CReSTIC

DIGITALIS

UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

DÉVELOPPEMENT D'ALGORITHMES DE RESTAURATION DE FRESQUES ANCIENNES PAR IA

ZINEDINE BOUZEKKAR – CHPS1006

ENCADRÉ PAR RÉMI ORVEAU ET ERIC DESJARDIN

ENTREPRISE ET CONTEXTE DU STAGE

- Les œuvres historiques, qu'elles soient peintures, sculptures ou manuscrits, s'abîment naturellement avec le temps :
  - Exposition à la lumière
  - Dégradation du support
  - Activité humaine, …

- Aujourd'hui, il y a des œuvres partiellement ou grandement endommagées. Il nous faut une méthode pour retrouver les détails qui ont été détériorés ou perdus.

- L'objectif de ce projet est de proposer une méthode de restauration pour des fresques anciennes (peinture réalisée directement sur un mur ou un plafond).

MODÈLE D'INTELLIGENCE ARTIFICIELLE – MODÈLE DE DIFFUSION

- Pour tenter de **restaurer** ces fresques, une méthode de restauration basée sur de **l'intelligence artificielle** est proposée.

- Plus précisément, notre méthode s'appuie sur les **modèles de diffusion**.

- Les modèles de diffusion sont des types de modèles **génératifs** utilisés en **apprentissage automatique** pour créer des données, comme des images ou du son.

- L'objectif du modèle est de retirer le **bruit** qui a été artificiellement ajouté aux données.

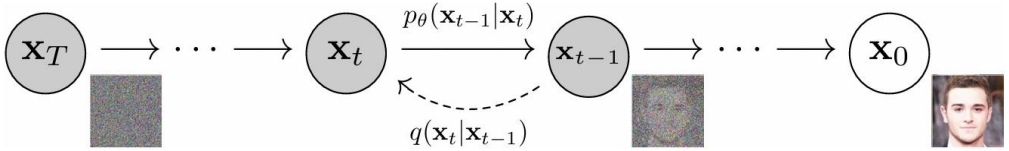


Figure 1 : Représentation visuelle du fonctionnement d'un modèle de diffusion

MODÈLE D'INTELLIGENCE ARTIFICIELLE – MODÈLE DE DIFFUSION DANS L'ESPACE LATENT

- Pour être plus précis, notre solution utilise un modèle de diffusion dans **l'espace latent**.
- Plutôt que de travailler dans l'espace des pixels, le modèle va travailler sur une **représentation compressée** de l'image.
- Cette représentation est obtenue à l'aide d'un puissant autre type de modèle appelé **auto-encodeur**.
- Cette représentation étant très proche de l'image encodée. Il est alors plus avantageux de travailler sur cet espace pour **réduire le coût computationnel** permettant ainsi d'accélérer les calculs.

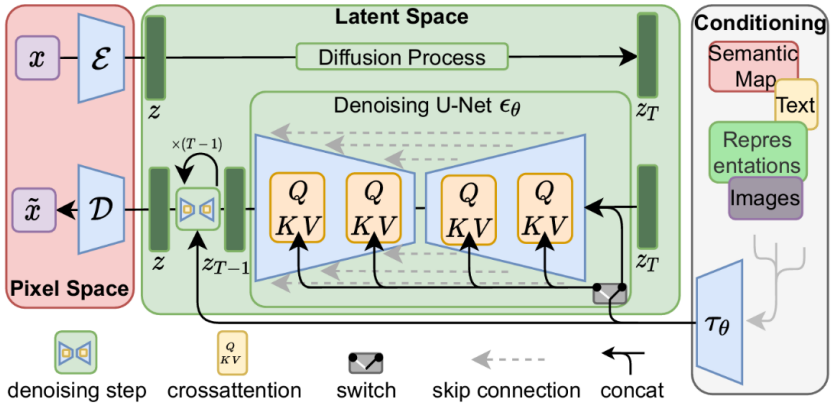


Figure 2 : Représentation visuelle du fonctionnement d'un modèle de diffusion qui utilise l'espace latent.

MODÈLE D'INTELLIGENCE ARTIFICIELLE – VISION TRANSFORMER

- Un **Vision Transformer** est un modèle qui apprend à analyser une image **en regardant toutes ses parties** en même temps, plutôt que morceau par morceau.

