

Rapport
HAI508I : Projet long intégrateur CMI Informatique 1
”Amusons-nous avec des images”

Etudiants :

Morgan Navel
Romain Gallerne
Baptiste Bronsin
Lisa Savy
Eric Gilles

Encadrant :

Pascal Poncelet

Janvier 2023

Table des matières

1	Introduction	3
2	Problématique	4
3	Présentation du Projet	4
3.1	Gestion du Projet	4
3.2	Planification	4
4	Travail réalisé	5
4.1	Les bases de l'apprentissage profond	5
4.2	La descente de Gradient	6
4.3	Les risques de l'apprentissage profond	7
5	Conclusion	8
6	Bibliographie	8

1 Introduction

L'IA est un domaine qui vise à créer des systèmes informatiques capables de réaliser des tâches qui nécessitent normalement une intelligence humaine, comme la reconnaissance vocale, la traduction automatique, la conduite autonome et bien d'autres encore. Il existe plusieurs catégories d'intelligences artificielles mais celle qui va ici nous intéresser est l'apprentissage automatique, de son nom anglais "machine learning", qui utilise des algorithmes pour apprendre à partir de données.

L'apprentissage profond, ou plus connu sous son nom anglais "deep learning", est une sous-catégorie d'apprentissage qui utilise des réseaux de neurones pour résoudre des problèmes complexes.

Dans ce rapport, nous allons expliquer brièvement ce qu'est l'apprentissage profond ainsi qu'ouvrir la voie sur le travail que nous réaliserons lors du prochain semestre où nous plongerons plus en profondeur dans la mise en place de cette technologie.

2 Problématique

Le sujet de ce stage est la reconnaissance d'images de tigres à l'aide de l'intelligence artificielle et plus précisément grâce à l'apprentissage profond. L'objectif principal est de développer un système capable de reconnaître des tigres sur des images, et éventuellement d'autres animaux ou objets. La deuxième étape consistera à utiliser les résultats de cette reconnaissance pour générer ou recolorer des images.

Pour cela, il nous faut tout d'abord acquérir une bonne compréhension théorique de ces technologies, c'est ce que nous nous sommes fixés comme objectif durant cette première partie du TER. Il nous faudra par la suite utiliser notre apprentissage afin de mettre en place par nous-même une telle intelligence artificiel puis d'organiser son entraînement afin de la rendre la plus performante possible.

3 Présentation du Projet

3.1 Gestion du Projet

Ce projet comporte cinq membres dont trois étudiants CMI : Morgan Navel, Baptiste Bronsin et Romain Galerne ainsi que deux étudiants de licence classique : Lisa Savy et Eric Gilles. Celui-ci est encadré par M. Pascal Poncelet, maître de conférence à l'université de Montpellier.

Les recherche réalisées consistent principalement en des documents et extraits de code Python d'intelligences artificielles à étudier fournis par notre encadrant ainsi qu'une présentation, réalisée elle aussi par notre encadrant afin de nous aider à débroussailler le sujet.

3.2 Planification

Pour atteindre nos objectifs, nous avons planifié de réaliser les étapes suivantes :

- Mettre en place une base de données d'images de tigres et éventuellement d'autres animaux ou objets.
- Utiliser des techniques de classifications automatiques pour apprendre à reconnaître les tigres dans les images.
- Entraîner un modèle de reconnaissance d'images sur ces données.
- Utiliser les résultats de la reconnaissance pour générer ou recolorer des images.

4 Travail réalisé

Au cours de ce semestre, nous avons appris les bases de l'IA, ainsi que les différentes notions liées à l'apprentissage automatique et profond. Nous avons également étudié comment fonctionne une IA, en nous concentrant sur les aspects mathématiques de base.

4.1 Les bases de l'apprentissage profond

Comme dit précédemment, l'apprentissage profond est une sous-catégorie d'apprentissage qui utilise des réseaux de neurones.

Les réseaux de neurones profonds utilisés en deep learning sont souvent appelés ainsi en raison de la profondeur de leur architecture, c'est-à-dire le nombre de couches de neurones qu'ils contiennent. Un réseau de neurones peut avoir des dizaines, voire des centaines de couches, chacune contenant des milliers de neurones.

