

Chapitre 3

Outils indispensables à la pratique

1. Objectifs de ce chapitre

Dans ce chapitre, nous allons lister et décrire les outils indispensables à la pratique de l'apprentissage profond pour le traitement d'images.

Nous commencerons par discuter du matériel (ordinateurs et cartes graphiques) le plus approprié selon les différents cas d'usage.

Nous parlerons ensuite des jeux de données, dont nous avons vu dans le chapitre précédent à quels points ils sont indispensables au développement d'un modèle performant. Nous présenterons certains jeux, très connus et couramment utilisés pour présenter les performances d'un modèle. Ensuite, des pistes pour trouver d'autres jeux de données plus spécifiques seront données.

Les langages et outils informatiques les plus utiles seront ensuite décrits, puis nous expliquerons comment les utiliser pour ouvrir les exemples de ce livre.

2. Ordinateurs et cartes graphiques

L'apprentissage profond requiert la réalisation de calculs complexes sur une grande quantité de données. Ainsi, il est préférable de disposer d'un ordinateur adapté.

Nous allons détailler les différents éléments à prendre en compte.

2.1 GPU et CPU

Classiquement, un ordinateur dispose d'un ou plusieurs processeurs (nommés CPU pour *Central Processing Unit*, ou unité centrale de traitement). Ces processeurs effectuent tous les calculs nécessaires à l'exécution des divers programmes.

Les ordinateurs disposent également d'une carte graphique, qui peut être intégrée au CPU, ou séparée.

Ces cartes, nommées GPU pour *Graphical Processing Units* (unités de traitement graphique en français), et initialement développées pour le rendu graphique, contiennent des milliers d'unités de calcul, quand un CPU n'en contient qu'une dizaine.

Ces dernières années, ces cartes graphiques ont été optimisées pour réaliser des calculs extrêmement rapidement. L'entraînement d'un réseau de neurones pour le traitement d'images étant extrêmement coûteux en temps de calcul, ces opérations sont généralement effectuées sur des cartes graphiques.

Ces cartes permettent de réaliser plus rapidement toutes les étapes d'entraînement et d'exploitation d'un réseau de neurones ; le facteur d'accélération dépend fortement de facteurs tels que la taille du réseau ou la taille du jeu de données. Dans certains cas, il est possible d'accélérer plus de cent fois les calculs en utilisant une carte GPU plutôt qu'une carte CPU.

L'utilisation de ces cartes a toutefois deux inconvénients.

Le premier inconvénient est que ces cartes sont plus coûteuses qu'un processeur. Leur prix peut atteindre plusieurs centaines, voire plusieurs milliers d'euros pour les plus performantes.

Le second inconvénient est que, du fait de leur puissance, ces cartes induisent une hausse de la consommation d'énergie de l'ordinateur.

■ Remarque

Cette hausse d'énergie est significative, ce qui est problématique dans le contexte de la crise climatique. Ainsi, nous assistons ces dernières années au développement de réseaux de plus en plus efficaces, nécessitant moins de ressources pour l'entraînement et l'application. Pour le moment, il reste néanmoins recommandé de s'équiper d'une carte graphique pour faire de l'apprentissage profond.

Il existe plusieurs constructeurs de cartes GPU, mais il est préférable de choisir une carte NVIDIA, qui est compatible avec les principaux frameworks d'apprentissage profond tels que TensorFlow et PyTorch. Ces frameworks sont des ensembles d'outils, de bibliothèques et d'interfaces ayant pour but de faciliter la création, l'entraînement et le déploiement de modèles d'apprentissage profond. Ils seront décrits plus en détail dans la section Langages, librairies et frameworks de ce chapitre.

Les deux critères les plus importants pour le choix d'une carte graphique sont la puissance et la taille de la mémoire.

La puissance dépend à la fois de la fréquence et du nombre d'unités de calcul. La fréquence, exprimée en MHz ou GHz, indique la vitesse à laquelle le GPU effectue ses calculs. Le nombre d'unités de calcul, soit le nombre de cœurs de calculs, indiquent combien d'unités sont dédiées au calcul numérique. La comparaison de deux cartes est donc compliquée, et le plus fiable est de se référer aux nombreux benchmarks disponibles sur les sites spécialisés. Parmi ces benchmarks, on peut citer FurMark, un protocole de test qui permet de détecter les problèmes de surchauffe et de consommation d'énergie, ainsi que GeekBench, qui permet de mesurer les performances globales de calcul.