- Pour réaliser cela, ce modèle se base sur le **mécanisme d'attention** qui permet au modèle de peser l'importance de différentes parties d'une séquence.

- Il va permettre **d'analyser l'image dans son ensemble** et identifier quelles zones sont les plus importantes à observer, et comment elles sont liées entre elles.

- Cela permet à ce modèle de construire une compréhension **plus riche** et plus précise de ce que contient l'image.

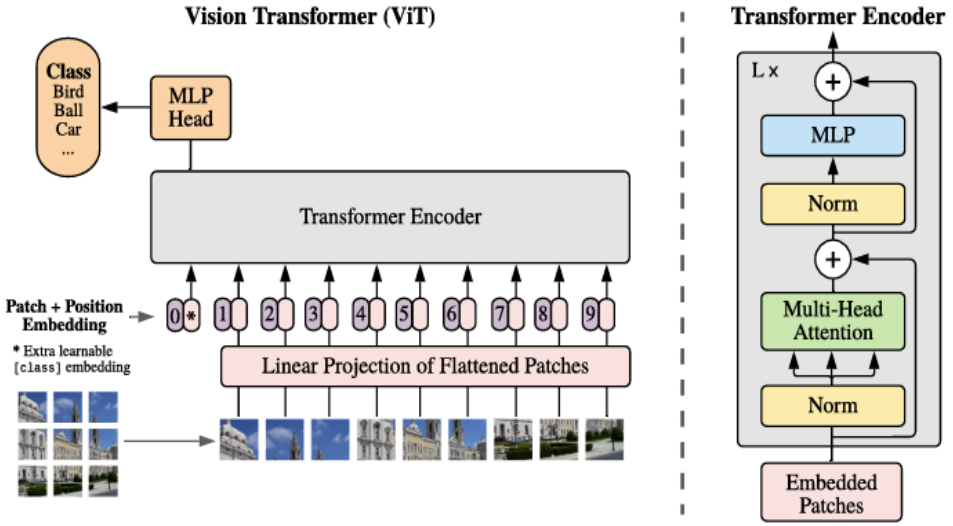


Figure 3 : Représentation visuelle d'un Vision Transformer

MÉTHODOLOGIE DU PAPIER DIFFIR

- Pour réaliser ce modèle de reconstruction, il a été décidé de reprendre un modèle de reconstruction existant nommée ***DiffIR (Efficient Diffusion Model for Image Restoration)***.

- Ce modèle a été **entrainé** pour restaurer des images abîmées (floues, basses qualité, incomplètes).

- Pour cela, les auteurs **dégradent** volontairement un ensemble d'images et entrainent le modèle à **générer** l'image d'origine.

- Lorsqu'on lui donne une image dégradée (avec un **masque** par exemple), il va chercher à reconstruire les parties de l'image qui sont perdus ou endommagées.

Originales Dégradées Reconstruit

MÉTHODOLOGIE DU PAPIER DIFFIR

- Le pipeline de DiffIR se divise en **deux parties** distinctes :
  - Le **pré-entraînement** : le modèle apprend à extraire une **représentation latente**, appelée Z, à partir d'images haute qualité, grâce à un **encodeur** appelé $CPEN_{S1}$. L'objectif du pré-entraînement est donc d'enseigner au modèle ce à quoi ressemble un bon Z lorsque l'on dispose de l'image originale intacte.
  - L'**entraînement** principal : ici, un modèle de diffusion est formé à **estimer** ce vecteur Z uniquement à partir de la **représentation latente** de l'image basse qualité, appelée D, obtenue par un autre encodeur appelé $CPEN_{S2}$. Pour cela, Z est volontairement **bruité**, puis le modèle apprend à inverser ce bruit en utilisant l'image dégradée comme condition.

