



UNIVERSITÉ PARIS 8 - VINCENNES À SAINT-DENIS

Master Informatique des Systèmes Embarqués

Mémoire de projet tuteuré

Fakhri YAHIAOUI - Roman BOURSIER

Date de soutenance : le 09/06/2020

Tuteur – Université : Farès BELHADJ

Résumé

A faire en dernier ...

Remerciements

Idem ...

Table des matières

Résumé	i
Remerciements	iii
Introduction	1
0.1 Contexte	1
0.2 Problématique	1
1 Etat de l'art	5
1.1 Généralité sur les GANs	5
1.2 Pix2Pix	6
1.3 GauGan	7
1.4 Le transfert de style neuronal	7
2 Propositions de solutions	9
2.1 Création du dataset en semi automatique	9
2.2 Photo-Sketching : Inferring Contour Drawings from Images	9
2.3 Abstraire une image via les CNN	9
3 Conclusion et Perspectives	11

Introduction

0.1 Contexte

Dans le cadre de notre projet de fin d'étude, nous souhaitons utiliser un modèle d'apprentissage automatique du type Deep Learning, afin de produire un moteur de rendu capable d'adopter une stylisation « type » telle que la peinture chinoise. Dans un premier temps, il s'agira de proposer un modèle d'abstraction des peintures sélectionnées comme base d'apprentissage et d'utiliser le couple « peinture originale » / « abstraction » pour l'entraînement. Par la suite, un moteur de rendu d'abstractions sera connecté au réseau profond qui produira une peinture sur la base de l'abstraction.

Le modèle généré devra d'une part adopter la stylisation retenue mais aussi interpréter l'abstraction d'origine.

0.2 Problématique

De nombreux modèles comme pix2pix ou Gaugan permettent déjà de transférer un style d'une image à l'autre avec de bons résultats. La principale difficulté est de constituer un dataset de pairs abstraction/peintures sur lequel entraîner ces modèles. Ce dataset n'existe pas et l'enjeux est de notre point de vue, de trouver un moyen de le constituer.

La figure n°1 montre qu'il est possible de générer un paysage à partir de quelques lignes. Cependant, nous sommes confronté à un problème avec la figure n°2.

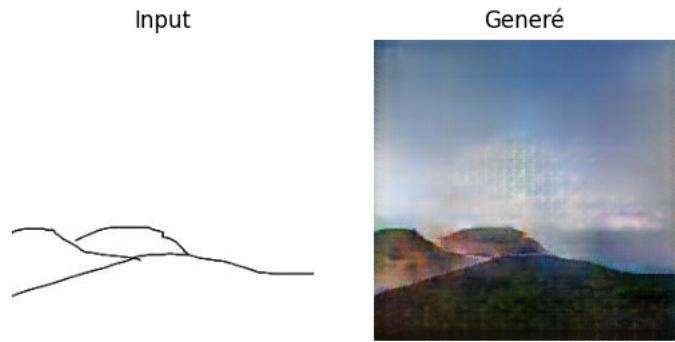


FIGURE 1 – Nous avons testé pix2pix sur un dataset composé de photos de paysages que nous avons labellisées en appliquant un filtre canny. Le résultat présenté montre le rendu une fois l'apprentissage terminé, avec un dessin fait à main levée en entrée

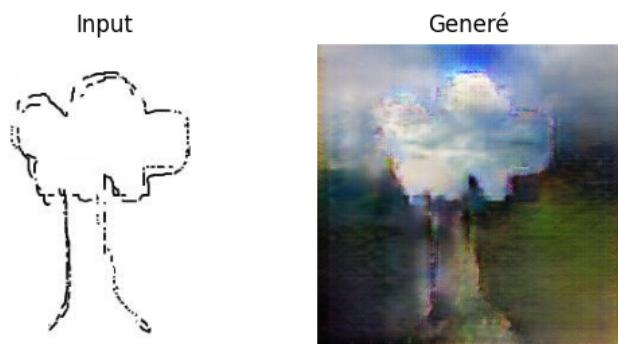


FIGURE 2 – Utilisation de pix2pix avec une abstraction d'arbre

En effet, le modèle n'a pas été confronté à ce type d'entrée lors de son apprentissage, il n'est donc pas en mesure de reconstituer un arbre.

