

Deep Learning et styles artistiques

Les cartes de vœux 2018 du CETIC sont basées sur des photos de nos locaux transformées selon le style d'œuvres d'art de Van Gogh, Kandinski, Hokusai et Picasso. Cette transformation d'images a été réalisée à l'aide d'algorithmes de réseaux de neurones.

L'approche choisie est issue de la recherche, écrite dans une publication scientifique, en apprentissage automatique profond ou deep learning. Elle illustre à la fois une thématique de recherche du CETIC, les logiciels et algorithmes d'apprentissage automatique et la vocation du centre à analyser et rendre exploitable des résultats de recherche fondamentale.

Nous vous proposons d'introduire brièvement les concepts liés au Deep Learning, de décortiquer cette technique de stylisation artistique, d'en faire des parallèles avec des travaux menés au CETIC et d'en lister les champs d'application.

DEEP LEARNING

Le Deep Learning ou apprentissage profond est un sous-ensemble du Machine Learning. Le Machine Learning est un outil de programmation utilisé en Data Science et en Intelligence Artificielle, et qui décrit l'ensemble des techniques d'apprentissage automatique permettant d'effectuer une tâche sans avoir à la programmer explicitement.

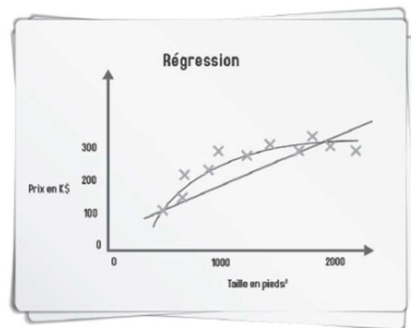


Fig. 1

Par exemple, dans la figure 1, l'exécution de deux algorithmes de Machine Learning permettrait de trouver automatiquement que l'ensemble de points est mieux modélisé par le résultat d'une régression quadratique (courbe) que par le résultat d'une régression linéaire (droite). Cela est rendu possible par un entraînement sur les données, par une optimisation d'une fonction-coût.

Le Deep Learning n'est pas foncièrement différent si ce n'est que cela repose sur des réseaux de neurones. Cette technologie est loin d'être nouvelle, mais l'augmentation de la puissance de calcul, ainsi que la découverte de solutions permettant l'entraînement de réseaux plus élaborés ont contribué à leur essor depuis une dizaine d'années. « Deep » fait d'ailleurs référence au nombre de couches et donc à la profondeur du réseau.

Il faut voir un neurone comme une fonction non-linéaire d'une somme pondérée d'entrées et le réseau comme la combinaison de plusieurs de ces fonctions. Pour faire l'analogie avec l'exemple de la figure 1, la régression linéaire serait équivalente à un simple neurone activé par une fonction linéaire.

Mentionnons encore l'existence de différents types de réseaux : les réseaux classiques, les réseaux convolutifs, récurrents... C'est un domaine qui jouit d'une très grande créativité dans tous ses aspects comme nous allons d'ailleurs le voir dans le détail de l'algorithme du moteur de stylisation.

FAST NEURAL STYLE

Le moteur de stylisation d'image est tiré de la librairie github opensource « fast-neural-style » [1]. Cette librairie fait référence à un travail de recherche de l'université de Stanford présenté lors de la 14^{ème} « European Conference on Computer Vision » à Amsterdam [2]. Ce travail propose une implémentation rapide et améliorée d'un algorithme de stylisation proposé initialement par Gatys et al. [3].

L'algorithme de stylisation artistique par réseaux de neurones est représenté dans le schéma de principe ci-dessous (fig. 2). Il consiste en la juxtaposition de deux réseaux de neurones profonds.

Le premier, appelé « **Image transformation network** », est chargé de transformer l'image d'entrée x en une image stylisée \hat{y} . Ce réseau est un « Deep Residual Convolutional Neural Network ». Il est dérivé du plus classique « Convolutional Neural Network » (CNN), composé de couches de convolution et souvent utilisé pour la classification d'images. « Residual » fait référence à une amélioration de ces réseaux qui consiste à introduire des « court-circuits » au niveau des couches de convolution, menant à la formation de blocs résiduels. Cette technique permet essentiellement de circonvenir aux problèmes récurrents lié à l'entraînement de réseaux de neurones profonds [4].