MÉTHODOLOGIE (SCHÉMA DE DIFFIR)

DIRformer

- Modèle qui utilise l'image d'origine
- Pré-entraînement du modèle

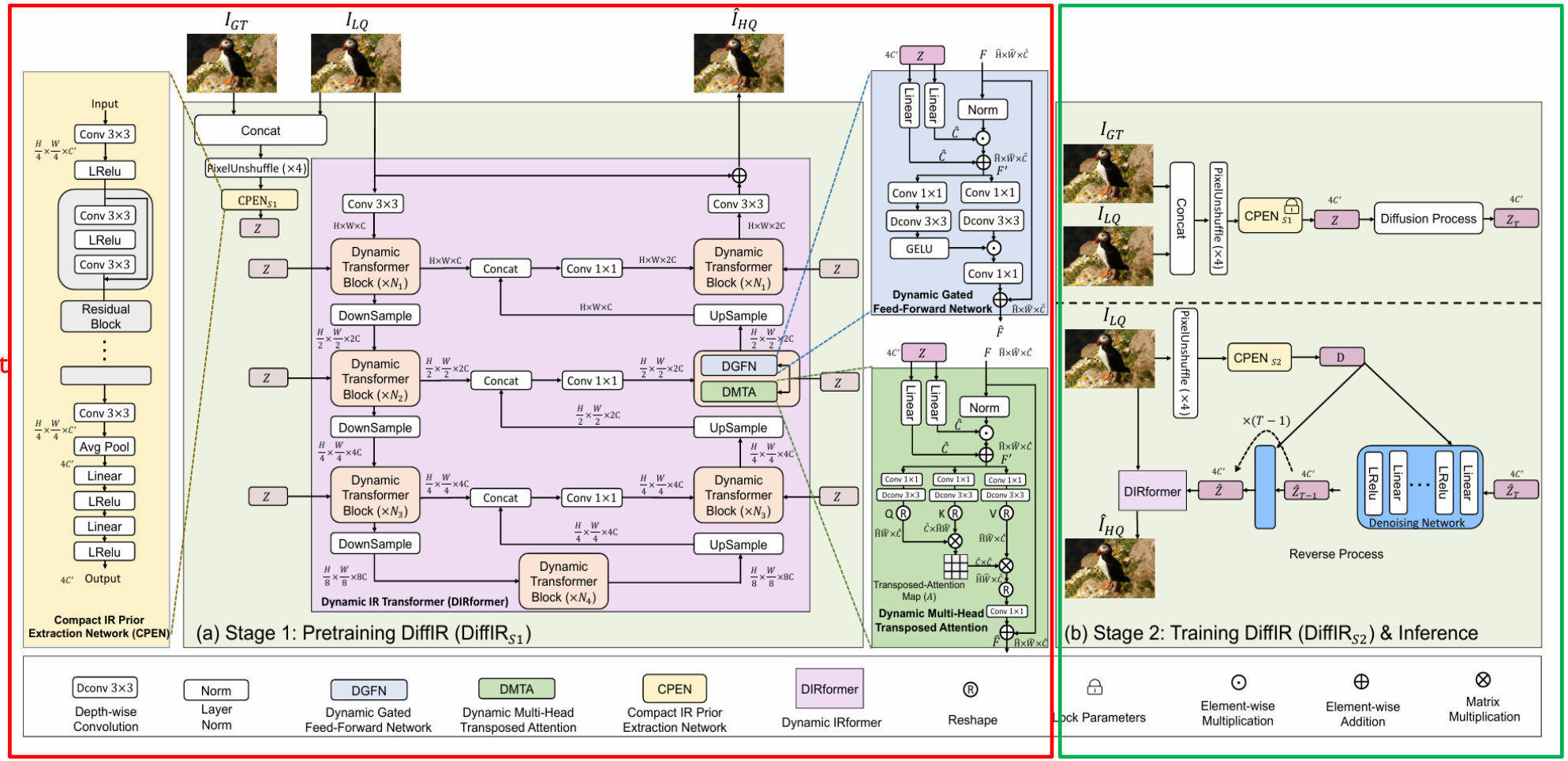


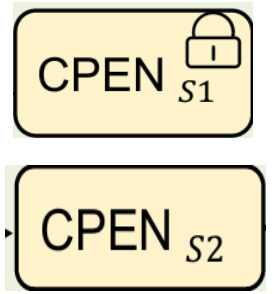
Figure 4 : Architecture du modèle de *DiffIR*

