

# Transfert de Style d'Image basé sur des Réseaux de Neurones Convolutifs

---

Ariane Alix, Marie Heurtevent

21 janvier 2020

# Introduction

Le transfert de style d'image est une méthode qui vise à modifier le style d'une image (lignes, formes, couleurs) tout en préservant l'intégrité de son contenu.

Leon Gatys publie en 2015 la première version se basant sur des réseaux de neurones convolutifs, qui lui permettent de différencier explicitement le contenu du style.

## État de l'art - séparation style/contenu

En 2014, D. P. Kingma et al. séparent le contenu du style des caractères manuscrits pour pouvoir mieux les identifier [Kin+14].

2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4
2 2 2 2 2 2 2 2 2 2	3 3 3 3 3 3 3 3 3 3	4 4 4 4 4 4 4 4 4 4

Handwriting styles for MNIST obtained by fixing the class label and varying the 2D latent variable

FIG. 1 : Extrait du travail de D. P. Kingma sur le style des chiffres manuscrits

## État de l'art - synthèse de texture

**Méthodes non-paramétriques** : se basent sur un ré-échantillonnage de pixels ou de patchs de l'image de texture. Exemple de A. Efros et W. Freeman [EF01] :



## État de l'art - synthèse de texture

**Méthodes non-paramétriques** : se basent sur un ré-échantillonnage de pixels ou de patchs de l'image de texture. Exemple de A. Efros et W. Freeman [EF01] :



**Méthodes paramétriques** : définissent un modèle explicite de la texture en observant différentes mesures statistiques de l'image. Gatys publie un papier en 2015 où les mesures sont les réponses aux couches d'un réseau neuronal.

# État de l'art - inversion de représentation d'image

Gatys se base sur une méthode publiée en 2014 par Aravindh Mahendran et Andrea Vedaldi ([MV14]), qui permet de reconstruire des images correspondant à chaque couche d'un CNN.

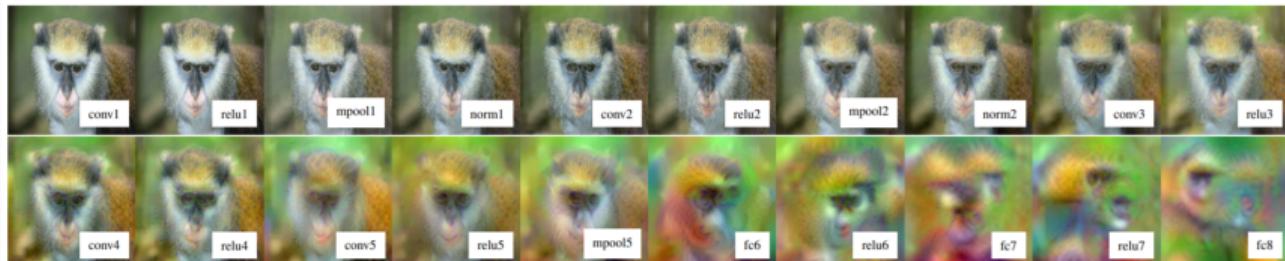


FIG. 2 : Reconstruction d'image depuis les réponses à chaque couche d'un CNN

## Représentation du style et du contenu d'une image

**Extraction du contenu d'une image** : depuis les couches profondes d'un CNN qui caractérisent la structure des objets, leur arrangement etc. mais pas leur représentation exacte.

## Représentation du style et du contenu d'une image

**Extraction du contenu d'une image** : depuis les couches profondes d'un CNN qui caractérisent la structure des objets, leur arrangement etc. mais pas leur représentation exacte.

**Extraction du style d'une image** : peut être extrait de toutes les couches mais nécessite la définition d'un espace de fonctions ; matrices de Gram représentant l'information de texture pour une image  $x$  à chaque couche  $l$  :

$$G_{x,ij}^l = \sum_k X_{ik}^l X_{kj}^l$$

## Principe de l'algorithme

Génération d'une nouvelle image proche à la fois de l'image de contenu et de l'image de style. Se formalise par la **minimisation d'une fonction de perte** :

$$\mathcal{L}_{\text{total}}(x, c, s) = \alpha \sum_l w_l^c \mathcal{L}_{\text{contenu}}(c, x, l) + \beta \sum_l w_l^s \mathcal{L}_{\text{style}}(s, x, l)$$

Avec

$$\begin{cases} \mathcal{L}_{\text{contenu}}(c, x, l) = \frac{1}{2} \sum_{ij} (C_{ij}^l - X_{ij}^l)^2 \\ \mathcal{L}_{\text{style}}(s, x, l) = \frac{1}{4N_l^2 M_l^2} \sum_{ij} (G_{s,ij}^l - G_{x,ij}^l)^2 \end{cases}$$

$\alpha, \beta, (w_l^c)_l$  et  $(w_l^s)_l$  paramètres à choisir.

