



---

UNIVERSITÉ PARIS 8 - VINCENNES À SAINT-DENIS

**Master Informatique des Systèmes Embarqués**

**Memoire 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

<b>Résumé</b>	<b>1</b>
<b>Remerciements</b>	<b>2</b>
<b>Introduction</b>	<b>4</b>
0.1 Problématique . . . . .	4
<b>1 Etat de l’art</b>	<b>7</b>
1.1 Généralités sur les GANs . . . . .	7
1.2 CGANs et traduction d’image . . . . .	8
1.2.1 Pix2Pix . . . . .	8
1.2.2 GauGan . . . . .	8
1.2.3 Transfert de style neuronal . . . . .	8
1.2.4 Learning to Sketch . . . . .	9
1.2.5 Sketching : Inferring Contour Drawings from Images . . . . .	9
1.3 Autre algorithmes n’utilisant pas le deep learning . . . . .	10
1.4 Datasets disponibles . . . . .	10
1.4.1 Croquis/abstractions . . . . .	10
1.4.2 Peinture chinoises . . . . .	10
<b>2 Propositions de solutions</b>	<b>11</b>
2.1 Création du dataset en semi automatique . . . . .	11
2.2 Photo-Sketching : Inferring Contour Drawings from Images . . . . .	11
2.3 Abstraire une image via les CNN . . . . .	11
<b>3 Conclusion et Perspectives</b>	<b>13</b>

# Introduction

Dans le cadre de notre projet de fin d'étude, nous souhaitons utiliser un modèle de 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.

## 0.1 Problématique

La "traduction image-image" (Image-to-image translation) permet d'apprendre le mapping entre une image d'entrée et de sortie. En Figure 1, nous testons la génération d'un paysage à partir d'un croquis simple. En Figure 2, c'est un échec.

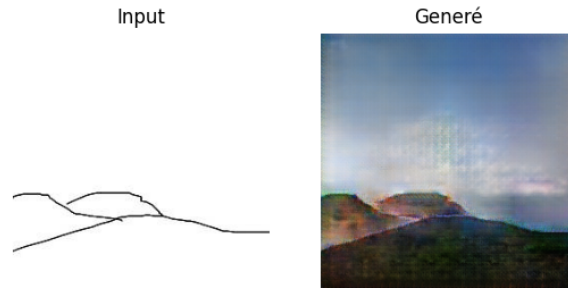


FIGURE 1 – Test du framework pix2pix [IZZE16] sur notre dataset composé de photos de paysages, labellisées en appliquant un filtre Canny [Can86] sur chacune d'elles.

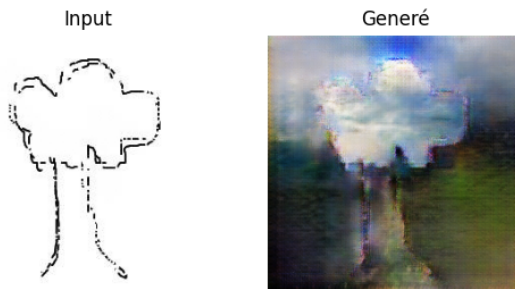


FIGURE 2 – Utilisation du même modèle avec une abstraction d'arbre

Comme évoqué dans [SPS<sup>+</sup>18] les algorithmes du type "traduction image-image" se basent essentiellement sur la corrélation d'une image à l'autre, et relève d'un apprentissage supervisé. Le rendu en Figure 2 s'explique par la nature du dataset et par la distance importante qui sépare une abstraction d'une photo.

En Figure 3 nous avons demandé à plusieurs personnes de dessiner schématiquement un paysage composé de montagnes et d'arbres, éléments courant de la peinture chinoise. Ces dessins sont des abstractions, que nous souhaitons traduire en peintures chinoises.

