

Réseaux de neurones convolutifs (CNN)

Convolutional Neural Networks (CNN)

April 2023

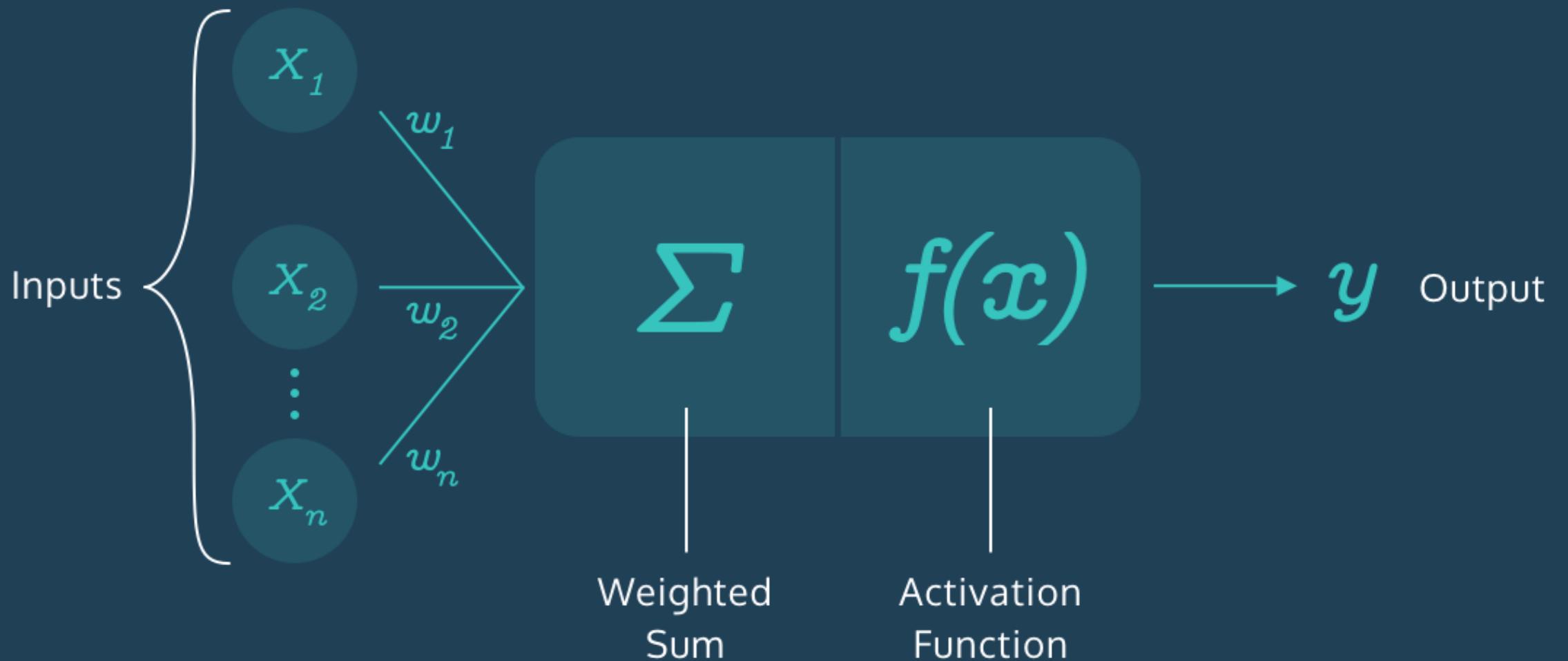
Sujets

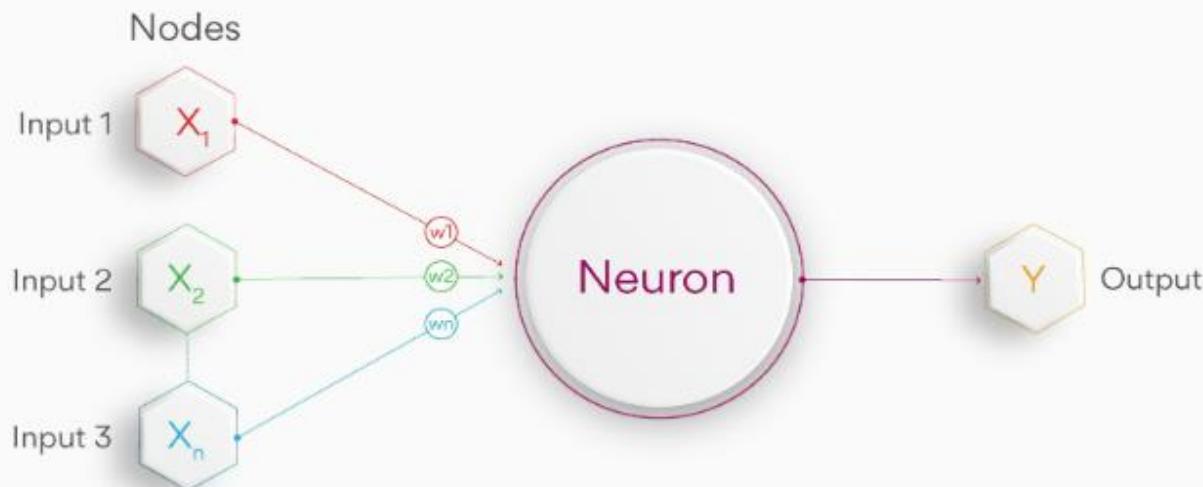
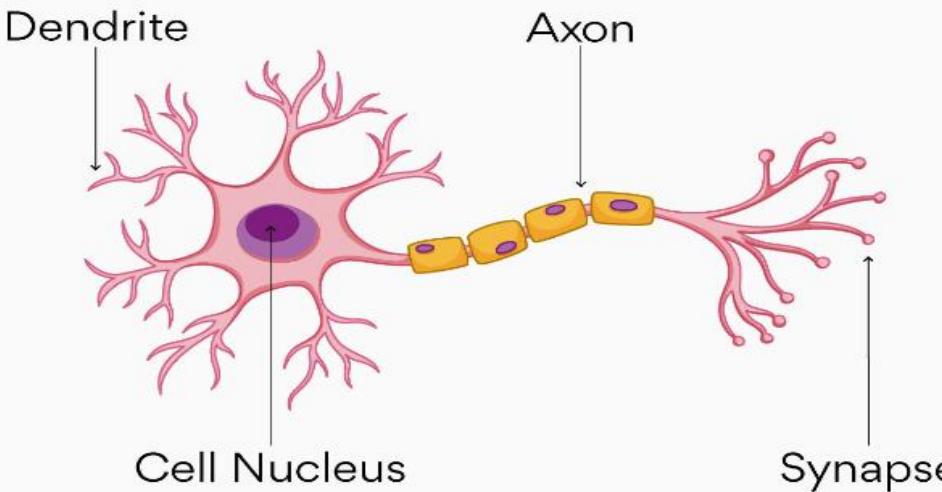
- I. Introduction aux CNN
 - A. Définition et applications des CNN
 - B. Vue d'ensemble de la structure d'un CNN
 - C. Avantages des CNN par rapport aux réseaux neuronaux traditionnels
 - D. Aperçu historique des CNN
- II. Principes fondamentaux des CNN
 - A. Opération de convolution
 - B. Fonctions d'activation
 - C. Mutualisation (pooling)
 - D. enjambée (Strides)
 - E. Rembourrage (padding)
- III. Création de CNN
 - A. Création d'une architecture CNN
 - B. Initialisation du poids
 - C. Propagation vers l'avant et vers l'arrière
 - D. Fonctions de perte
 - E. Optimiseurs
- IV. Mise au point des CNN
 - A. Apprentissage par transfert
 - B. Modèles pré-entraînés
 - C. Techniques de réglage fin
 - D. Régularisation
- V. Amélioration des performances CNN
 - A. Augmentation des données
 - B. Réglage des hyperparamètres
 - C. Normalisation des lots
 - D. Marginal (dropout)
- VI. Études de cas et applications
 - A. Détection d'objets
 - B. Segmentation d'image
 - C. Traitement du langage naturel
 - D. Analyse de séries chronologiques
- VII. Orientations futures des CNN
 - A. CNN profonds
 - B. CNN récurrents
 - C. Mécanismes d'attention
 - D. Modèles hybrides