Ces réseaux sont des modèles dit "inspirés du cerveau humain" car ceci sont composés de nombreux nœuds (ici appelés neurones) connectés entre eux. Chacun de ces nœuds représente une unité de calcul qui effectue une opération mathématique sur les entrées qu'il reçoit, avant de transmettre le résultat à d'autres neurones dans le réseau.

Ainsi, lorsque l'on qualifie ces unités de neurones, cela pourrait être considéré comme un abus de langage. En effet, ce ne sont en réalité que des fonctions mathématiques effectuant une opération bien précise, le parallèle fait avec le cerveau humain est plutôt à faire sur la méthode d'apprentissage.

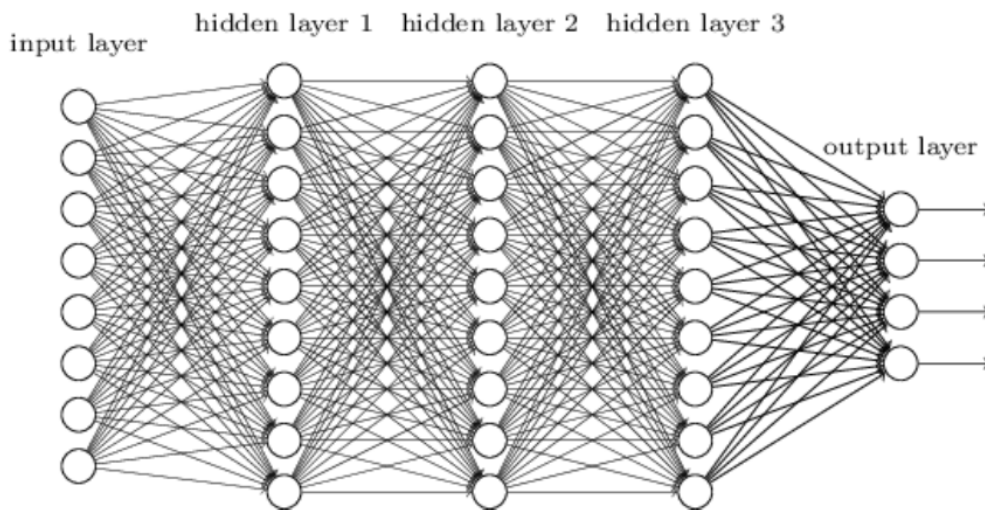


FIGURE 1 – Représentation graphique d'un réseau de neurones

En effet, les réseaux de neurones sont entraînés en utilisant un grand nombre de données. Ces données peuvent être de différents types selon la reconnaissance voulue (des images, du texte, des sons...). On n'oubliera pas de fournir dans le jeu de données la catégorie à laquelle appartient chaque donnée.

Par exemple, des images pourraient être classées selon :

- Chien
- Chat
- Requin
- Ours polaire
- Tigre
- ...

L'algorithme essaiera ensuite de faire des regroupements d'images d'après ces catégories, de leur trouver des points communs.

Plus formellement, pour chaque image l'algorithme va traiter l'image au travers de son réseau de neurones, lui appliquant ainsi différentes fonctions afin d'obtenir une sortie. Dans le cas qui nous intéresse, on pourrait imaginer une sortie binaire :

- 1 : C'est un tigre
- 0 : Ce n'est pas un tigre

Dans le cas qui nous intéresse, c'est-à-dire une image, l'algorithme ne verra pas l'image de la même façon qu'un oeil humain. En réalité, celle-ci va être décomposée sous la forme d'une matrice (voir Figure 2). Le traitement appliqué est donc purement mathématique.