Les cartes graphiques sont souvent limitées en taille mémoire, ce qui peut ralentir les calculs. Généralement, 8 à 16 Go de mémoire sont recommandés pour traiter des images bidimensionnelles. Pour le traitement d'images tridimensionnelles, qui sont beaucoup plus lourdes, 16 Go sont nécessaires au minimum.

Pour une puissance de calcul accrue, il est possible d'installer plusieurs cartes graphiques dans un même ordinateur. Notez que l'accumulation ne compense pas la faible puissance d'une carte : il sera toujours plus efficace d'avoir une unique carte graphique de bonne qualité que plusieurs cartes médiocres.

Ces configurations à plusieurs cartes, extrêmement coûteuses, utilisent généralement les cartes graphiques les plus puissantes. Il est alors nécessaire d'adapter le système de refroidissement de l'ordinateur, ce qui augmente encore le coût de la configuration finale. Le coût total d'une telle configuration démarre autour de 5 000 euros, et peut atteindre plusieurs dizaines de milliers d'euros.

Quant au processeur, il n'est utilisé que pour effectuer des prétraitements sur les images, et pour envoyer des données à la carte graphique. Deux éléments doivent guider votre choix : le nombre de cœurs, qui devrait être au moins égal à cinq, et le nombre de lignes PCIe, qui permettent de connecter le processeur à la carte graphique. Si vous comptez installer plusieurs cartes GPU dans votre ordinateur, il faut donc adapter le processeur en conséquence.

2.2 Taille de la mémoire vive

La mémoire vive (appelée aussi RAM pour *Random Access Memory*) est un autre facteur important pour le choix d'un ordinateur dédié à l'apprentissage profond. Plus la mémoire vive est grande, plus elle peut gérer une quantité élevée de données, ce qui accélère le traitement.

Il est généralement conseillé d'avoir une taille de RAM au moins équivalente à la plus grande taille mémoire des cartes graphiques présentes dans votre ordinateur.

2.3 Investir dans un ordinateur performant ou réaliser tous les calculs sur des plateformes déportées ?

Il existe plusieurs services proposant des plateformes de calcul sur des plateformes déportées. Le paiement se fait alors en fonction du temps de calcul, et de la puissance des unités de calculs mises à disposition. Les plateformes de calcul les plus couramment utilisées sont AWS (*Amazon Web Services*), Microsoft Azure, et GCP (*Google Cloud Platform*). Les tarifs varient de un euro de l'heure pour la mise à disposition des serveurs les moins puissants à plusieurs dizaines d'euros de l'heure pour les serveurs plus puissants.

L'avantage principal de ces plateformes est leur flexibilité : il est très facile de modifier la configuration selon ses besoins. Ceci permet d'augmenter la puissance de calcul presque en temps réel, mais aussi de la diminuer si vos besoins diminuent.

A contrario, avec un ordinateur local, augmenter votre puissance de calcul ne pourra se faire instantanément, et il vous sera impossible de diminuer cette puissance.

La solution déportée a toutefois deux inconvénients. Le premier inconvénient est qu'il peut être difficile de prédire les coûts, contrairement à un ordinateur local, pour lequel les seuls coûts sont le coût initial, facilement maîtrisé, et le coût de l'énergie.

Le second inconvénient est lié aux données : réaliser les calculs sur un serveur déporté nécessite d'y envoyer les données, ce qui peut être délicat pour des données sensibles comme des données médicales ou d'autres données confidentielles.

Pour résumer, si vous travaillez avec des données sensibles (comme des données médicales), et que vous avez une idée précise de vos besoins en termes de puissance de calcul, choisissez un ordinateur local.

Si vous travaillez avec des données non sensibles, que vous n'avez pas une idée précise de vos besoins et que vous êtes ouvert à une certaine incertitude au sujet du budget total, effectuez les calculs sur une plateforme déportée. Il vous sera ensuite possible de vous équiper localement une fois que vous saurez plus précisément de quelle puissance de calcul vous avez besoin.

Enfin, il est également possible de constituer une petite plateforme de calcul à l'échelle de l'entreprise. Ainsi, certaines entreprises investissent dans un ou plusieurs ordinateurs très puissants, auxquels les membres du personnel peuvent se connecter selon leurs besoins.

3. Jeux de données ouverts

Les modèles d'apprentissage profond sont très gourmands en données. Heureusement, il existe aujourd'hui une grande quantité de jeux de données facilement disponibles.