Nous avons demandé à un plusieurs personnes de dessiner schématiquement un paysage composé de montagnes et d'arbres éléments typiques d'une peinture chinoise :



Nous pouvons dire que ce type de dessins constitue des "abstractions", que nous souhaitons convertir en peinture chinoises. Nous pouvons déjà noter qu'il n'existe pas une seule façon de représenter un paysage mais plusieurs. Selon le dessinateur, les dessins possèdent des niveaux "d'informations" plus ou moins élevé, c'est à dire plus ou moins "abstraits". La figure n°1 nous indique des arbres, mais nous ne savons rien de la forme potentiel que ceux-ci pourraient avoir (bien que nous pouvons dire qu'il s'agit plutôt d'un sapin que d'un pin). Tandis que sur la figure n°2 nous avons déjà plus de détails.

Chapitre 1

Etat de l'art

Nous présentons dans un premier temps l'architecture GANs, qui est au cœur de notre problématique. Nous présenterons également différents modèles existants sur lesquels nous avons testé nos données.

Enfin, nous évoquerons, de manière plus succincte, des recherches liées à notre sujet mais n'utilisant pas les réseaux de neurones afin de pouvoir identifier les avantages et inconvénients des deux mondes.

1.1 Généralité sur les GANs

Les réseaux adverses génératifs sont une classe d'algorithme d'apprentissage non-supervisé. Ces algorithmes ont été introduits dès 2014 par Goodfellow [GPAM⁺14] et permettent de générer des images avec un fort degré de réalisme. La structure des GANs se compose de deux parties, le générateur et le discriminateur. Le générateur génère des données et les soumet au discriminateur dont le but est d'évaluer le degré de crédibilité de la donnée.

$G(z, 01)$ se charge de "mapper" le bruit z vers l'espace désiré $D(x, 02)$ retourne et la probabilité dans l'intervalle $[0, 1]$ que x vient du dataset original. Enfin $0i$ représente les poids définis par chacun des modèles.

Les poids sont donc mis à jours pour :

1 - Maximiser la probabilité que les données x , soient classifiées comme appartenant au dataset d'origine. Soit maximiser la fonction de perte $D(x)$

2 - Minimiser la probabilité que de fausses images appartiennent au dataset d'origine. Soit minimiser la fonction de perte $D(G(z))$

Pour résumer c'est une sorte de jeu, dans lequel le générateur essaie de maximiser la probabilité que sa sortie soit reconnue comme réel, tandis que le discriminateur essaie de minimiser la même valeur.

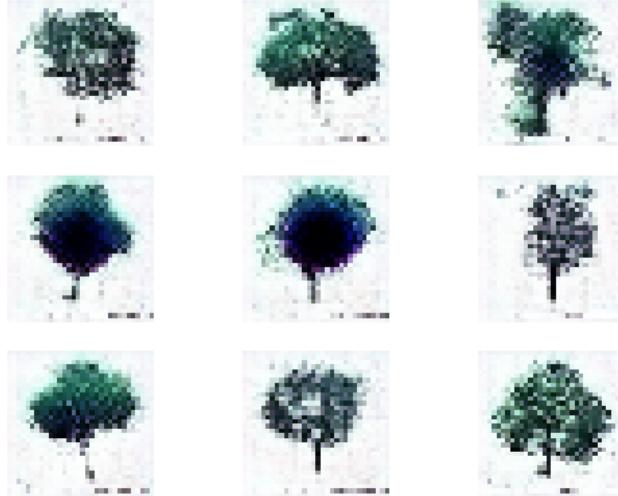


FIGURE 1.1 – Test de génération d’arbres utilisant l’architecture DCGAN

Il existe aujourd’hui une très grande variété de travaux de recherches implémentants les Gans (CGAN, Bayesian GAN, GYCLE GAN ...). Nous présentons ici des solutions existantes permettant de répondre partiellement à notre problématique.