Modèle de diffusion

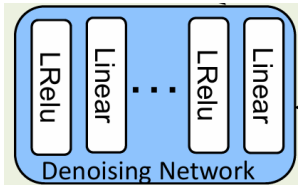
- L'ajout de ce modèle permet de se passer de l'image d'origine
- Entrainement et inférence

DÉVELOPPEMENT DE NOTRE MÉTHODE

- L'objectif est d'**améliorer** certaines parties du modèle afin d'obtenir de meilleurs résultats.
  - Utiliser de meilleurs **encodeurs** pour obtenir une meilleure **compression** de l'image.
  - Modifier le modèle de diffusion afin qu'il propose une meilleure **reconstruction** de l'image.



Encodeurs



Modèle de diffusion

- De plus, notre modèle sera entraîné sur notre propre **jeu de données**.
  - En effet, les modèles pré-entraînés sont généralement formés sur des jeux de données **génériques** qui peuvent ne pas refléter les caractéristiques particulières des images que l'on veut utiliser.
  - L'objectif est donc de l'**adapter** pour l'analyse d'images de **fresques**.

Image de fresque

DÉVELOPPEMENT DE NOTRE MÉTHODE – UTILISATION DES TRANSFORMERS

- Pour améliorer ces parties du modèle, des **Transformers** vont être utilisés.
- Les encodeurs ($CPEN_{S1}, CPEN_{S2}$) seront alors remplacés par des **Vision Transformers**, qui vont permettre de fournir la meilleure représentation latente pour nos images.
- L'objectif est aussi de **reconstruire** le modèle de diffusion de l'étape d'entraînement en y incluant ces Transformers, afin qu'il analyse les vecteurs d'espace latent de manière efficace.
- Cela permettra d'identifier **quelles zones sont les plus importantes** à observer, et **comment elles sont liées entre elles**.

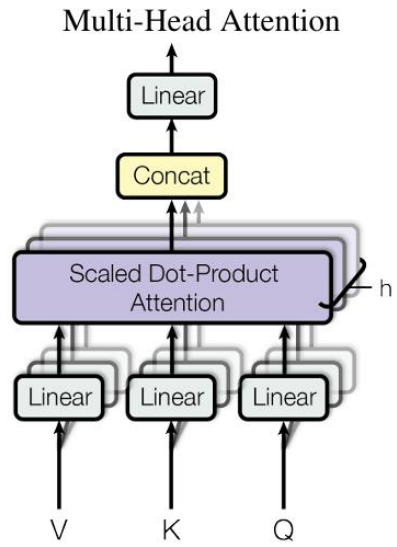


Figure 5 : Illustration du mécanisme d'attention dans les Transformers

DÉVELOPPEMENT DE NOTRE MÉTHODE – UTILISATION DES TRANSFORMERSCONCEPTION DU MODÈLE DE DIFFUSION

- L'architecture choisie est un U-Net car il correspond bien au format **encodeur-décodeur** d'un modèle de diffusion.
- Des **Transformers** sont intégrés dans ce modèle de diffusion qui utilisent comme mécanisme d'attention la **cross-attention**.
- Ce mécanisme va permettre de faire un lien entre l'image dégradée et l'image d'origine**.**
- Le modèle indiquera alors la quantité de bruit à retirer afin d'obtenir la version d'origine.