3.1 Fondamentaux d'un jeu de données d'entraînement

Pour être le plus performant possible, un modèle d'apprentissage profond doit être entraîné sur un ensemble dont les images remplissent les conditions suivantes :

- Images très nombreuses : on estime généralement qu'un réseau de classification nécessite au moins mille images par classe pour l'entraînement. Le chiffre exact dépend évidemment du problème et peut varier largement. Il est conseillé de démarrer la recherche de données par une recherche bibliographique. Si vous trouvez d'autres personnes qui ont résolu un problème similaire au vôtre, vous pourrez voir quelles performances ont été obtenues avec différentes quantités de données.
- Images accompagnées d'annotations fiables : les annotations des images sont les informations que le modèle va chercher à prédire. Par exemple, pour un jeu de données utilisé en classification, l'annotation de chaque image est un mot désignant sa classe. Avoir un faible nombre d'annotations erronées reste acceptable, mais un plus grand taux d'erreur pénalisera fortement le modèle.
- Images en nombre à peu près équivalent pour chaque classe : c'est ce qu'on appelle des classes équilibrées.
- Images suffisamment différentes les unes des autres à l'intérieur de chaque classe pour illustrer la diversité possible à l'intérieur d'une classe.

De plus, il est important de prendre en compte les biais éventuels du modèle dès la constitution du jeu de données qui sera utilisé pour l'entraînement.

Par exemple, les images trouvées sur Internet contiennent souvent des différences liées au genre, à l'âge ou à la couleur de peau. Si ces biais ne sont pas corrigés dans les données d'entraînement, ils se répercuteront dans le modèle. Pour minimiser ces biais potentiels, il est donc important d'avoir une représentation équilibrée des différents groupes dans les données d'entraînement.

Pour cela, il est nécessaire de réfléchir en amont aux biais potentiels dont le modèle pourrait souffrir, afin de s'assurer que les données d'entraînement sont non biaisées.

Nous parlerons plus longuement des biais dans le chapitre Bonnes pratiques pour l'entraînement.

3.2 Avantages d'utiliser un jeu de données ouvert

Il existe deux avantages principaux à utiliser un jeu de données ouvert plutôt que d'essayer de constituer son propre jeu.

Le premier avantage est qu'il est extrêmement laborieux de constituer un ensemble de données de bonne qualité, et satisfaisant aux conditions énumérées plus haut.

Certains ensembles de données ouverts sont de mauvaise qualité et ne remplissent pas tous ces critères, mais la plupart les satisfont.

Le second avantage des ensembles de données ouverts est qu'ils permettent de comparer les performances de différents réseaux développés par des personnes différentes sur les mêmes données.

Comme nous l'avons vu au chapitre précédent, c'est l'existence même de ces ensembles de données ouverts qui a permis la démocratisation de l'apprentissage profond.

La plupart des données disponibles concernent la classification. Historiquement, c'est la première application (parce que c'est la plus simple) pour laquelle des algorithmes d'apprentissage profond ont été développés. Mais de plus en plus de jeux de données se développent pour des applications plus complexes comme la détection, la segmentation, ou des tâches encore plus spécifiques comme l'estimation de pose. Rappelons que la détection, décrite dans le chapitre Intelligence artificielle et apprentissage profond, a pour but de détecter quels objets d'intérêt sont présents sur une image, et de tracer une boîte englobante autour de chacun. La segmentation, décrite dans le même chapitre, a pour but de délimiter précisément le contour de ces objets. L'estimation de pose est une tâche plus récente qui vise à détecter la position et l'orientation d'un objet dans l'espace. Une des applications de cette tâche est l'analyse de mouvement, qui peut être utilisée dans des domaines tels que la rééducation fonctionnelle des personnes handicapées, l'amélioration des performances sportives, ou encore l'optimisation ergonomique des chaînes de montage.

3.3 Ensembles de référence

Certains ensembles de données sont très classiques, et les nouveaux réseaux sont presque systématiquement entraînés sur eux. Nous allons en présenter quelques-uns ici.

3.3.1 Iris

« Les iris de Fisher » sont un célèbre ensemble de données numériques utilisé depuis les années 1930 pour la classification. Il s'agit de reconnaître le type d'iris (*setosa*, *virginica*, et *versicolor*) à partir de la longueur et de la largeur de ses pétales et sépales. Le fichier contient cinquante fleurs de chaque type.