Introduction aux CNN

Réseaux neuronaux - revue

- Les réseaux neuronaux modélisent la structure et la fonction du cerveau humain et sont constitués de couches interconnectées de neurones artificiels.
- Chaque neurone reçoit une ou plusieurs entrées, applique une fonction de transformation et produit une sortie qui est transmise à la couche suivante.
- Les poids des connexions entre les neurones déterminent leur force et leur direction, et sont ajustés pendant l'entraînement pour optimiser les performances.
- Les réseaux neuronaux peuvent gérer différents types de données et apprendre des modèles et des relations complexes dans les données sans ingénierie explicite des fonctionnalités.
- Il existe différents types de réseaux neuronaux, tels que feedforward, convolutif et récurrent, chacun avec ses propres forces et applications.
- Les réseaux neuronaux ont atteint des performances de pointe sur de nombreuses tâches difficiles et sont largement utilisés dans divers domaines.
- Les progrès récents dans les techniques de visualisation et d'explicabilité ont facilité la compréhension et l'interprétation des réseaux de neurones.
- Dans l'ensemble, les réseaux de neurones sont un outil puissant qui a le potentiel de révolutionner de nombreux domaines.

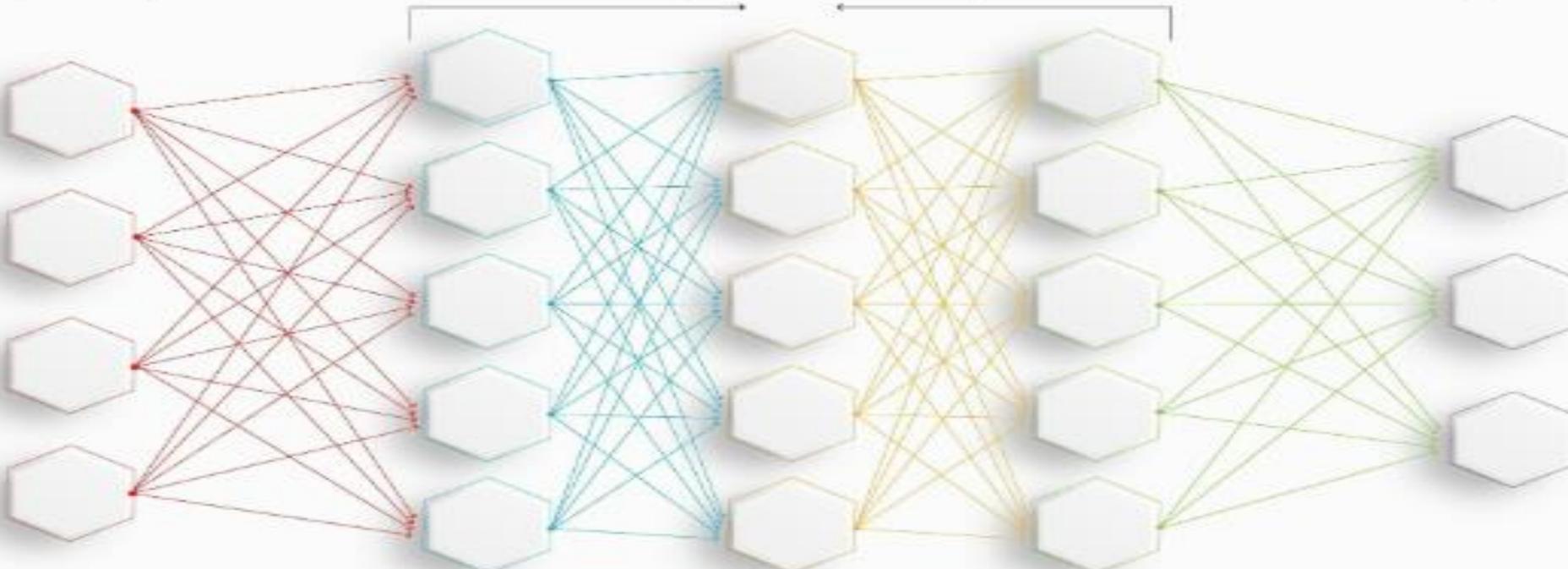




Input Layer

Multiple Hidden Layers

Output Layer

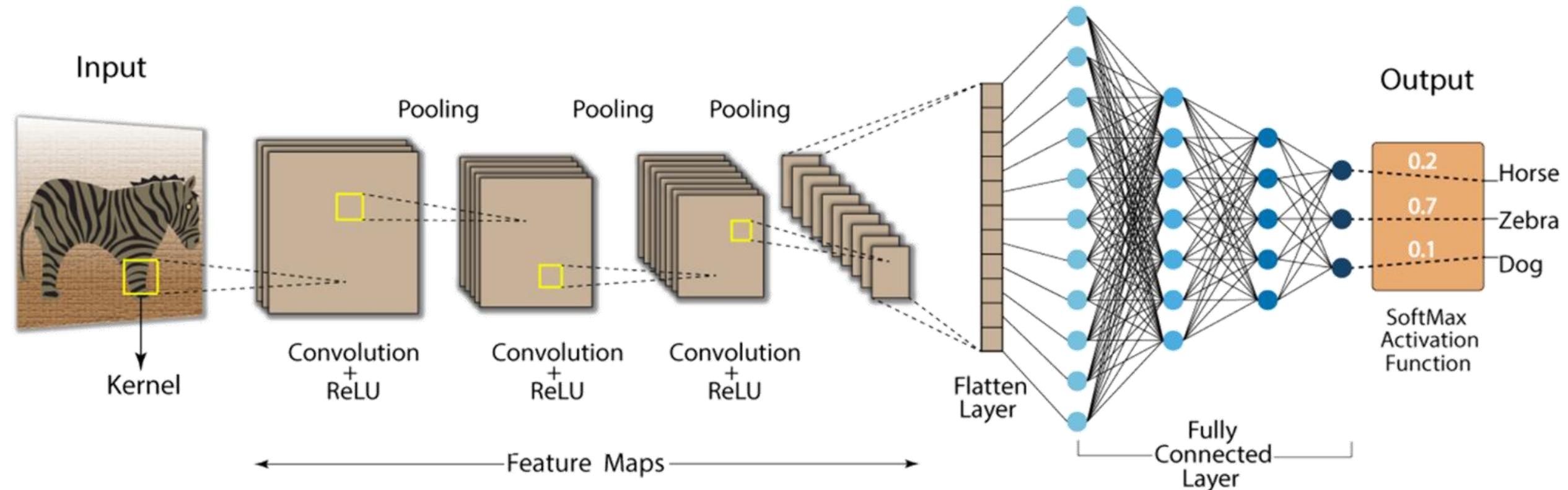


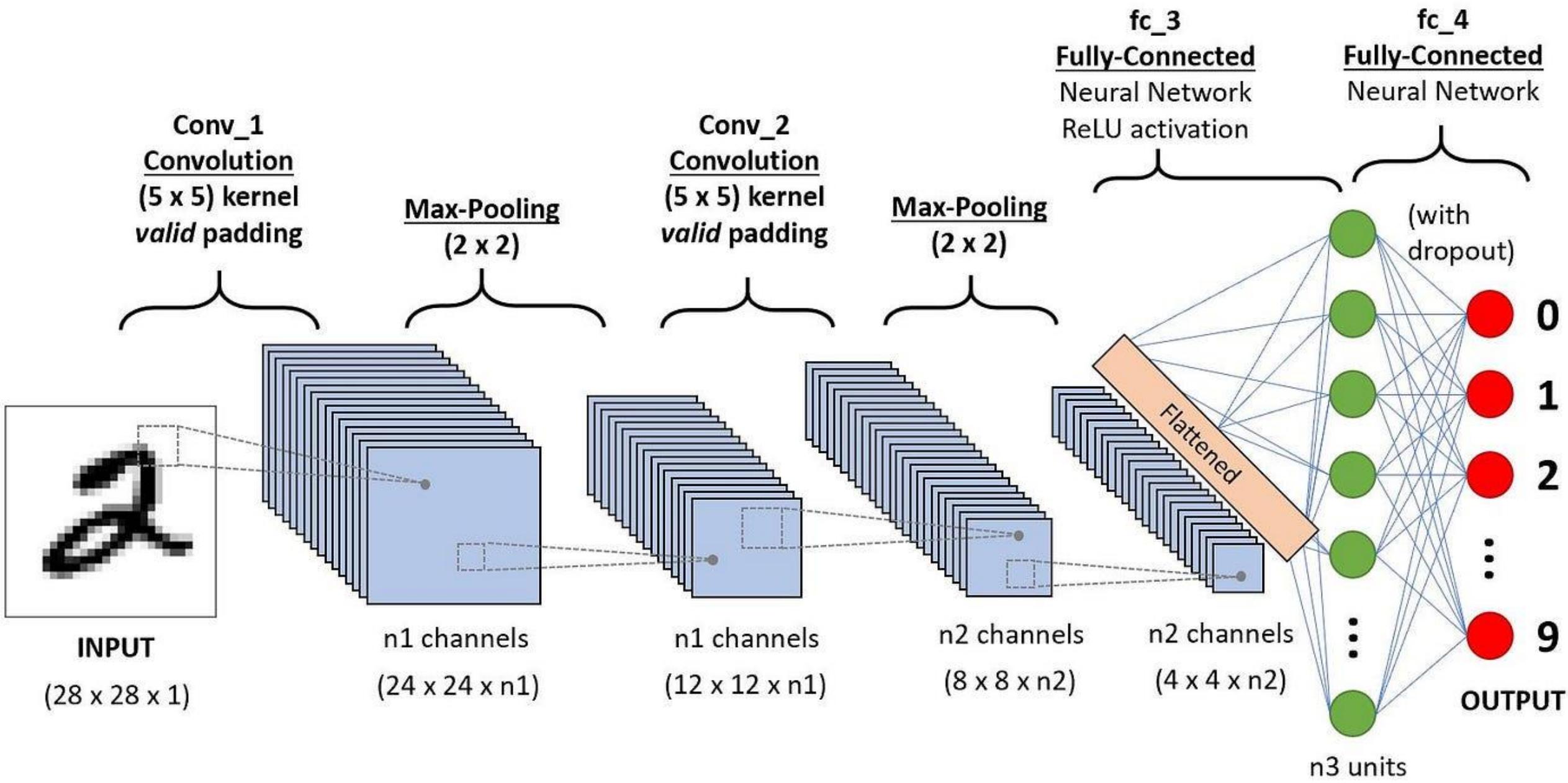
Définition et applications des CNN

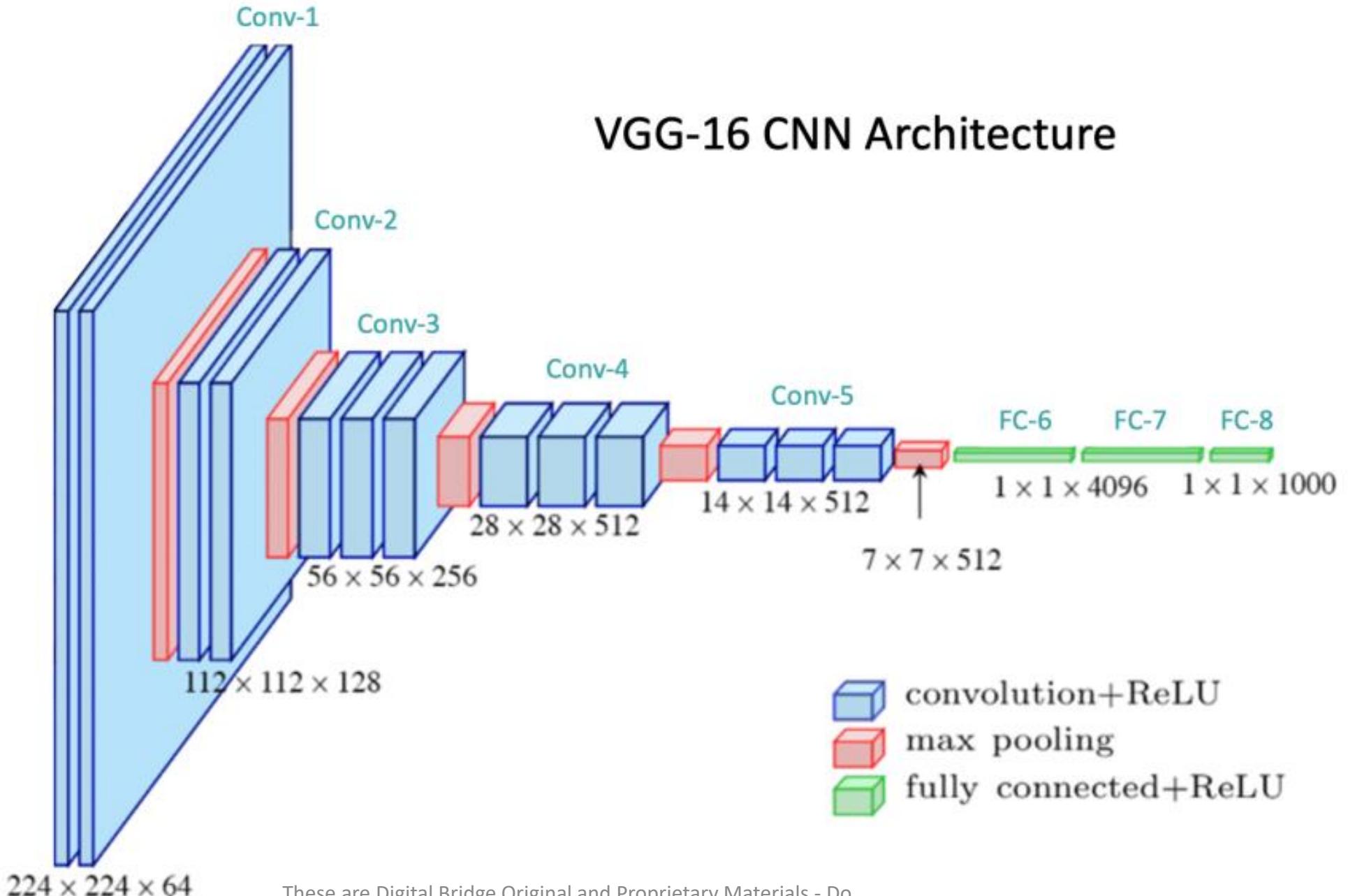
- Les réseaux de neurones convolutifs (CNN) sont un type de réseau neuronal couramment utilisé pour la reconnaissance d'images, la classification et d'autres tâches de vision par ordinateur.
- Les CNN sont conçus pour apprendre automatiquement et de manière adaptative les hiérarchies spatiales des caractéristiques à partir d'images d'entrée ou d'autres sources de données.
- Certaines applications courantes des CNN incluent la classification d'images, la détection d'objets, la reconnaissance faciale et l'analyse d'images médicales.