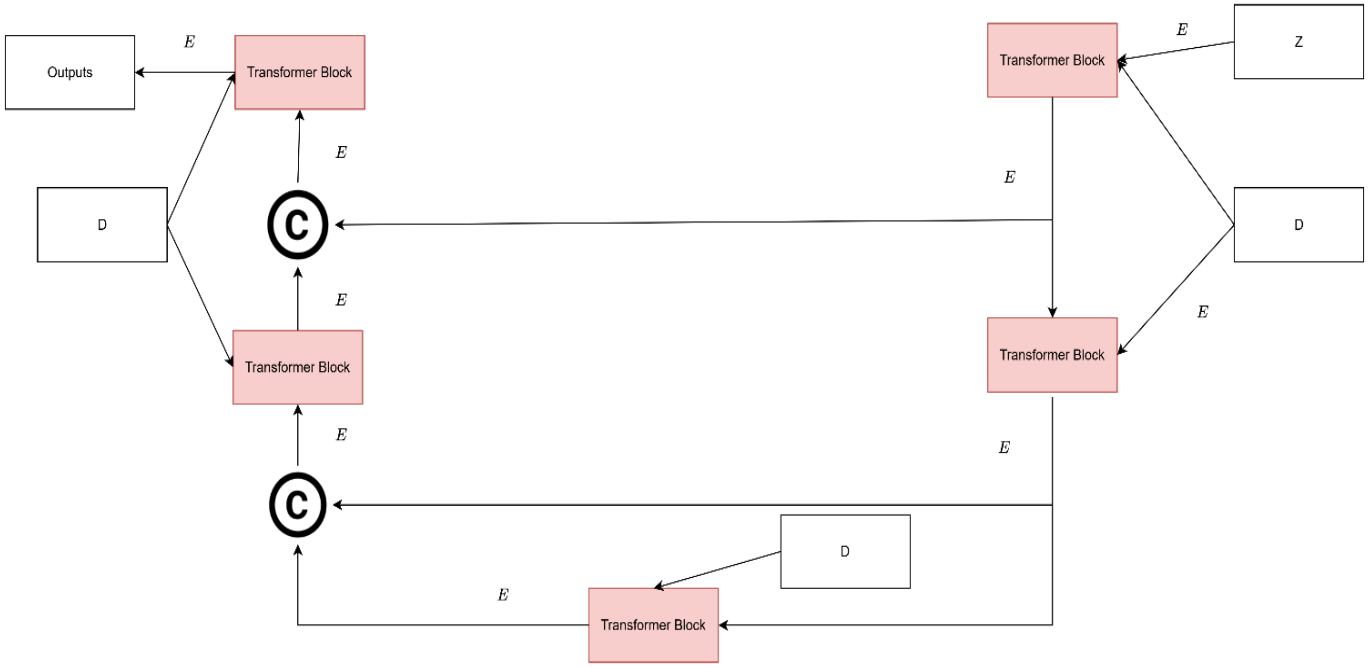


Figure 6 : Représentation visuelle de l'architecture du modèle de diffusion proposé

DÉVELOPPEMENT DE NOTRE MÉTHODE – FONCTIONS DE PERTE

- Afin que le modèle s'auto-corrige pendant l'entraînement, un ensemble de **fonctions de perte** a été utilisé.

- Pour le pré-entraînement, les fonctions utilisées sont les suivantes :
  - L2 Loss
  - TV Loss
  - SSIM Loss
  - Perceptual Loss
  - La perte totale est définie comme suit :

$$Total\ Loss\ =\ L2\ loss + TV\ Loss + SSIM\ Loss + Perceptual\ Loss$$

- Dans le cadre de l'entraînement, les fonctions de perte du pré-entraînement sont aussi utilisées en plus d'une autre qui est la Diffusion Loss.
  - La perte pendant l'entraînement est donc :

$$Total\ Loss =\ L2\ loss + TV\ Loss + SSIM\ Loss + Perceptual\ Loss + Diffusion\ Loss$$

RÉSULTATS

- La phase de pré-entraînement est en cours de recherche.
- Au départ, seule une fonction de **L1 Loss** était utilisée, donnant alors les courbes de perte suivantes.
- La courbe commençait très bas et convergeait au bout de 400 epochs. L'observation de ces courbes pourrait laisser penser que le modèle est prêt pour la reconstruction.

