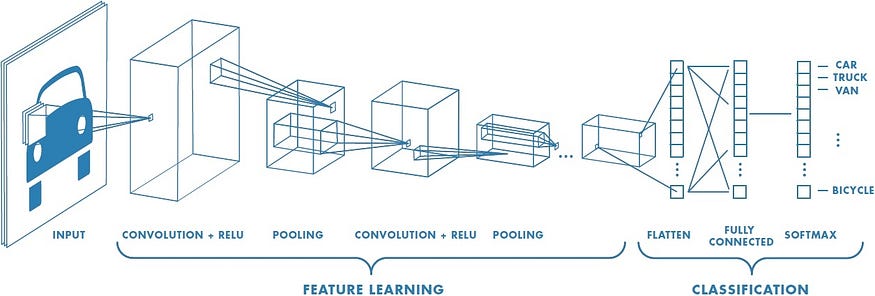
# TP7 : Classification d'Images avec CNN et MNIST - Ressources

Un **réseau neuronal convolutionnel (CNN)** est un type de réseau neuronal d'apprentissage profond qui convient parfaitement à l'analyse d'images et de vidéos. Les CNN utilisent une série de couches de convolution et de regroupement pour extraire des caractéristiques des images et des vidéos, puis utilisent ces caractéristiques pour classer ou détecter des objets ou des scènes.

## Architecture CNN:



* + **Convolution** : Opération qui permet d'extraire les caractéristiques locales d'une image (bords, textures, motifs).
  + **Pooling** : Technique de réduction de dimension qui consiste à prendre le maximum ou la moyenne de sous-régions d'une image.
  + **Flattening** : Conversion de la matrice de caractéristiques 2D en un vecteur 1D.
  + **Couches denses** : Couches entièrement connectées qui suivent la couche de flattening et permettent la classification finale.

## Fonctionnement du CNN

Les CNN fonctionnent en appliquant une série de couches de convolution et de regroupement à une image ou une vidéo d'entrée. Les couches de convolution extraient les caractéristiques de l'entrée en faisant glisser un petit filtre, ou noyau, sur l'image ou la vidéo et en calculant le produit scalaire entre le filtre et l'entrée. Les couches de regroupement sous-échantillonnent ensuite la sortie des couches de convolution pour réduire la dimensionnalité des données et les rendre plus efficaces en termes de calcul.

## Avantages des CNN :

* + Bon pour détecter des modèles et des caractéristiques dans les images, les vidéos et les signaux audio.
  + Formation de bout en bout, pas besoin d'extraction manuelle de fonctionnalités.
  + Peut gérer de grandes quantités de données et atteindre une grande précision.

## Inconvénients des CNN :

* + La formation est coûteuse en termes de calcul et nécessite beaucoup de mémoire.
  + Peut être sujet à un surajustement si les données ne sont pas suffisantes ou si une régularisation appropriée est utilisée.
  + Nécessite de grandes quantités de données étiquetées.

## Architecture CNN pour MNIST:

1. **Entrée (image)** :
   * Les images du jeu de données MNIST sont de taille 28x28 pixels. Chaque image est représentée comme une matrice de 28x28 avec une seule canal (niveaux de gris).
2. **Couches de Convolution** :
   * **Première couche de convolution** : Elle applique 32 filtres (ou noyaux) de taille 3x3 à l'image d'entrée. Cela produit 32 nouvelles images, chacune représentant une caractéristique détectée.
3. **Couches de Pooling (Sous-échantillonnage)** :
   * **Première couche de pooling** : Utilise le MaxPooling avec une taille de 2x2, ce qui réduit la dimension spatiale de l'image de moitié (14x14). Cela permet de diminuer la taille des images et de réduire le nombre de paramètres, tout en conservant les informations importantes.
4. **Couches de Convolution et Pooling supplémentaires** :
   * **Deuxième couche de convolution** : Applique 64 filtres de taille 3x3 aux sorties de la première couche de pooling.
5. **Deuxième couche de pooling** :
   * Réduit encore la dimension des images à 7x7.
6. **Couche de Flattening** :
   * Cette couche convertit les matrices 2D résultantes en un vecteur 1D. Par exemple, les images de 7x7x64 sont converties en un vecteur de 3136 éléments.
7. **Couche dense** : Comprend 64 neurones.
8. **Couche de sortie** : Contient 10 neurones avec une activation Softmax, correspondant aux 10 classes de chiffres (0-9). La fonction Softmax transforme les sorties en probabilités pour chaque classe.

## Q/R:

Q- Pourquoi choisir 32 neurones ?

**Suffisant pour capturer les caractéristiques de base** : Avec des images de petite taille 28x28, 32 filtres sont généralement suffisants pour détecter les caractéristiques principales des chiffres (comme les bords et les contours).

**Efficacité** : 32 est un nombre qui offre un bon équilibre entre la capacité de détection de caractéristiques et le coût computationnel.

**Pratiques courantes** : Il est courant dans la littérature de commencer avec 32 filtres et d'augmenter ce nombre dans les couches suivantes. Cela permet de capturer des caractéristiques de plus en plus complexes à mesure que l'on avance dans le réseau.

**Éviter le surajustement** : Utiliser un trop grand nombre de filtres dès le début pourrait entraîner un surajustement du modèle, surtout avec un jeu de données relativement simple comme MNIST.

**Q**- Pourquoi choisir 64 neurones ?

**Profondeur des caractéristiques :** La première couche de convolution (32 filtres) détecte des caractéristiques simples comme les bords et les textures de base, la deuxième couche (64 filtres) est conçue pour détecter des motifs plus complexes en combinant ces caractéristiques simples.

**Progression graduelle** : En augmentant le nombre de filtres dans les couches successives, le modèle peut capter des niveaux de complexité croissants dans les images. Une progression typique pourrait être 32 -> 64 -> 128 filtres dans les couches de convolution successives.

**Expérimentation et empirisme** : Les valeurs 32 et 64 sont souvent utilisées car elles ont montré de bons résultats empiriquement sur de nombreux jeux de données, y compris MNIST.

**Q- C’est quoi la différence entre accuracy et val\_accuracy ?**

**accuracy : C'est la précision du modèle sur les données d'entraînement. Elle mesure la proportion de prédictions correctes faites par le modèle lorsqu'il est appliqué aux mêmes données utilisées pour le former. C'est un bon indicateur pour voir comment le modèle s'ajuste aux données qu'il connaît.**

**val\_accuracy : C'est la précision du modèle sur les données de validation. Cet ensemble de données est distinct de l'ensemble d'entraînement et est utilisé pour évaluer la performance du modèle de manière plus objective, en vérifiant comment il se généralise à des données nouvelles qu'il n'a jamais vues auparavant. Cela aide à détecter des problèmes comme le surapprentissage (overfitting).**

**Q- Quelle est la différence entre une couche de convolution et une couche de pooling ?**

**Une couche de convolution (Conv2D) extrait les caractéristiques d'une image ou d'une vidéo en entrée, tandis qu'une couche de regroupement (MaxPooling2D) sous-échantillonne la sortie des couches de convolution. Les couches de convolution utilisent une série de filtres pour extraire les caractéristiques, tandis que les couches de regroupement utilisent diverses techniques pour sous-échantillonner les données, telles que le regroupement maximal et le regroupement moyen.**