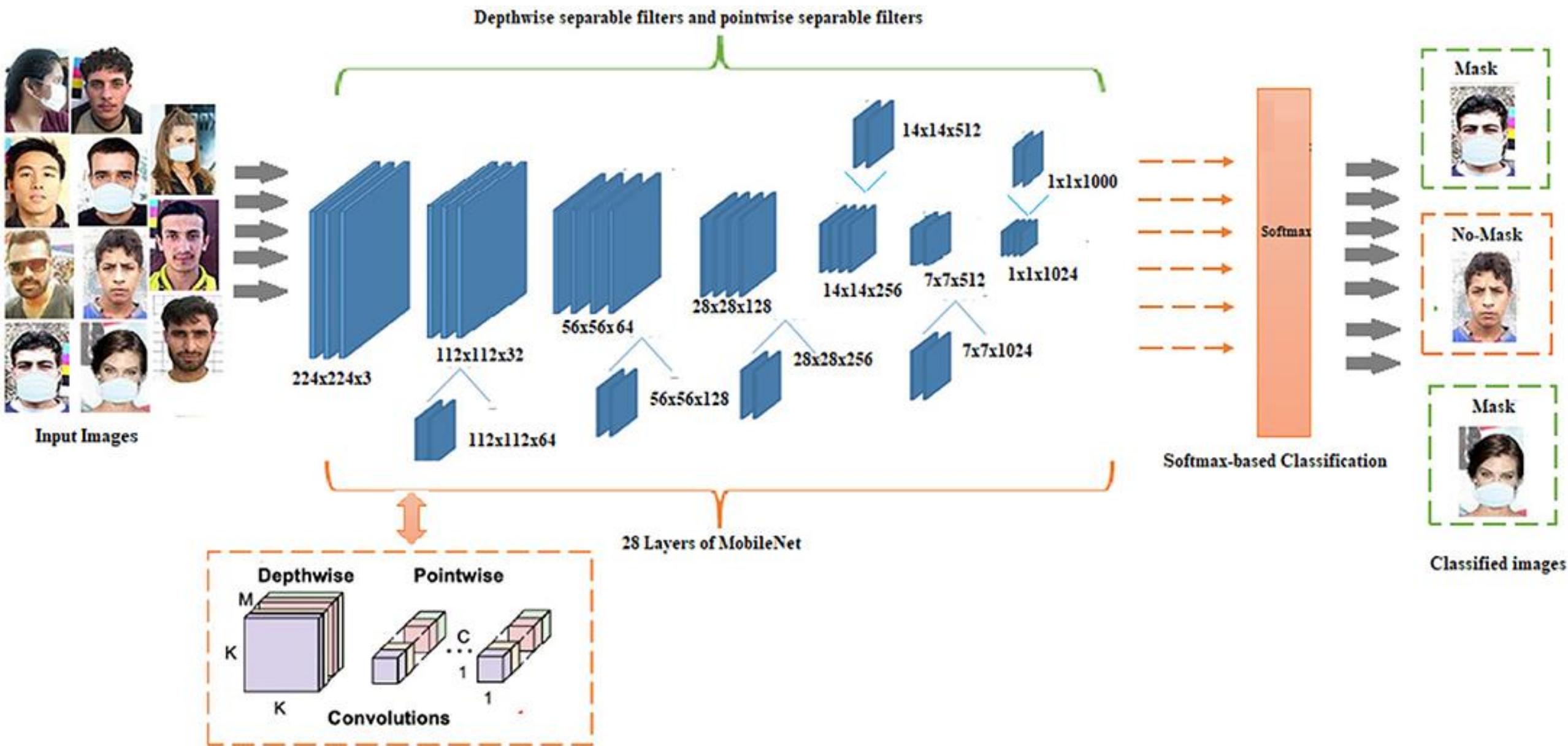
Vue d'ensemble de la structure d'un CNN

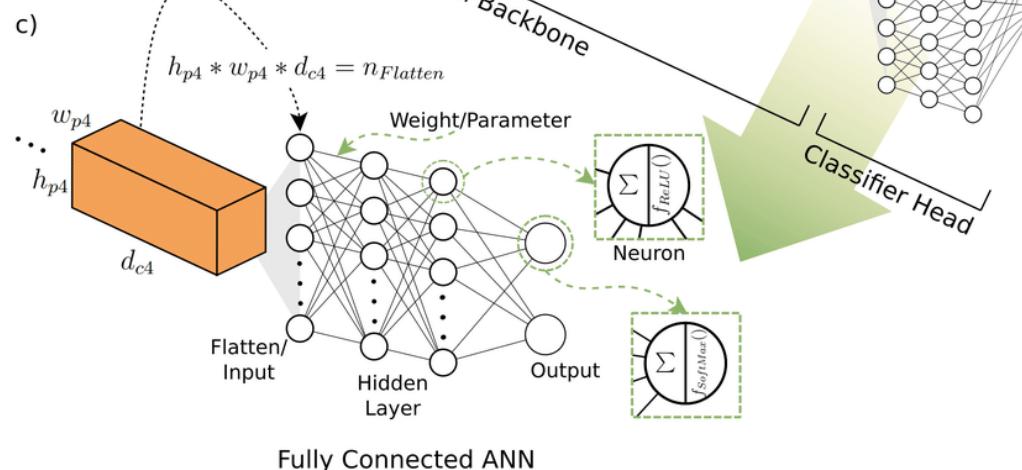
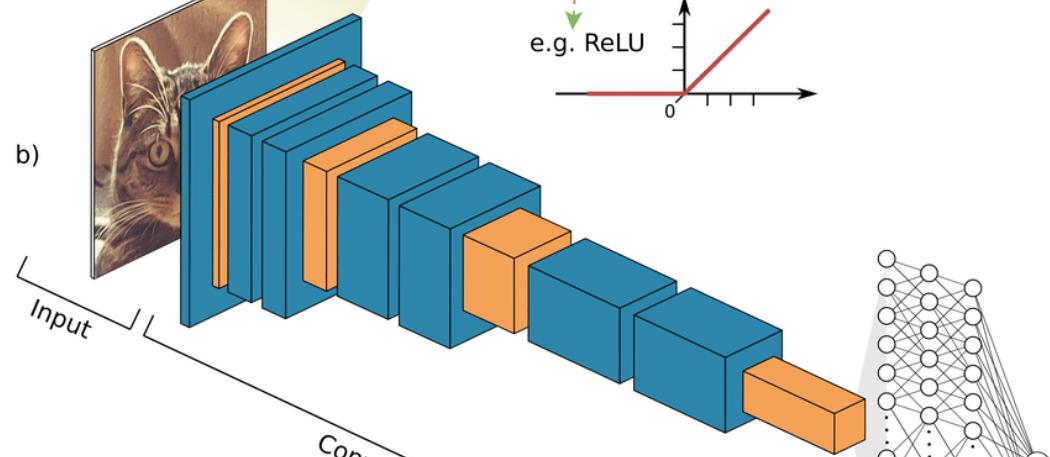
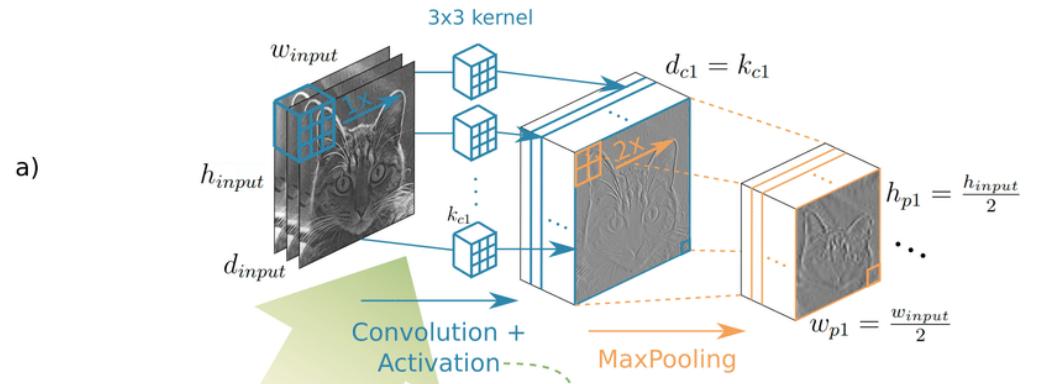
- La structure de base d'un CNN se compose de plusieurs couches de nœuds interconnectés qui traitent les données d'entrée de manière hiérarchique.
- La première couche d'un CNN est généralement une couche convolutionnelle, qui applique un ensemble de filtres à l'image d'entrée pour extraire les caractéristiques pertinentes.
- La sortie de la couche convulsive est généralement introduite dans une ou plusieurs couches de pooling, qui sous-échantillonnent les cartes d'entités pour réduire leur taille et améliorer l'efficacité de calcul.
- Les couches finales d'un CNN sont généralement des couches entièrement connectées, qui utilisent les caractéristiques extraites pour faire des prédictions ou des classifications.