# Résultats- compromis entre contenu et style

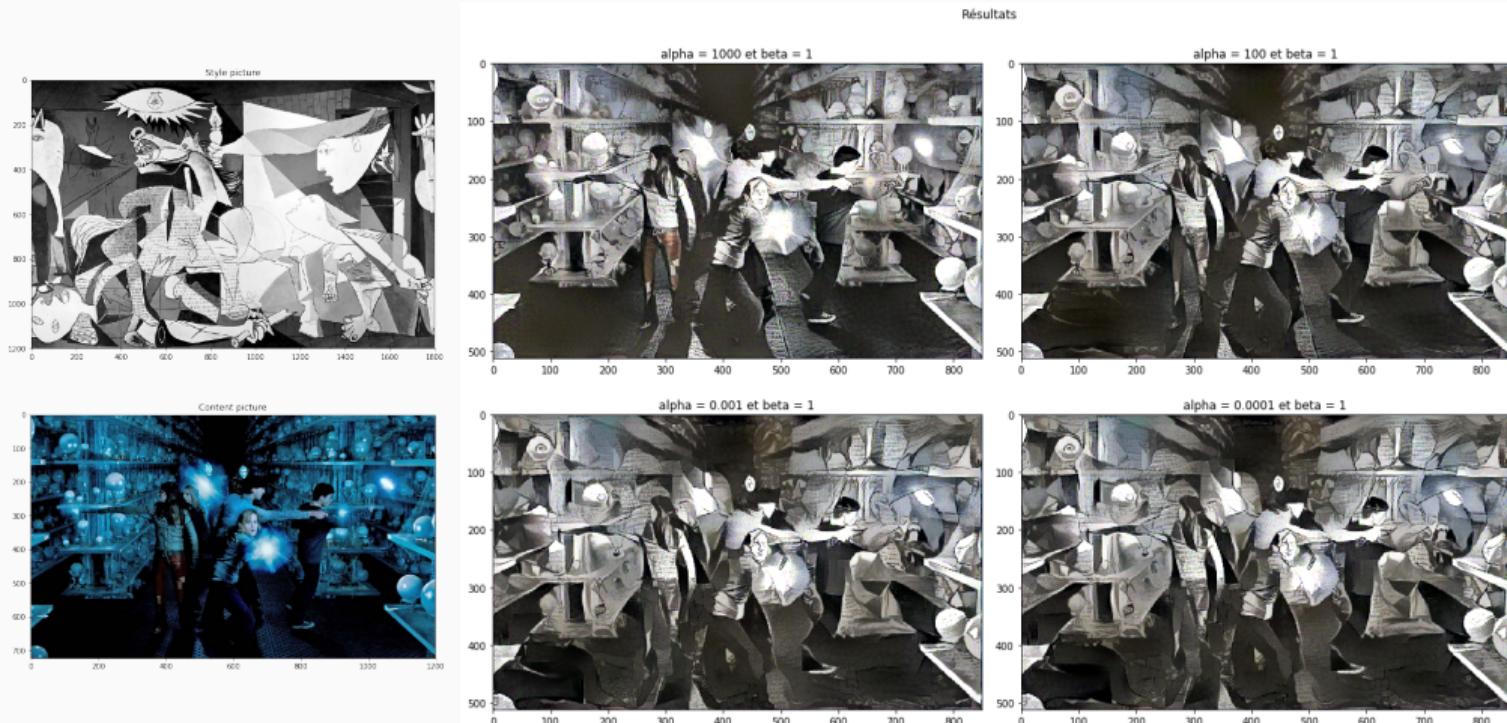


FIG. 3 : Transfert de style pour différents ratios  $\frac{\alpha}{\beta}$

## Résultats- limitation

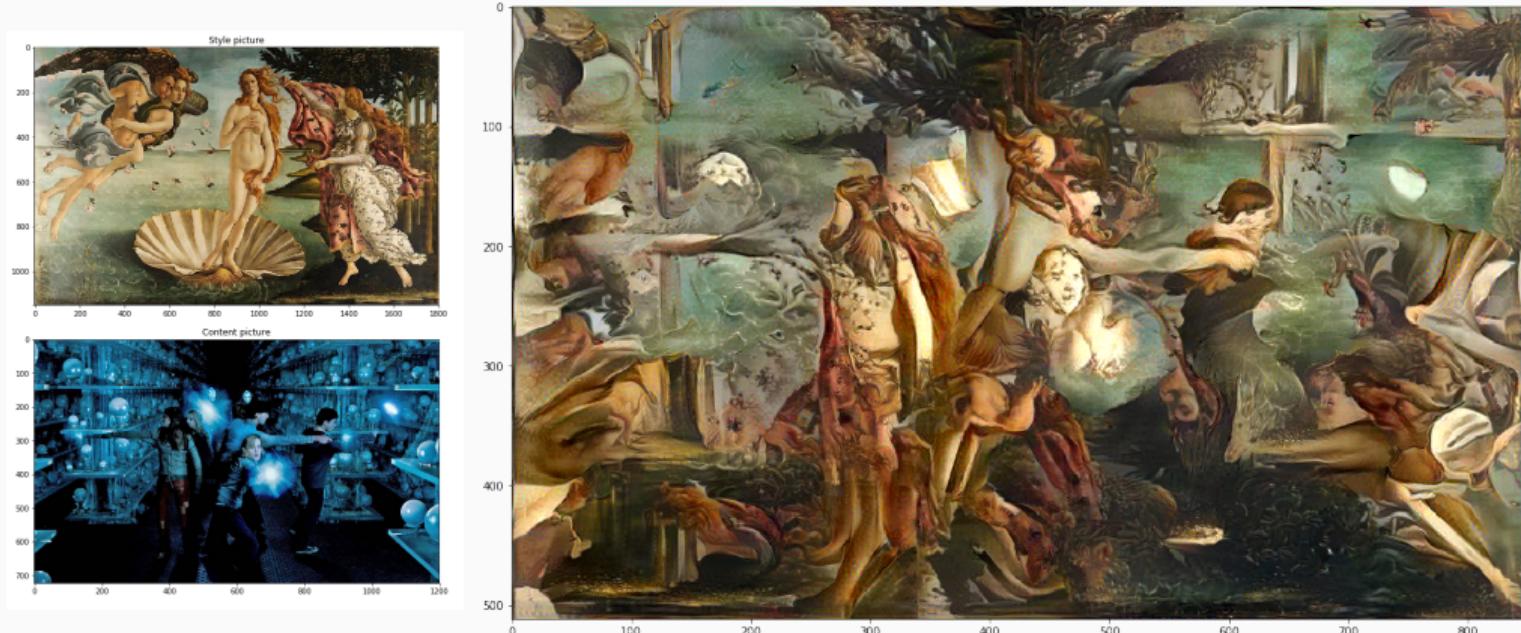


FIG. 4 : Transfert de style avec  $\frac{\alpha}{\beta} = 1000$

## Comparaison à une égalisation d'histogrammes

---

# Égalisation d'histogramme

## Motivation

Le style d'une image repose sur les textures associées à son contenu sémantique mais aussi sur sa palette de couleurs

# Égalisation d'histogramme

## Motivation

Le style d'une image repose sur les textures associées à son contenu sémantique mais aussi sur sa palette de couleurs

→ Pour les images avec peu de textures, une simple égalisation d'histogrammes pourrait suffire

# Comparaison sur des images à texture



**TAB. 1 :** A. Photo de Tübingen dans le style de "Paysage marin à Port-en-Bessin" de Georges Seurat. B. "Le Déjeuner des canotiers" de Auguste Renoir dans le style de "Mahana No Atua" de Paul Gauguin.

## Comparaison sur des photos

C.



Transfert de style



Égalisation d'histogramme



D.



**TAB. 2 : C.** Photo de New York de nuit dans le style d'une photo de Londres de jour. **D.** Photo de Sharbat Gula dans le style d'une photo de Frida Kahlo.

## Comparaison à l'utilisation d'un SinGAN

---

## Génération d'images

Un SinGAN permet de générer des images de différentes tailles et ratios respectant la distribution de patchs de l'image originale.



FIG. 5 : Random samples of different sizes and ratios

# Impact de l'échelle de départ lors de l'inférence

La génération d'images peut commencer à n'importe quelle échelle.