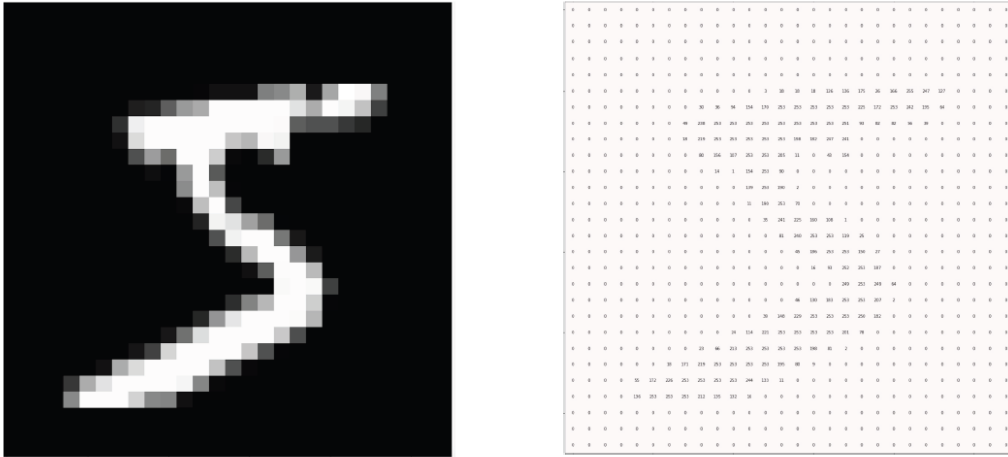


FIGURE 2 – Exemple d'une image après sa transformation en matrice

Au passage de chaque image, l'algorithme va ajuster le poids de ces noeuds, c'est-à-dire l'impact de chacune des fonctions mathématiques afin de forcer le réseau à rendre la valeur souhaitée. Passage après passage, le réseau va naturellement s'ajuster afin de reconnaître les similitudes entre les images et ainsi se perfectionner.

4.2 La descente de Gradient

Afin que l'algorithme soit capable de s'améliorer tout seul au passage chaque image, celui-ci doit pouvoir se corriger. Cette correction peut être plus ou moins forte. La "force" de cette correction est appelée le pas, ou en anglais "Learning Rate". Ce pas définit à quel point l'algorithme va se corriger fortement lorsque il doit s'adapter à chaque image. Par corriger on entend modifier le poids des différents neurones. L'objectif ici est d'atteindre le taux d'erreurs, appelé erreur quadratique moyenne, le plus bas. Cette erreur quadratique moyenne peut être représentée comme une fonction quadratique (voir Figure 3). Le but ici est de se rapprocher du minimum de la fonction, plus on en sera proche, plus l'algorithme sera performant (comprendre ici, aura un meilleur taux de bonne réponse). Le pas définit donc de combien on avance sur l'axe x à chaque correction.