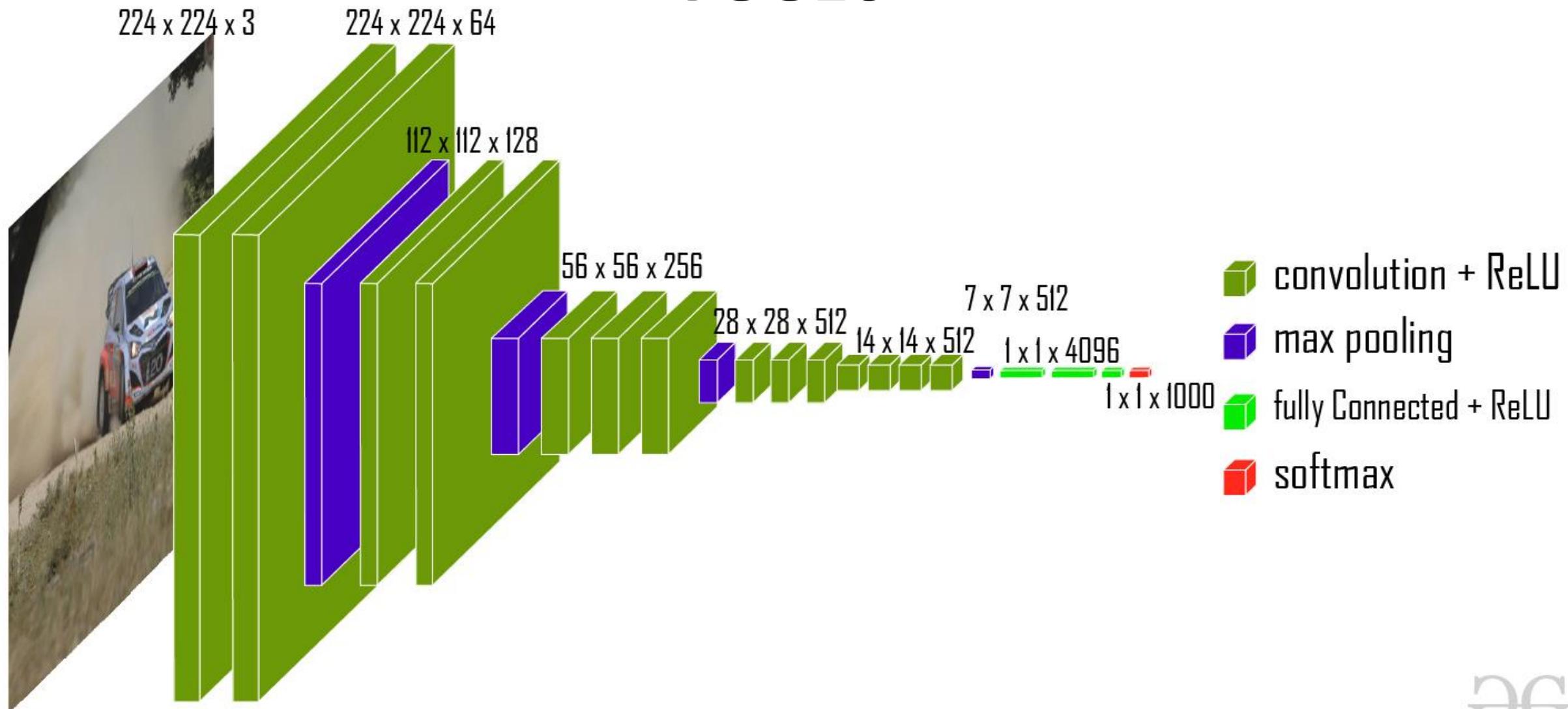






I Bridge Original and Proprietary Materials - Do
use without explicit authorization

VGG16



Avantages des CNN par rapport aux réseaux neuronaux traditionnels

- Les CNN sont conçus pour gérer de grandes entrées de grande dimension comme des images ou des vidéos, qui peuvent être difficiles à traiter pour les réseaux neuronaux traditionnels.
- En utilisant des couches convolutives et de pooling, les CNN peuvent automatiquement apprendre et extraire les fonctionnalités pertinentes des données d'entrée, sans nécessiter d'ingénierie explicite des fonctionnalités.
- Les CNN peuvent également atteindre des niveaux élevés de précision sur un large éventail de tâches de vision par ordinateur, ce qui en fait un choix populaire pour de nombreuses applications du monde réel.

Aperçu historique des CNN

- Les CNN ont été introduits pour la première fois dans les années 1980 et 1990 par des chercheurs comme Yann LeCun, qui a développé l'architecture LeNet-5 pour la reconnaissance numérique manuscrite.
- Cependant, les CNN ne sont devenus largement utilisés qu'au début des années 2010, lorsque les progrès de la puissance de calcul et des ensembles de données à grande échelle ont permis de former des modèles plus profonds et plus complexes.
- Depuis lors, les CNN sont devenus un outil clé dans le domaine de la vision par ordinateur et ont été utilisés dans un large éventail d'applications, des voitures autonomes à l'imagerie médicale.

VGG16

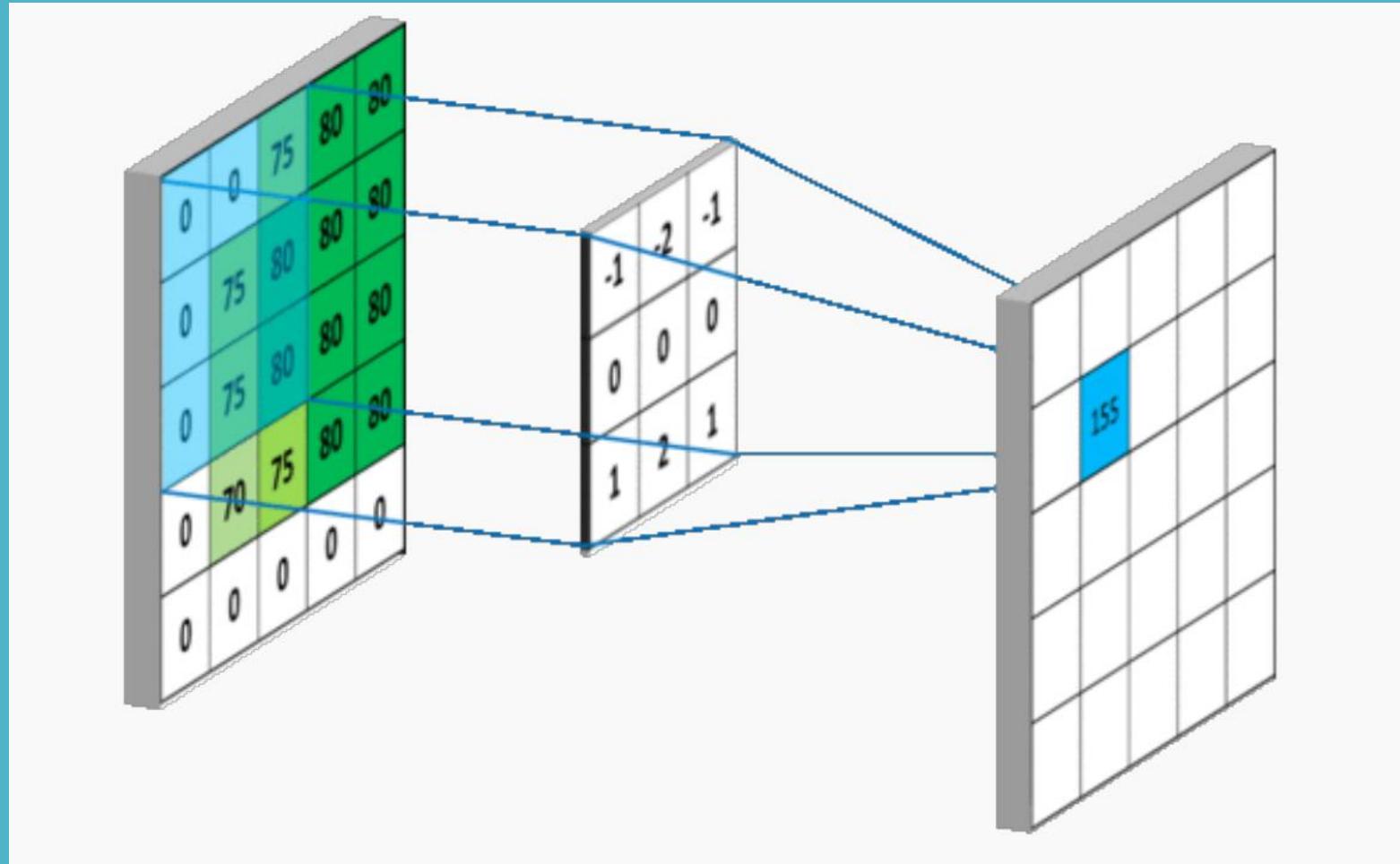
- VGG16 est une architecture de réseau neuronal convolutif qui a été introduite en 2014 par le Visual Geometry Group de l'Université d'Oxford.
- C'est un réseau neuronal profond avec 16 couches, d'où le nom VGG16.
- VGG16 est important car il a obtenu des résultats de pointe sur le jeu de données ImageNet, qui est un grand ensemble de données d'images utilisées pour la reconnaissance et la classification d'images.
- VGG16 utilise une architecture simple qui se compose de couches convolutives répétées 3x3 suivies de couches de pooling 2x2 max, ce qui est devenu un modèle commun dans de nombreux réseaux neuronaux convolutifs.
- L'architecture VGG16 peut être facilement modifiée ou adaptée à différentes tâches, telles que la détection d'objets, la segmentation et l'apprentissage par transfert, ce qui en fait un choix populaire pour de nombreux chercheurs et praticiens.
- VGG16 a également été utilisé comme architecture de référence pour comparer les performances d'autres réseaux neuronaux convolutifs sur diverses tâches et ensembles de données.
- Le succès de VGG16 et d'autres réseaux neuronaux profonds a conduit à l'adoption généralisée de techniques d'apprentissage profond dans de nombreux domaines, notamment la vision par ordinateur, le traitement du langage naturel et la reconnaissance vocale.
- Dans l'ensemble, VGG16 est une étape importante dans le développement de l'apprentissage profond et des réseaux neuronaux convolutifs, et son impact continue de se faire sentir dans de nombreux domaines de la recherche et de l'industrie.
- VGG signifie Visual Geometry Group, un groupe de recherche de l'Université d'Oxford qui a créé les architectures de réseaux neuronaux convolutifs VGG16 et VGG19.

