*ENSEEIHT*

*Rapport du projet de classification des images des pièces d'ordinateur*

Hamid Oukhnini

Paula Sánchez

Houssem Radhouane

Pierre-Louis de Villers

Table des matières

[Base de données 0](#_Toc99397643)

[Description du sujet choisi 0](#_Toc99397644)

[Description de la méthodologie pour acquérir, puis annoter les données. 0](#_Toc99397645)

[Méthodologie pour partitionner les images 0](#_Toc99397646)

[Votre pronostic : votre problème est-il simple ou compliqué ? Quels résultats obtiendrez-vous d’après vous ? 1](#_Toc99397647)

[Script de chargement de vos données. 1](#_Toc99397648)

[Et bien sûr des images de votre BD ! 1](#_Toc99397649)

# **Base de données**

Lien vers la base de données  
<https://github.com/houssemradhouane/CNN-Algorithm-for-Image-Classification/tree/main/Dataset>

# **Description du sujet choisi**

Sur les sites de vente de pièces d'ordinateur, la classification est toujours un problème lorsqu'on vend des pièces d'ordinateur. Il arrive souvent que le vendeur classe la pièce qu'il veut vendre dans la mauvaise case. Par exemple, il peut vouloir vendre un refroidisseur d'unité centrale mais le mettre dans la catégorie des claviers. Dans notre projet, la classification des pièces d'ordinateur sera automatique en faisant appel à l'apprentissage profond.

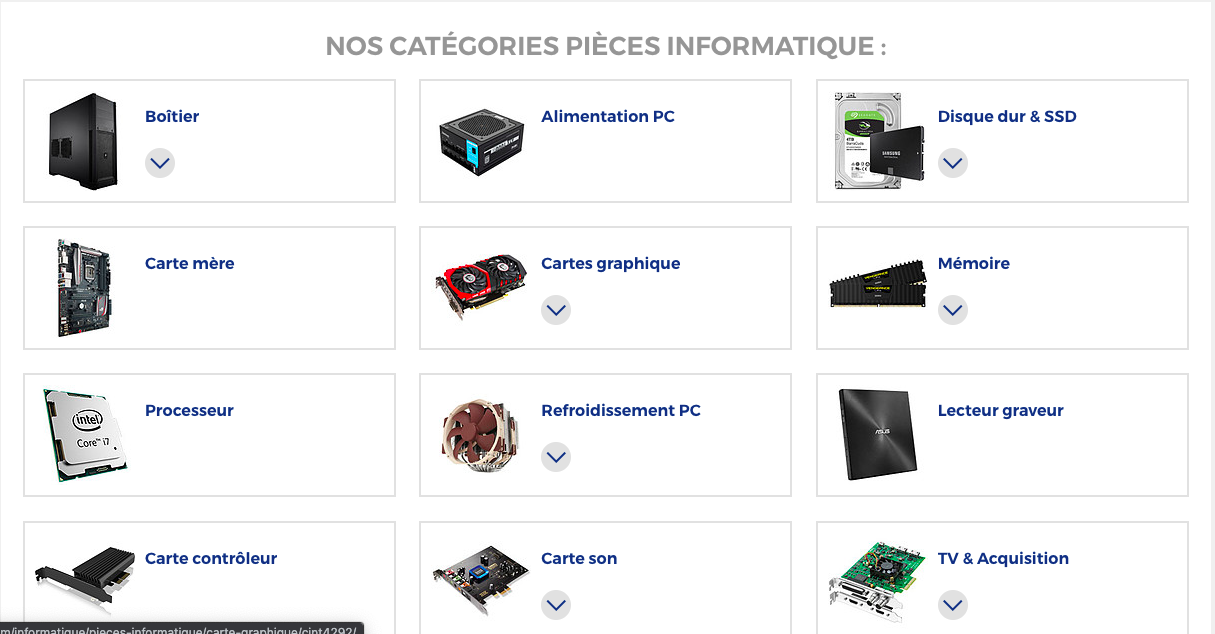


Figure Exemple de classification sur le site LDLC

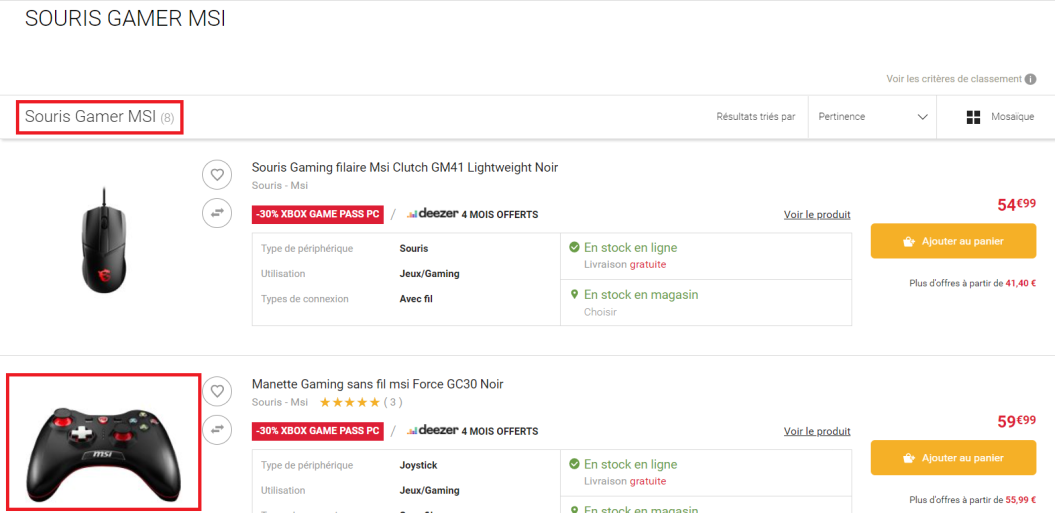


Figure Exemple de classification sur le site fnac

# **Description de la méthodologie pour acquérir, puis annoter les données**

On a d'abord commencé par chercher des sites de vente riches en images de pièces d'ordinateurs, ceci nous évite la procédure d'élimination des doublons puisque ces sites offrent rarement le même produit deux fois. Le site qu'on a choisi est pcpartpicker.com. On a ensuite utilisé une extension de chrome qui s'appelle 'Fatkun Batch Download', elle permet de télécharger des images en batch et nous facilite la collecte des images. Ensuite, on a utilisé deux script python pour renommer ces images et les ramener à une même résolution ( 256 x 256 ), nous pensons que cette résolution offre un bon compromis entre qualité d'apprentissage et volume de ressources nécessaire pour l'apprentissage.

# **Méthodologie pour partitionner les images**

Les données acquises (images des parties de l'ordinateur) sont divisées pour la formation, les tests et la validation. Il existe six parties d'un ordinateur : le refroidisseur de l'unité centrale, le boîtier, les écouteurs, le clavier, la souris et le séparateur. Pour chaque partie, nous avons la même quantité de données (images). Ces données doivent être divisées en : formation, test et validation.

Comme il est important de disposer d'une grande quantité de données pour l'entraînement, 70% d'entre elles seront utilisées pour cette première partie. Pour le test et la validation, respectivement 15% des données seront utilisées.

# **Votre pronostic**

Dans un premier temps, nous considérons que le problème n'est pas très difficile car les pièces que nous essayons de classer sont très différentes les unes des autres, comme les claviers et les souris. Un problème que nous avons rencontré est que les images utilisées pour l'entraînement ont un fond blanc, de sorte que l'algorithme peut ne pas identifier la partie de l'ordinateur lorsque le fond n'est pas blanc.

# **Script de chargement de vos données.**