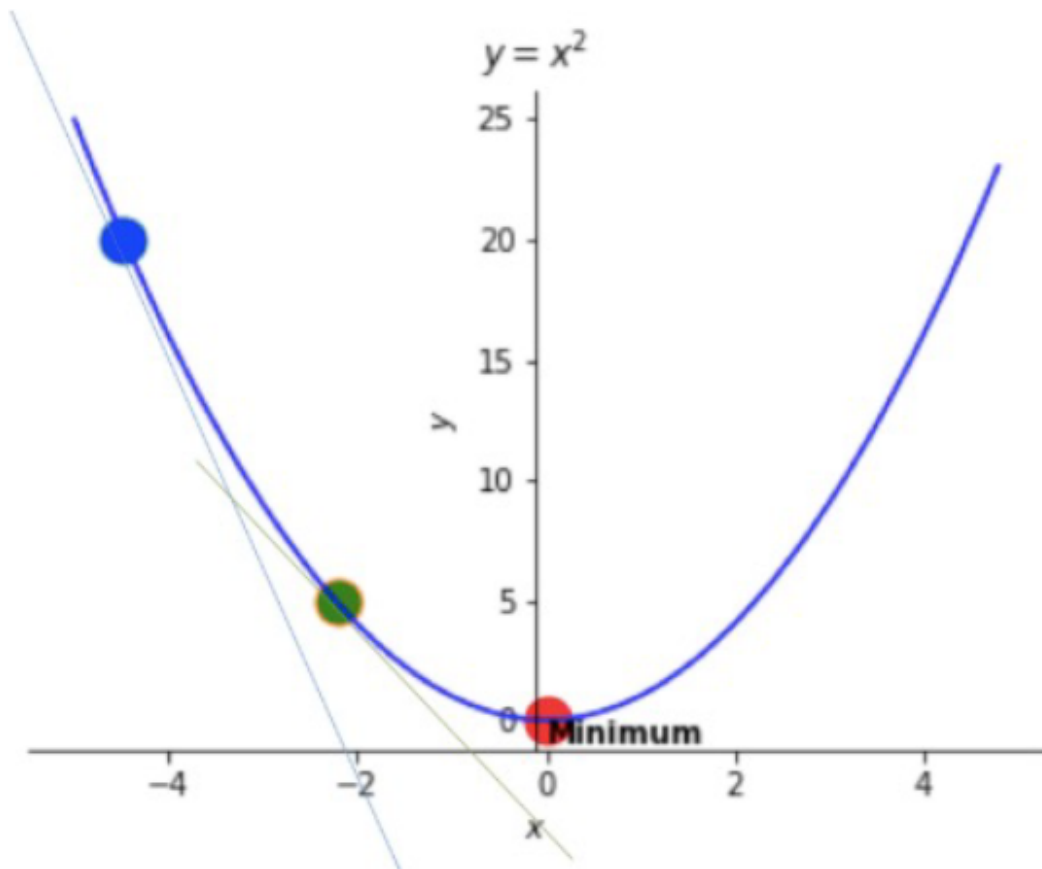


FIGURE 3 – Graphique de la Descente de Gradient, on voit ici indiqué le pas qui correspond à la correction appliqué.

Pour simplifier on a ici choisi de ne montrer que la version quadratique "simple" de la descente de gradient. Pour plus de précisions on pourrait plutôt, dans certain cas, la comparer à un fonction en trois dimensions.

On comprend ainsi bien qu'un pas très grand permet de se rapprocher rapidement du minimum, en revanche, il sera difficile de l'atteindre précisément (car on risque de le dépasser vite).

A l'inverse, un très petit pas permettra de trouver une valeur de x très proche du minimum de la fonction, cependant le temps d'apprentissage en sera décuplé car les corrections seront très minimes (donc très précise).

4.3 Les risques de l'apprentissage profond

Attention cependant, plusieurs choses sont bien à prendre en compte :

Premièrement, un tel apprentissage bien que pouvant atteindre de très haut niveau de fiabilité (+97%), celui-ci ne pourra jamais être infaillible car il ne peut pas être entraîné sur toutes les images de tigres possibles. Il est donc fort probable qu'il existe quelques failles permettant de tromper la reconnaissance alors que l'oeil humain aurait reconnu l'objet.

De plus, même si l'on possédait toutes les images de tigres existantes cela ne serait pas bénéfique car il faut faire attention au risque dit de "sur-apprentissage". Si on laisse trop tourné l'algorithme en mode apprentissage, celui-ci risque de n'apprendre qu'à reconnaître par coeur le jeu de données de test. Il faut donc faire attention à garder une marge de manoeuvre.

5 Conclusion

Au cours de ce premier semestre, nous avons appris de nombreuses choses sur l'IA et sur les techniques utilisées pour la reconnaissance d'images. Nous nous sommes familiarisés avec les concepts de base de l'intelligence artificielle et de l'apprentissage profond qu'il nous faudra explorer plus en profondeur et implémenter lors du prochain semestre. Nous aimerions maintenant continuer à explorer ces sujets pour développer notre propre système de reconnaissance d'images et atteindre nos objectifs initiaux.

6 Bibliographie

- "Les réseaux de neurones", "Descente de Gradient", "Les réseaux de neurones convolutifs (CNN)", "Sous apprentissage – Sur apprentissage" par Pascal Poncelet, Transparence de conférence dispensés à l'Université de Montpellier.
- "Deep Learning for Computer Vision : A Brief Review" par Kuang-Huei Lee, publié dans Computer Science Review en 2018, qui donne un aperçu des techniques de deep learning utilisées pour la reconnaissance d'images.
- "A Review on Deep Learning Techniques Applied to Object Detection" par Shivanand K. Shettar, publié dans Pattern Recognition Letters en 2018, qui examine les différentes techniques de deep learning utilisées pour la détection d'objets dans des images.
- "Deep Learning for Object Recognition : A Review" par Ahmed Elgammal, publié dans ACM Computing Surveys en 2017, qui examine les différentes techniques de deep learning utilisées pour la reconnaissance d'objets dans des images.
- "Deep Learning for Animal Detection and Classification" par Ali Borji, Ross Girshick, publié dans IEEE Transactions on Pattern Analysis and Machine Intelligence en 2018, qui examine comment les techniques de deep learning ont été utilisées pour la détection et la classification d'animaux dans des images.