Principes fondamentaux des CNN

Opération de convolution

- L'opération de convolution est le bloc de construction principal d'un CNN et implique l'application d'un ensemble de filtres ou de noyaux à l'image d'entrée pour extraire les caractéristiques pertinentes.
- Chaque filtre est une petite matrice de poids qui est convolvée (ou glissée) à travers l'image d'entrée pour produire une carte d'entités.
- La taille et le nombre de filtres peuvent être ajustés pour contrôler le nombre et la complexité des entités extraites.

$$\begin{matrix} 0 & 25 & 75 & 80 & 80 \\ 0 & 75 & 80 & 80 & 80 \\ 0 & 75 & 80 & 80 & 80 \\ 0 & 70 & 75 & 80 & 80 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \times \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} = \begin{matrix} 0 & 0 & 75 \\ 0 & 0 & 80 \\ 0 & 0 & 80 \end{matrix} \Sigma \begin{matrix} & & & & \\ & & & & \\ & & 235 & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{matrix}$$



| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

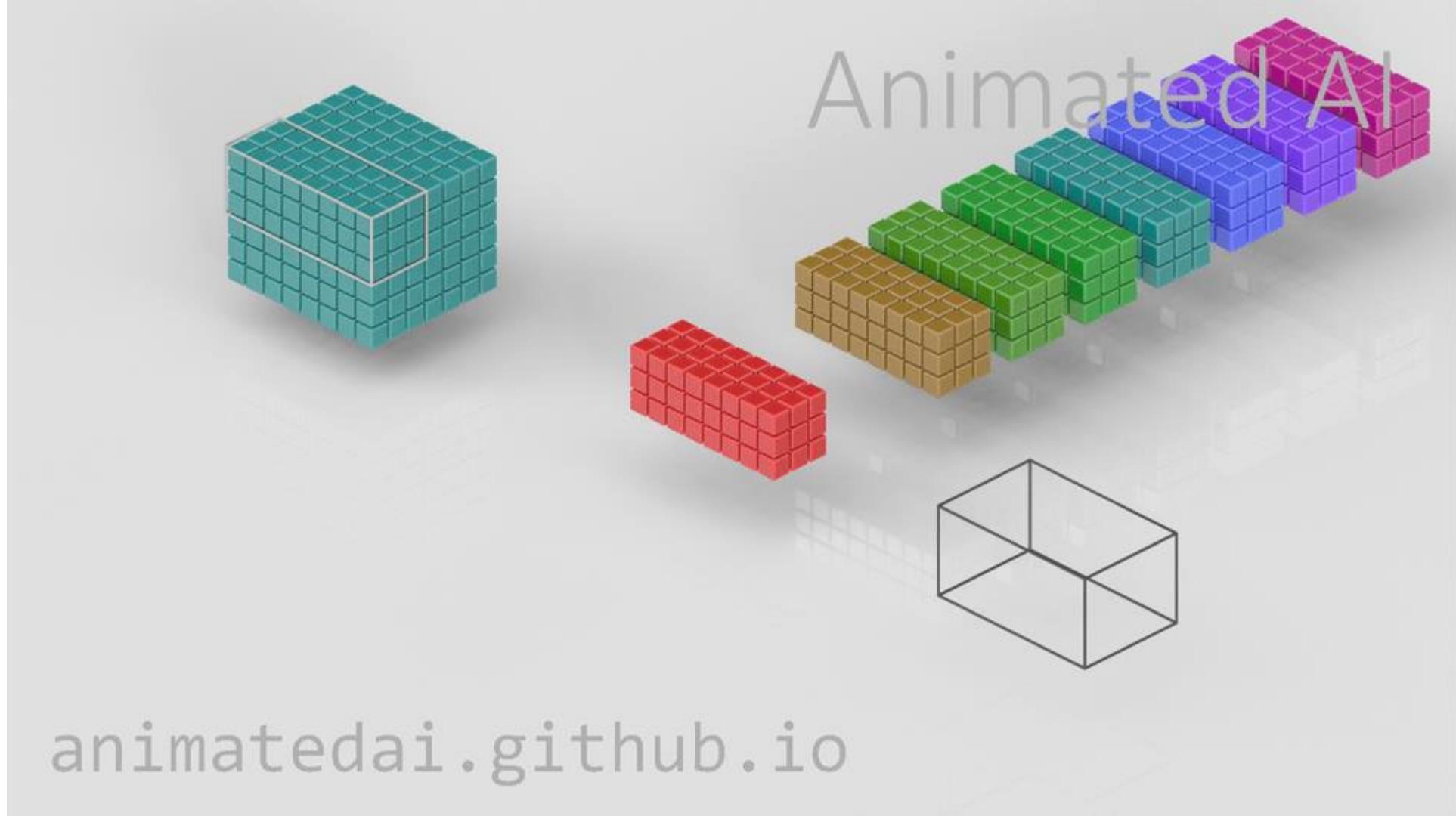
| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