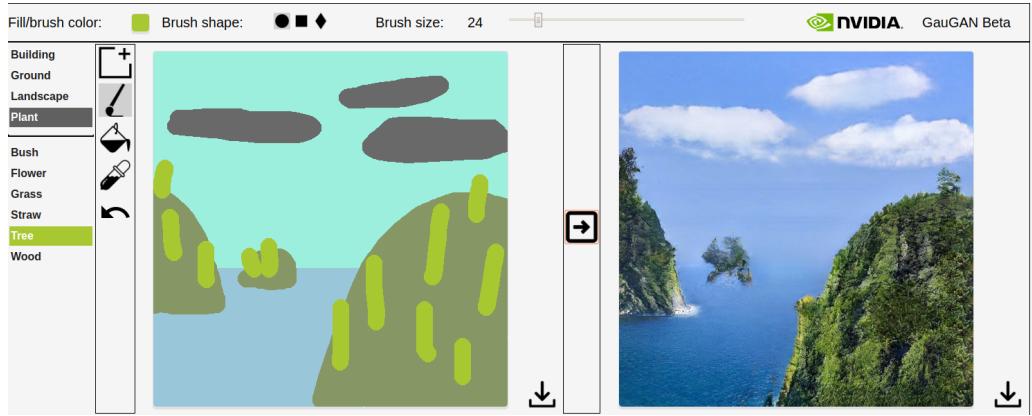
1.2 Pix2Pix

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou et Alexei A. Efros proposent en 2014 une nouvelle approche intitulée Image-to-Image Translation with Conditional Adversarial Networks.[GEB15] Leur contribution principale fut de montrer en quoi les cGans peuvent permettre de résoudre efficacement les problèmes de translation d’image et ainsi de proposer un framework pouvant être appliqué à n’importe quel domaine.

Les Gans conditionnels apprennent une cartographie à partir d’images observées x et d’un vecteur de bruit aléatoire z , vers $y : G : x, z \rightarrow y$.

L’architecture utilise "U-Net" pour le générateur et "PatchGAN" pour le discriminateur.

1.3 GauGan



1.4 Le transfert de style neuronal

Cette technique fut décrite par Leon A. Gatys dans son article "A Neural Algorithm of Artistic Style". [GEB15]

Celle-ci utilise trois images, la première est l'image d'entrée (bruit), la seconde représente le style de référence (comme une peinture par exemple), la dernière correspond à l'image que l'on souhaite transformer.

Le principe consiste à définir deux fonctions de distance. La première décrit comment le contenu de deux images sont différentes, et la seconde représente la différence de style. On obtient alors deux images, une pour le style souhaité, une pour le contenu.

L'objectif est de transformer l'image d'entrée en minimisant la distance avec l'image de contenu et avec l'image de style. On obtiens alors par rétro-propagation, une image qui correspond au contenu de l'image d'origine et au style souhaitée.





Cela nécessite la création d'un réseau convolutionnel, en effet le modèle doit être en mesure de capturer les invariances et de définir les caractéristiques de l'images (chats vs chien) afin d'en obtenir une compréhension générale.

Chapitre 2

Propositions de solutions

Voici les solutions envisagées :

2.1 Cr eation du dataset en semi automatique

En utilisant le logiciel LabelBox, il est possible de labeliser une image, en cr eant des s emantique layout[1]. Nous avons d j  un collection de plus de 1000 images de peinture chinoises, les labelis s nous prendrait environ 2 jours hommes  raison d'une minute par image. Une fois cette tape r aliss e, nous pourrions g n rer une ou plusieurs abstractions pour chaque peintures, en utilisant le dataset "quick draw" entre autres. Outre l'inconv nient du temps pass , cette solution n'est pas g n ralisable  autre chose que notre probl matique.

2.2 Photo-Sketching : Inferring Contour Drawings from Images

2.3 Abstraire une image via les CNN

Nous savons qu'un CNN est capable de d etecter des caract ristiques d'une image

Comme le prouve le travail de <https://github.com/dribnet/perceptionengines>, il est peut- tre possible de g n rer un repr sentation abstraite en utilisant cette propri t  des CNN.

Chapitre 3

Conclusion et Perspectives

Bibliographie

- [GEB15] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.
- [GPAM⁺14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.