FIG. 6 : Échelle de départ = 0

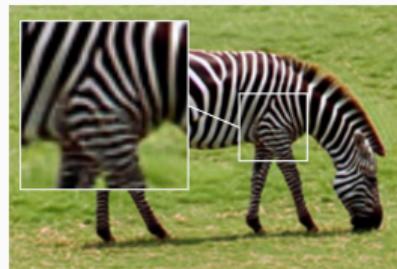
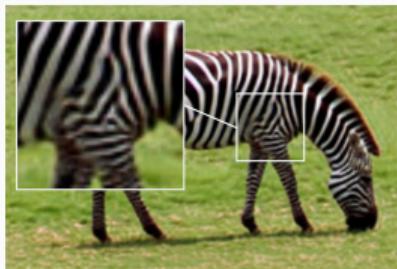


FIG. 7 : Échelle de départ = 2

## Harmonisation - sur l'exemple de Renoir / Gauguin



**TAB. 3 :** Transfert de style utilisant un SinGAN : "Le Déjeuner des canotiers" de Auguste Renoir dans le style de "Mahana No Atua" de Paul Gauguin. En commençant de l'échelle 1 [en haut à gauche] à l'échelle 8 [en bas à droite].

# Comparaison sur des images à texture



**TAB. 4 :** A. Photo de Tübingen dans le style de "Paysage marin à Port-en-Bessin" de Georges Seurat. B. "Le Déjeuner des canotiers" de Auguste Renoir dans le style de "Mahana No Atua" de Paul Gauguin.

## Comparaison sur des photos

C.



Transfert de style



SinGAN



D.



**TAB. 5 : C.** Photo de New York de nuit dans le style d'une photo de Londres de jour. **D.** Photo de Sharbat Gula dans le style d'une photo de Frida Kahlo.

## Conclusion

Le modèle de Gatys reste une référence et a des performances remarquables.

Cependant,

## Conclusion

Le modèle de Gatys reste une référence et a des performances remarquables.

Cependant,

- Si le modèle associe bien texture et contenu sémantique, il pourrait ajouter une contrainte sur **le lien entre couleurs et contenu sémantique**  
→ Le transfert de couleurs est parfois trop important

## Conclusion

Le modèle de Gatys reste une référence et a des performances remarquables.

Cependant,

- Si le modèle associe bien texture et contenu sémantique, il pourrait ajouter une contrainte sur **le lien entre couleurs et contenu sémantique**  
→ Le transfert de couleurs est parfois trop important
- Le modèle fonctionne mieux sur **des images avec de fortes textures**  
→ Dans le cas d'images peu texturées, d'autres algorithmes pourraient présenter de meilleurs résultats

## Références

---

-  Alexei A. EFROS et William T. FREEMAN. « Image Quilting for Texture Synthesis and Transfer ». In : *SIGGRAPH '01 : Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), p. 341-346.
-  Leon A. GATYS, Alexander S. ECKER et Matthias BETHGE. « Image Style Transfer Using Convolutional Neural Networks ». In : *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). DOI : [10.1109/cvpr.2016.265](https://doi.org/10.1109/cvpr.2016.265).

- 
- Diederik P. KINGMA et al. « Semi-Supervised Learning with Deep Generative Models ». In : *CoRR* abs/1406.5298 (2014). arXiv : 1406 . 5298. URL : <http://arxiv.org/abs/1406.5298>.
- 
- Alex KRIZHEVSKY, Ilya SUTSKEVER et Geoffrey E. HINTON. « ImageNet classification with deep convolutional neural networks ». In : *Communications of the ACM* 60.6 (2017), p. 84-90. DOI : [10.1145/3065386](https://doi.org/10.1145/3065386).
- 
- Aravindh MAHENDRAN et Andrea VEDALDI. « Understanding Deep Image Representations by Inverting Them ». In : *CoRR* abs/1412.0035 (2014). arXiv : 1412 . 0035. URL : <http://arxiv.org/abs/1412.0035>.
- 
- Manuel RUDER, Alexey DOSOVITSKIY et Thomas BROX. « Artistic Style Transfer for Videos ». In : *Lecture Notes in Computer Science Pattern Recognition* (2016), p. 26-36. DOI : [10.1007/978-3-319-45886-1\\_3](https://doi.org/10.1007/978-3-319-45886-1_3).



Joshua B. TENENBAUM et William T. FREEMAN. « Separating Style and Content with Bilinear Models ». In : *Neural Computation* 12.6 (2000), p. 1247-1283. DOI : [10.1162/089976600300015349](https://doi.org/10.1162/089976600300015349).