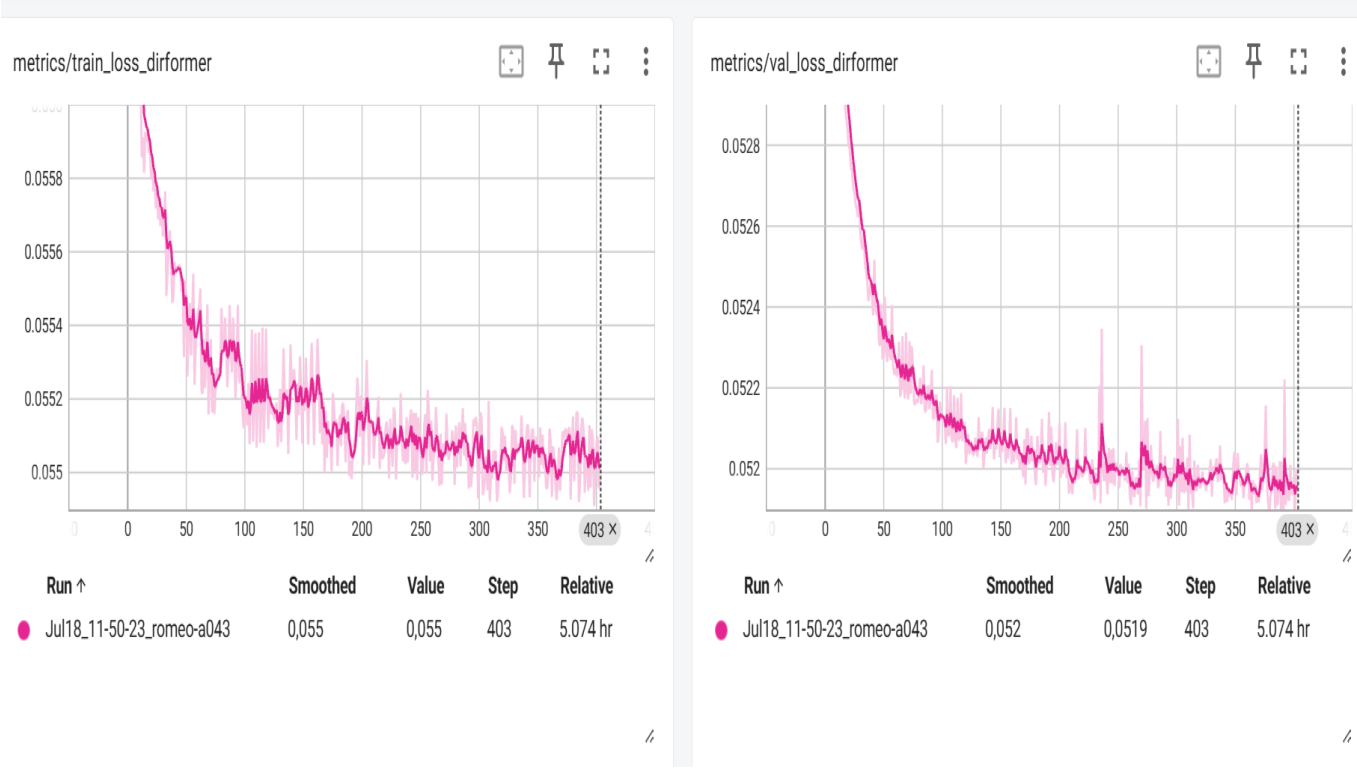


Figure 7 : Courbes de pertes avec seulement une L1 loss

RÉSULTATS

Image originale

Image dégradée

Image restaurée

Figure 8 : Tentative de restauration avec la fonction de perte L1

RÉSULTATS

- Après avoir ajouté les fonctions de perte afin de complexifier le modèle, ces courbes de perte peuvent être observées.
- Comme montré sur la Figure ci-dessus, la courbe de perte commence bien plus.
- Le fait que la courbe commence plus haut est mieux, mais elle converge beaucoup trop rapidement vers 10 epochs.

