dire les zones nécessitant une réparation par inpainting. Cette étape de détection permet de générer un masque binaire indiquant les régions de l'image à restaurer. Ensuite, un réseau de restauration basé sur une architecture U-Net est utilisé pour effectuer l'inpainting des zones identifiées comme endommagées par le masque.

3.2 Pix2Pix

C'est une méthode de génération d'images reposant sur un réseau antagoniste génératif conditionnel (Conditional GAN) dont l'objectif est de traduire une image d'une représentation à une autre (par exemple, d'une image en niveaux de gris à une image en couleur).

Pix2Pix combine un réseau U-Net en tant que générateur et un PatchGAN comme discriminateur. Le générateur produit des images, tandis que le discriminateur évalue la qualité des images générées. En ajoutant une fonction de perte perceptuelle basée sur la base de données VGG, on pourrait améliorer la qualité des images restaurées.



FIGURE 5 – Résultat obtenu avec Pix2Pix

3.3 Deep Image Prior

Deep Image Prior (DIP) est un type de réseau de neurones convolutifs spécialisé dans la résolution de problèmes inverses tels que la réduction de bruit, la super-résolution et l'inpainting. Initialement, les poids sont fixés de manière aléatoire, puis, au fur et à mesure, ils sont optimisés, ce qui entraîne une amélioration de la qualité de l'image restaurée.

4 Planning prévisionnel

Nous allons affiner nos recherches sur les méthodes VAE, Pix2Pix et DIP afin de choisir quelle méthode implémenter dans le cadre de ce projet.

Notre planning prévisionnel, sur les semaines à venir, est le suivant :

- **CR3** (semaine 4) : Approfondissement des recherches sur les méthodes VAE, Pix2Pix et DIP et préparation pour implémenter la méthode sans CNN.
- **CR4** (semaine 5) : Implémentation de la méthode sans CNN.

Références

- [1] Neel PATEL. *Demystifying Deep Image Prior*. 2020. URL : <https://towardsdatascience.com/demystifying-deep-image-prior-7076e777e5ba>.
- [2] Tinghui Zhou PHILLIP ISOLA Jun-Yan Zhu et Alexei A. EFROS. "Image-to-Image Translation with Conditional Adversarial Networks". In : (2017). URL : <https://arxiv.org/pdf/1611.07004>.
- [3] Toon Stolk PIETER BIJL et Ruben MEESTER. *Image Restoration Using Pix2Pix*. 2019. URL : <https://medium.com/@28group28/image-restoration-using-pix2pix-83d07d09fe8e>.

- [4] Dmitry ULYANOV, Andrea VEDALDI et Victor LEMPITSKY. “Deep Image Prior”. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), p. 9446-9454. URL : https://sites.skoltech.ru/app/data/uploads/sites/25/2018/04/deep_image_prior.pdf.
- [5] Ziyu WAN et al. “Bringing Old Photos Back to Life”. In : *arXiv preprint arXiv :2004.09484* (2020). Work done during internship at Microsoft Research Asia. URL : <https://arxiv.org/pdf/2004.09484>.
- [6] Sheng ZHANG et al. *Pik-Fix : Restoring and Colorizing Old Photos*. 2022. URL : https://www.researchgate.net/publication/360383689_Pik-Fix_Restoring_and_Colorizing_Old_Photo.