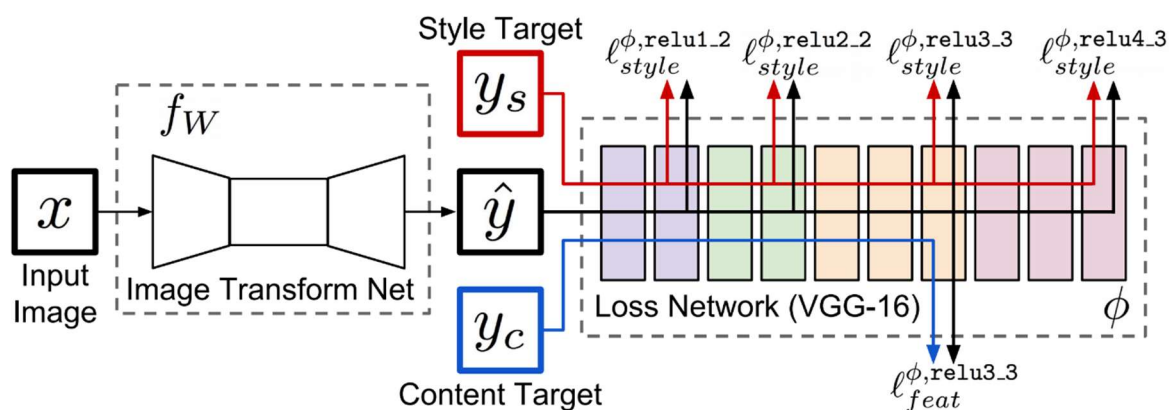


Fig. 2

La deuxième partie de cet algorithme est constituée d'un réseau neuronal convolutif pré-entraîné pour la classification d'images. Dans ce réseau, appelé **Loss Network**, certaines couches sont utilisées pour générer des fonctions de pertes perceptuelles en fonction du besoin. Ce dernier étant de styliser « à la manière de » une image tout en gardant sa structure spatiale, ces fonctions de pertes perceptuelles sont de deux types :

- Fonction de perte liée aux caractéristiques (« feature ») :
L'idée est de minimiser la différence des transformations par les plus hautes couches des images \hat{y} et y_c (égale à l'image d'entrée x). La sélection des plus hautes couches et la minimisation de la fonction de perte a pour effet de tendre à préserver le contenu de l'image initiale et sa structure spatiale, tandis que les couleurs, les textures et la forme exacte ne sont pas préservées.
- Fonction de perte liées au style :
L'idée est de minimiser la somme des différences des transformations des images \hat{y} et y_s (l'image cible stylisée) par diverses couches du réseau convolutif liées au style. Ce sont les caractéristiques stylistiques qui sont préservées et non la structure spatiale.

Enfin, il est important de comprendre que seul le premier réseau est modifié par l'entraînement global de l'algorithme, le deuxième étant figé. L'entraînement de ce réseau de transformation a donc pour but de minimiser les fonctions de pertes mentionnées ci-dessus et donc de converger vers une transformation balancée entre les caractéristiques spatiales de l'image de départ et les caractéristiques stylistiques de l'image cible. Un entraînement signifie une image stylisée cible, 80000 images d'entraînement du dataset MS-COCO et environ 4h sur une carte GPU très puissante comme le rapportent les auteurs [2].

LE CETIC et le Deep Learning

L'approche d'analyse de données du Deep Learning est présente depuis plusieurs années dans plusieurs travaux et projets de recherche. Les exemples les plus récents sont :

- Reconnaissance de panneaux routiers par réseaux de neurones convolutifs (2017)
- Utilisation de réseaux autoencoders pour la détection d'anomalie sur des signaux d'accéléromètre pour la détection de chutes (2018)
- Reconnaissance faciale par réseaux de neurones convolutifs (2018)
- Utilisation de réseaux de neurones récurrents RNN-LSTM pour la détection d'anomalie sur capteurs industriels (2018)
- Utilisation des réseaux de neurones récurrents RNN-LSTM pour la détection de crises d'épilepsie (2018)

Le CETIC est engagé dans des projets de recherche utilisant le Deep Learning dans les champs d'applications tels que la santé, la mobilité, l'industrie ... et peut-être le votre.

Références

[1] <https://github.com/jcjohnson/fast-neural-style>

[2] [Perceptual Losses for Real-Time Style Transfer and Super-Resolution](#), Justin Johnson, Alexandre Alahi, Li Fei-Fei, Presented at ECCV 2016

[3] A Neural Algorithm of Artistic Style, Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

[4] <http://torch.ch/blog/2016/02/04/resnets.html>, Training and investigating residual nets, Gross, S., Wilber, M. (2016)