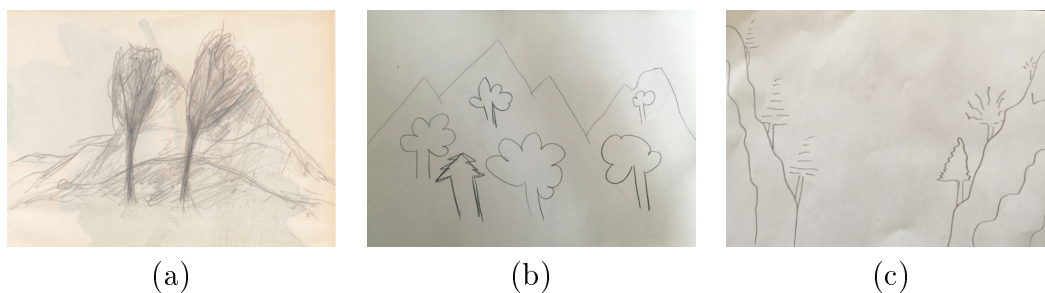


FIGURE 3 – Dessin de paysages réalisés par des personnes différentes. On note que les niveaux d'informations sont plus ou moins élevés. Le dessin (b) est très abstrait tandis que nous possédons plus de détails au niveau de (a) .

# Chapitre 1

## Etat de l'art

Nous présentons dans un premier temps les GANs qui sont au coeur de notre problématique. En un second lieu, nous étudions différents modèles existants basés sur les GANs conditionnels. Nous évoquons aussi, de manière plus succincte, des travaux liés à notre sujet mais n'utilisant pas les réseaux de neurones afin de pouvoir identifier les avantages et inconvénients des deux mondes. Nous terminons par une liste des datasets utiles à notre sujet.

### 1.1 Généralités sur les GANs

"Les GANs (en anglais generative adversarial networks) sont une classe d'algorithmes d'apprentissage non-supervisé. Ces algorithmes ont été introduits dès 2014 par Goodfellow et permettent de générer des images avec un fort degré de réalisme." [Wik20b]

Un générateur fabrique des données et les soumet au discriminateur dont le but est d'évaluer leurs degré de crédibilité.

Le générateur  $G(z, \theta_1)$  représente un réseau de neurones capable de mapper du bruit  $z$  vers l'espace désiré  $x$ . Le discriminateur  $D(z, \theta_2)$  retourne la probabilité dans l'intervalle  $[0, 1]$  que  $x$  vient du dataset original.  $\theta_i$  représente les poids définis par chacun des modèles. Le générateur tente de maximiser la probabilité que les données  $x$ , soient classifiées comme appartenant au dataset d'origine et inversement, le discriminateur minimise la probabilité que de fausses images appartiennent au dataset d'origine.

Il existe aujourd'hui une très grande variété de travaux de recherches basés les GANs (BCGAN, AmbientGAN, ORGAN, Perceptual GAN). Le dépôt "The GANs Zoo" [hin18] référençait déjà en 2018 plus de 502 noms de GANs différents !



## 1.2 CGANs et traduction d'image

Les GANs conditionnels ajoutent une information supplémentaire  $y$ , partagée par le discriminateur et le générateur. Grâce aux cGANs il est possible de générer des images réalistes basées sur des labels de classes, des textes ou des images.

### 1.2.1 Pix2Pix

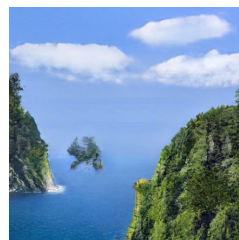
[GEB15] nous montre en quoi les cGANs peuvent permettre de résoudre efficacement les problèmes de traductions d'images et propose un framework applicable à n'importe quel domaine.

### 1.2.2 GauGan

[PLWZ19] permet de générer des paysages réalistes basés sur des masques de segmentation sémantiques. Les auteurs introduisent la "normalisation conditionnel" ou "adaptative", permettant de prendre en compte notamment les informations spatiales. Les couches de normalisations ont tendance à faire perdre de l'information contenu dans les masques sémantiques d'entrées car ils ne dépendent pas de données externes.