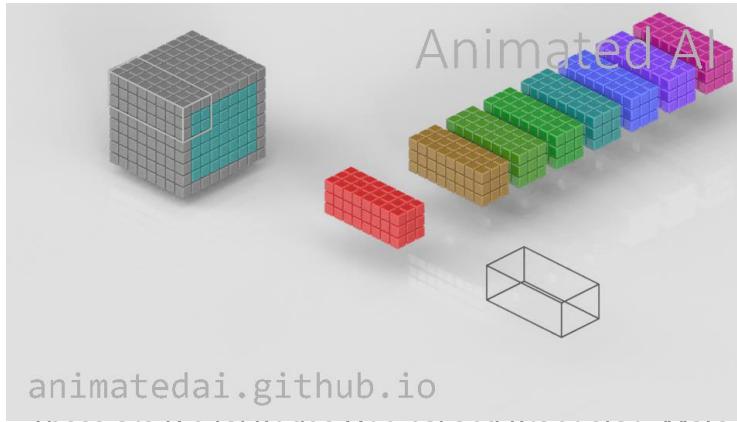
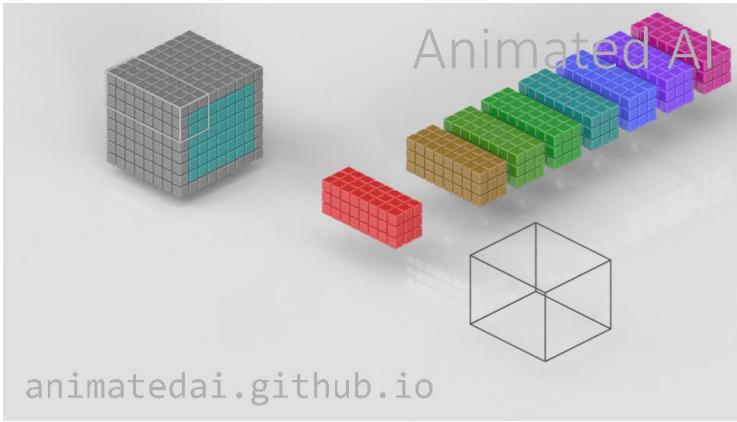
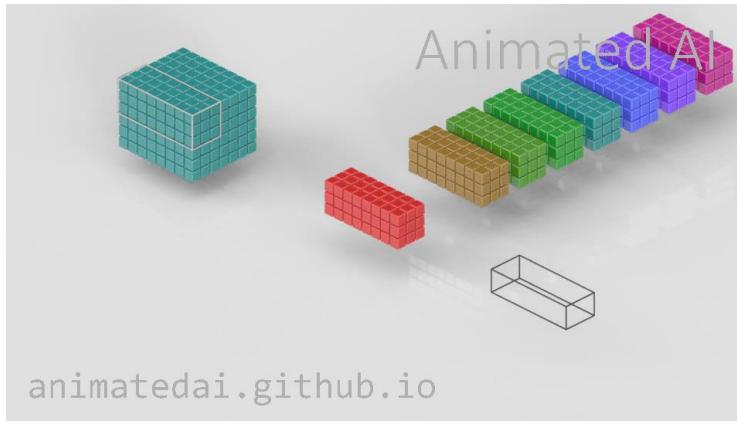
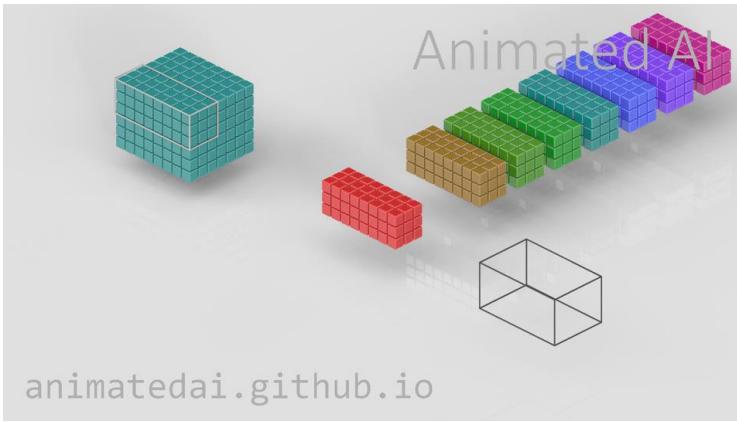
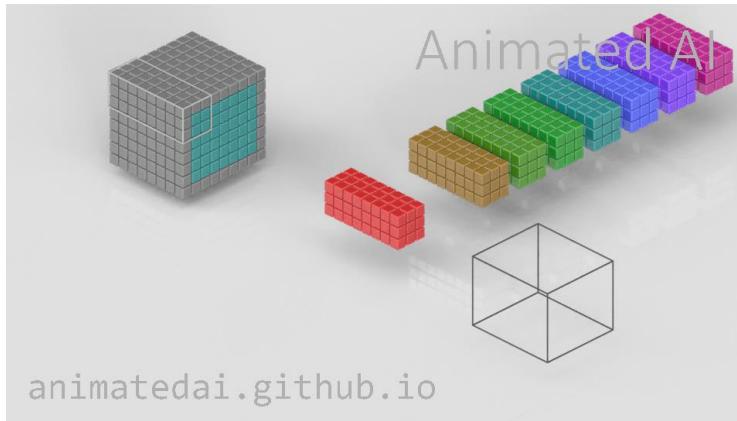
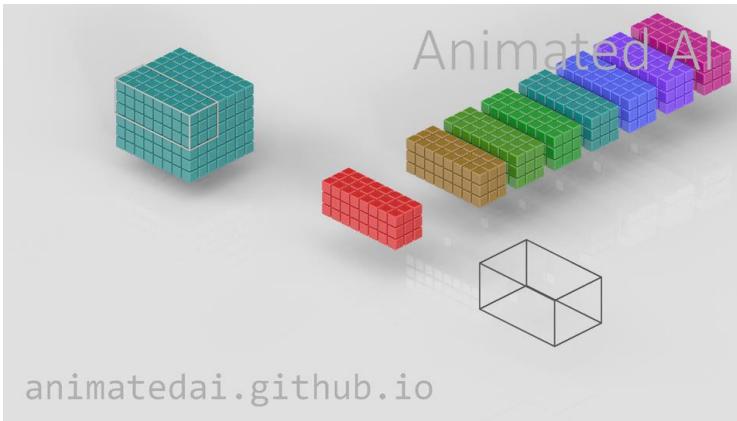
| | | | | |
|---|---|---|---|---|
| 3 | 3 | 2 | 1 | 0 |
| 0 | 0 | 1 | 3 | 1 |
| 3 | 1 | 2 | 2 | 3 |
| 2 | 0 | 0 | 2 | 2 |
| 2 | 0 | 0 | 0 | 1 |

| | | |
|-----|-----|-----|
| 1.7 | 1.7 | 1.7 |
| 1.0 | 1.2 | 1.8 |
| 1.1 | 0.8 | 1.3 |

<https://animatedai.github.io/>



animatedai.github.io



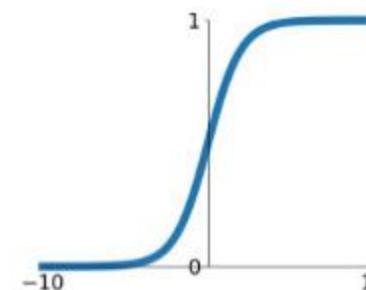
These are Digital Bridge Original and Proprietary Materials - Do
not use without explicit authorization

Fonctions d'activation

- Les fonctions d'activation sont utilisées pour introduire la non-linéarité dans le CNN, ce qui lui permet d'apprendre des mappages complexes et non linéaires entre les données d'entrée et de sortie.
- Les fonctions d'activation courantes utilisées dans les CNN incluent ReLU (Rectified Linear Unit), sigmoïde et tanh.
- ReLU est la fonction d'activation la plus largement utilisée dans les CNN, car elle est efficace sur le plan informatique et a fait ses preuves dans la pratique.

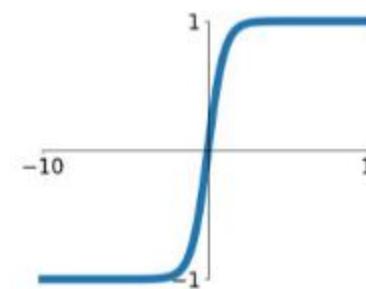
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



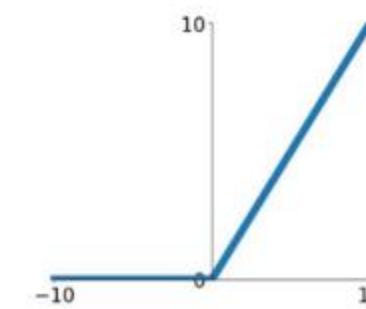
tanh

$$\tanh(x)$$



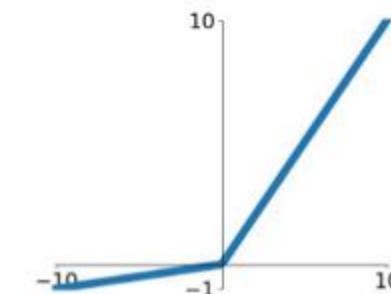
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