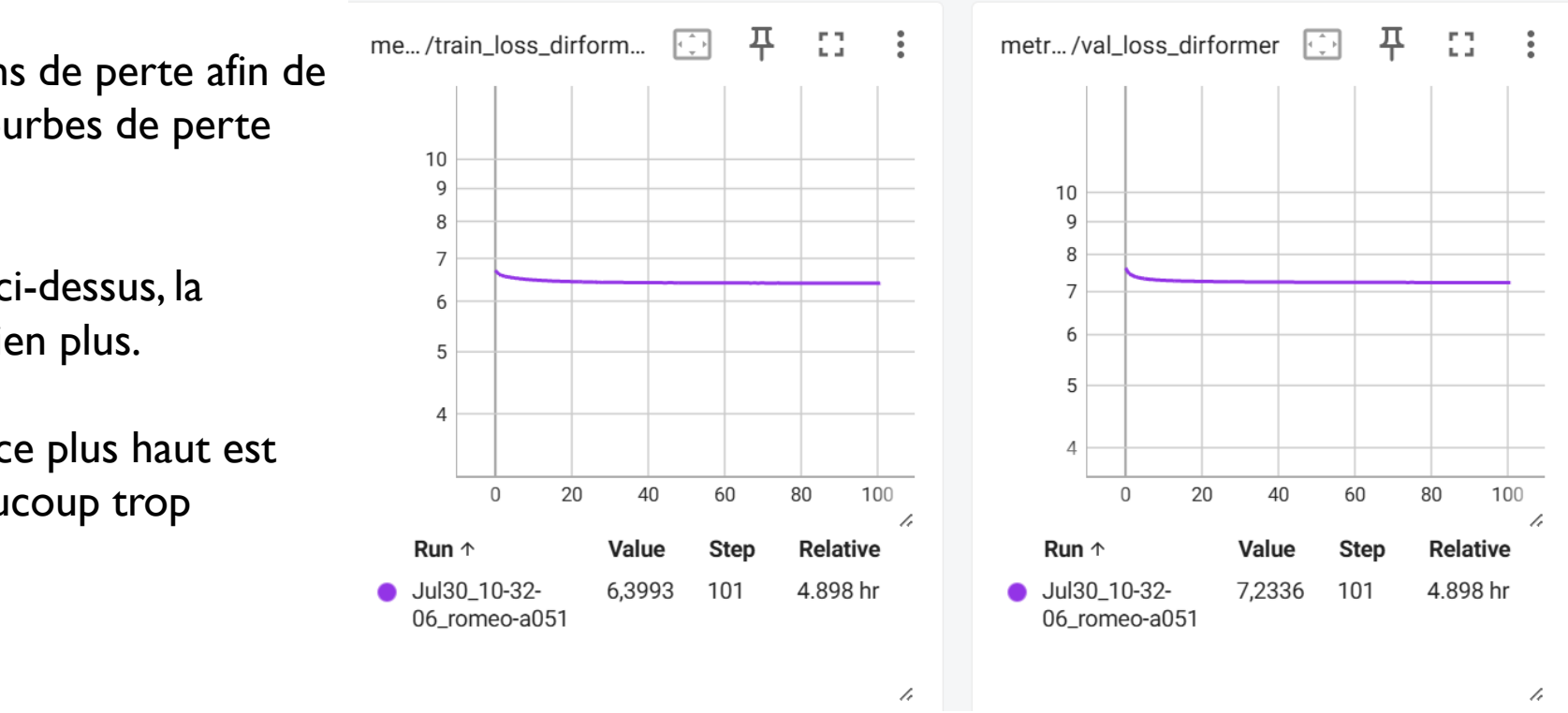


Figure 9 : Courbes de pertes pendant le préentrainement

RÉSULTATS

Image originale

Image dégradée

Image restaurée

Figure 10 : Tentative de restauration avec l'ensemble des fonctions de perte

CONCLUSION

- Beaucoup de choses sont à faire, il va nous falloir trouver pourquoi la perte stagne très rapidement.

- Plusieurs pistes sont envisagées :
  - La qualité du jeu de données,
  - Le choix des hyperparamètres.

- Lorsque le pré-entraînement du modèle fonctionnera, il faudra alors tester de réaliser l'entraînement et vérifier si le modèle de diffusion permet alors à DiffIR de fournir des restaurations recevables.

- Il faudra alors faire évaluer les résultats par des historiens associés au projet.

- Ensuite, il faudra comparer le modèle final par rapport à d'autres méthodes de restaurations qui ont été proposées dans diverses publications, puis proposer des axes d'améliorations, dans le but d'une future publication.

RÉFÉRENCES

- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.

- Nichol, A. Q., & Dhariwal, P. (2021, July). Improved denoising diffusion probabilistic models. In *International conference on machine learning* (pp. 8162-8171). PMLR.

- Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., ... & Van Gool, L. (2023). Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 13095-13105).

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

MERCI DE VOTRE ÉCOUTE

AVEZ-VOUS DES QUESTIONS ?