(a) Input - masques de segmentation sémantiques



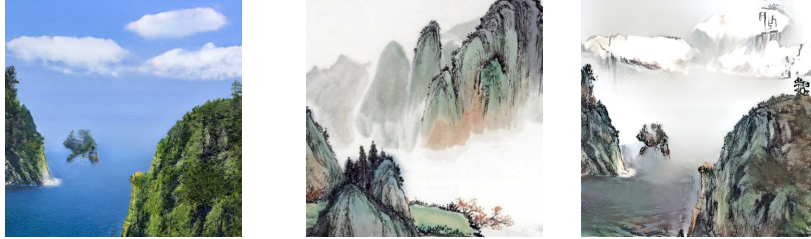
(b) Output

FIGURE 1.1 – Test de rendu de paysage à partir de l'application GauGAN : <http://nvidia-research-mingyuliu.com/gaugan/>

### 1.2.3 Transfert de style neuronal

Introduit par Leon A. Gatys [GEB15] le TDN consiste à transférer un style à partir d'une image de référence vers une image de contenu. L'objectif est de transformer l'image d'entrée (bruit) en minimisant la distance avec l'image de contenu et avec l'image de style. On obtient alors par rétropropagation, une image qui correspond au contenu de l'image d'origine et au style souhaitée.

L'avantage de cette technique est quelle ne nécessite pas de dataset, seulement deux images sont nécessaires.



(a) Image de contenu (b) Image de style (c) Résultat

FIGURE 1.2 – Test de transfert de style réalisé à partir de la figure 1.1(b)

### 1.2.4 Learning to Sketch

[SPS<sup>+</sup>18] présente une méthode permettant de transformer une photo en croquis, en essayant d'imiter la façon de faire d'un humain. Le modèle est capable de réaliser le croquis séquentiellement, c'est à dire trait par trait. Les auteurs proposent de résoudre le problème des styles subjectifs et variés des dessins humain en utilisant un modèle hybride supervisé/non-supervisé. L'objectif étant de palier "au signal de supervision faible et bruyant" induit par l'écart important entre un croquis et sa photo correspondante.

L'architecture est décomposé en 4 sous-modèles contenant chacun leur propres sous-réseaux d'encodeurs et de décodeurs. Deux réseaux supervisés traduisent respectivement une photo en croquis  $D(E(photo)) \rightarrow sketch$  et un croquis en photo  $(D(E(sketch)) \rightarrow photo)$ . Deux autres réseaux non-supervisés se chargent de la reconstruction.  $D(E(photo)) \rightarrow photo$  et  $D(E(sketch)) \rightarrow sketch$

### 1.2.5 Sketching : Inferring Contour Drawings from Images

[LLM<sup>+</sup>19] propose une nouvelle approche concernant la détection des contours dans une image. L'article montre que les solutions traditionnelles comme Canny [Can86], captent uniquement les signaux de haute fréquence dans l'image sans la comprendre. Les auteurs ont collecté un dataset de 5000 paires "croquis humain/photos", créé manuellement via la plateforme de crowdsourcing "Amazon Mechanical Turk". En effet aucun dataset existant ne convenait (nombre d'éléments dans l'image, limites internes manquantes, le contenu non reconnaissable, les zones ombrées vides etc ..).

Ce modèle également du type "traduction image-image", permet d'avoir plusieurs labels différents pour la même photo, ce qui correspond bien à la nature subjective du dessin humain.

## 1.3 Autre algorithmes n'utilisant pas le deep learning

TODO

## 1.4 Datasets disponibles

Il est évident qu'il n'existe pas à l'heure actuel de dataset de paires abstraction/peinture chinoise.

### 1.4.1 Croquis/abstractions

Un des principal problème de la recherche dans le domaine de la génération de dessins à la main est le manque de dataset. Aujourd'hui les plus grosses bases de données que nous avons trouvées sont TU-Berlin [EHA12], Sketchy [SBHH16] et Quickdraw. A noter que tous ces datasets sont constitués d'objets "isolés" et non composés comme peut l'être une peinture chinoise.

### 1.4.2 Peinture chinoises

"La peinture chinoise désigne toute forme de peinture originaire de Chine ou pratiquée en Chine ou par des artistes chinois hors de Chine." [Wik20a] Il existe beaucoup de styles, courants et genres différentes (paysage, peinture narratives, faunes, flore, personnage). Nous prévoyons d'exploiter un dataset de plus de 5000 peintures chinoises au format 256x256, récupérées depuis [ych18]

# Chapitre 2

## Propositions de solutions

Malgré nos recherches, nous n'avons à ce jour trouvé aucun modèle nous permettant de réaliser totalement notre projet. Ci-dessous une liste non définitive des nouvelles pistes que nous souhaitons explorer.

### 2.1 Création du dataset en semi automatique

En utilisant le logiciel LabelBox, il est possible de labelliser manuellement une image à travers l'utilisation des sémantiques layout. A titre d'information, labelliser 1000 images prendrait environ 2 jours hommes à raison d'une minute par image. Une fois cette étape réalisée, nous pourrions générer une ou plusieurs abstractions pour chaque peintures, en utilisant le dataset "quick draw" entre autres. Outre l'inconvénient du temps passé, cette solution n'est pas généralisable.

### 2.2 Photo-Sketching : Inferring Contour Drawings from Images

Todo : Fakhri

### 2.3 Abstraire une image via les CNN

Nous savons que les RCNNs, grâce aux couches de convolutions, sont capables de détecter les caractéristiques visuelles communes d'un ensemble d'images. Comme le prouve ce travail : <https://github.com/dribnet/perceptionengines>, il est peut-être possible de générer une représentation abstraite de nos peintures en utilisant ce principe. Il serait alors théoriquement possible d'obtenir

des paires abstractions/peintures, permettant l'entraînement d'un cGAN type "traduction image-image".

## Chapitre 3

### Conclusion et Perspectives

# Bibliographie

- [Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6) :679–698, 1986.
- [EHA12] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4) :44 :1–44 :10, 2012.
- [GEB15] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.
- [hin18] hindupuravinash. The gan zoo. <https://github.com/hindupuravinash/the-gan-zoo/>, 2018.
- [IZZE16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.
- [LLM<sup>+</sup>19] Mengtian Li, Zhe Lin, Radomír Mech, Ersin Yumer, and Deva Ramanan. Photo-sketching : Inferring contour drawings from images. *CoRR*, abs/1901.00542, 2019.
- [PLWZ19] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. *CoRR*, abs/1903.07291, 2019.
- [SBHH16] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database : Learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, 2016.
- [SPS<sup>+</sup>18] Jifei Song, Kaiyue Pang, Yi-Zhe Song, Tao Xiang, and Timothy M. Hospedales. Learning to sketch with shortcut cycle consistency. *CoRR*, abs/1805.00247, 2018.
- [Wik20a] Wikipedia. Peinture chinoise — Wikipedia, the free encyclopedia. <http://fr.wikipedia.org/w/index.php?title=Peinture\%20chinoise&oldid=167921234>, 2020. [Online; accessed 17-May-2020].

- [Wik20b] Wikipedia. Réseaux antagonistes génératifs — Wikipedia, the free encyclopedia. <http://fr.wikipedia.org/w/index.php?title=R%C3%A9seaux%20antagonistes%20g%C3%A9n%C3%A9ratifs&oldid=170457375>, 2020. [Online; accessed 17-May-2020].
- [ych18] ychen93. The gan zoo. <https://github.com/ychen93/Chinese-Painting-Dataset>, 2018.