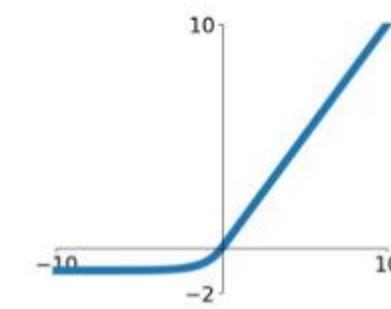


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Input

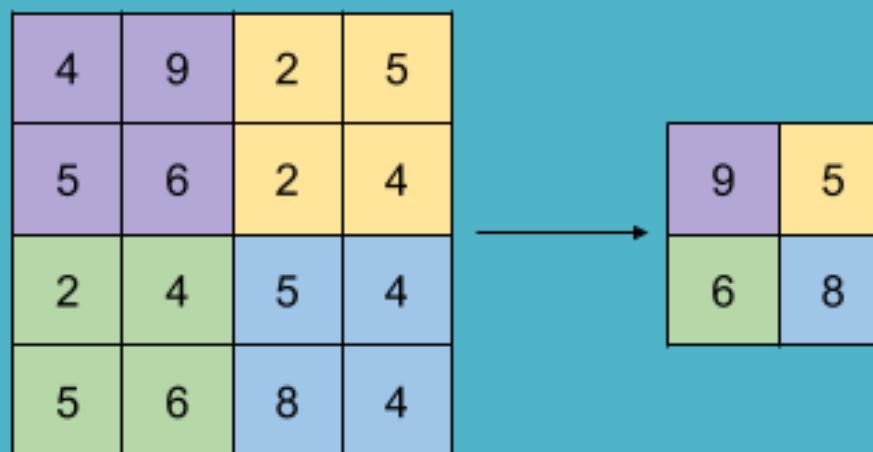
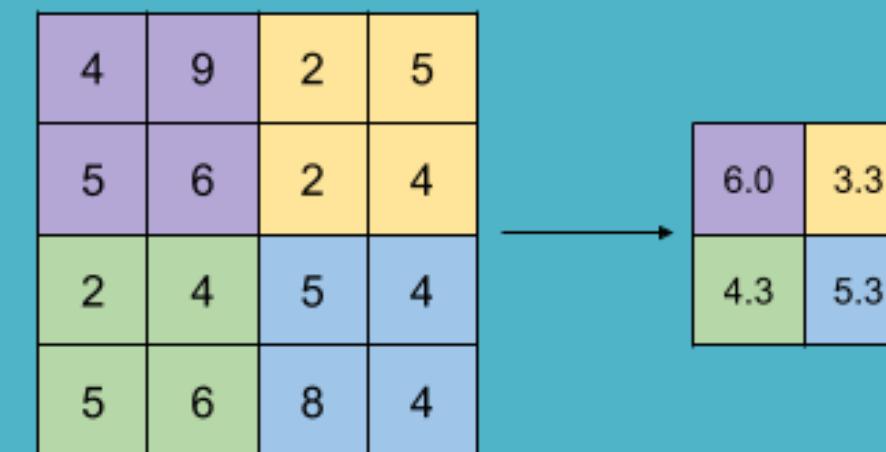
| | | |
|------|------|------|
| -249 | -91 | -37 |
| 250 | -134 | 101 |
| 27 | 61 | -153 |

ReLU

| | | |
|-----|----|-----|
| 0 | 0 | 0 |
| 250 | 0 | 101 |
| 27 | 61 | 0 |

Pooling

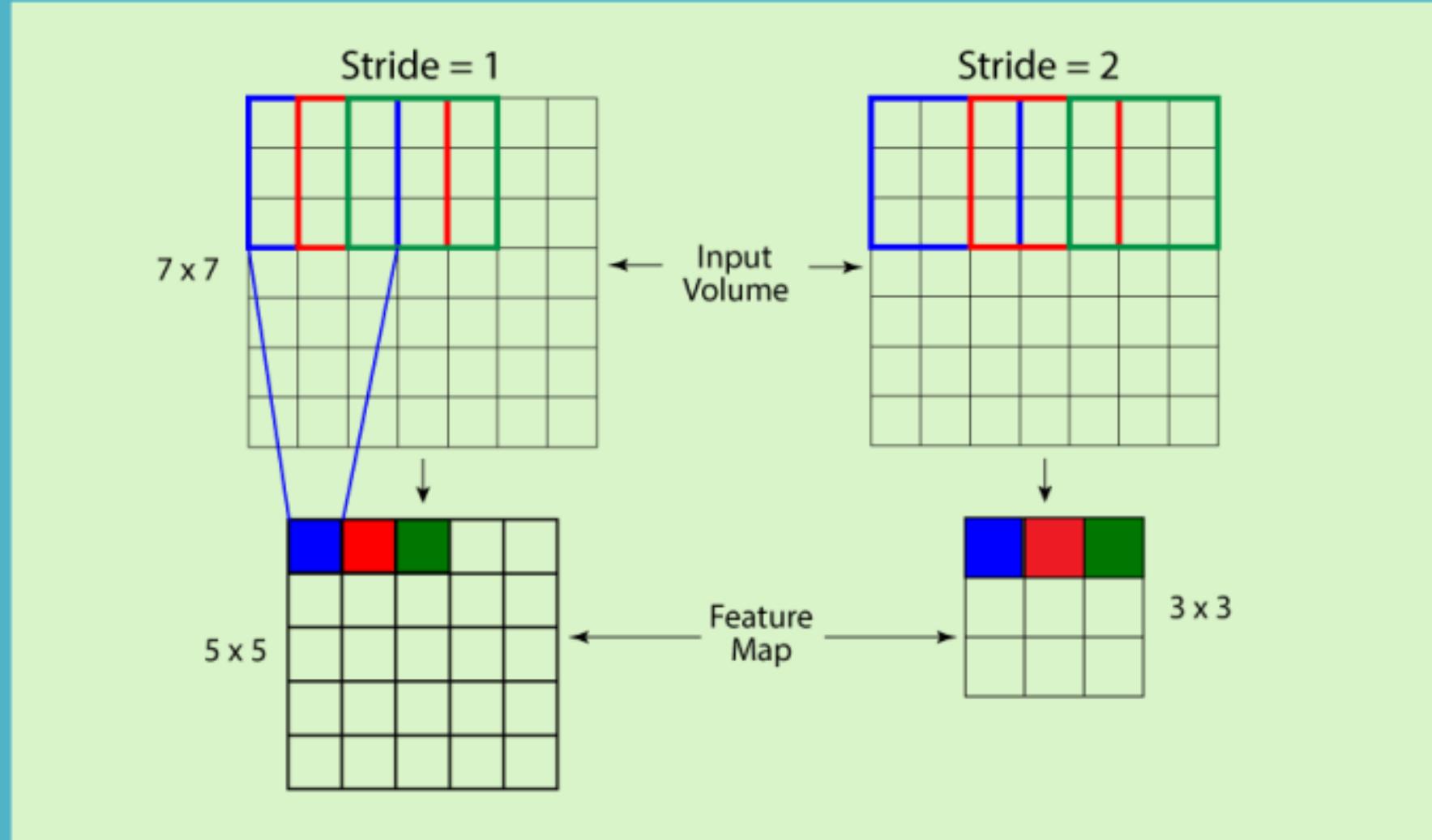
- Le pooling est une technique utilisée pour sous-échantillonner les cartes d'entités produites par les couches convolutionnelles, réduisant ainsi leur taille et améliorant l'efficacité de calcul.
- L'opération de regroupement la plus courante est la mise en pool maximale, qui sélectionne la valeur maximale de chaque région de la carte d'entités et ignore le reste.
- D'autres types d'opérations de mise en commun comprennent la mise en commun moyenne et la mise en commun L2.

Max Pooling***Avg Pooling***

<https://indoml.com>

Strides

- Strides Reportez-vous au nombre de pixels par lesquels le filtre est déplacé sur l'image d'entrée pendant l'opération de convolution.
- Une valeur de foulée plus élevée se traduit par une carte d'entités de sortie plus petite, car moins de circonvolutions sont effectuées.
- Les foulées peuvent être ajustées pour contrôler la quantité de sous-échantillonnage et le niveau de détail des caractéristiques extraites.

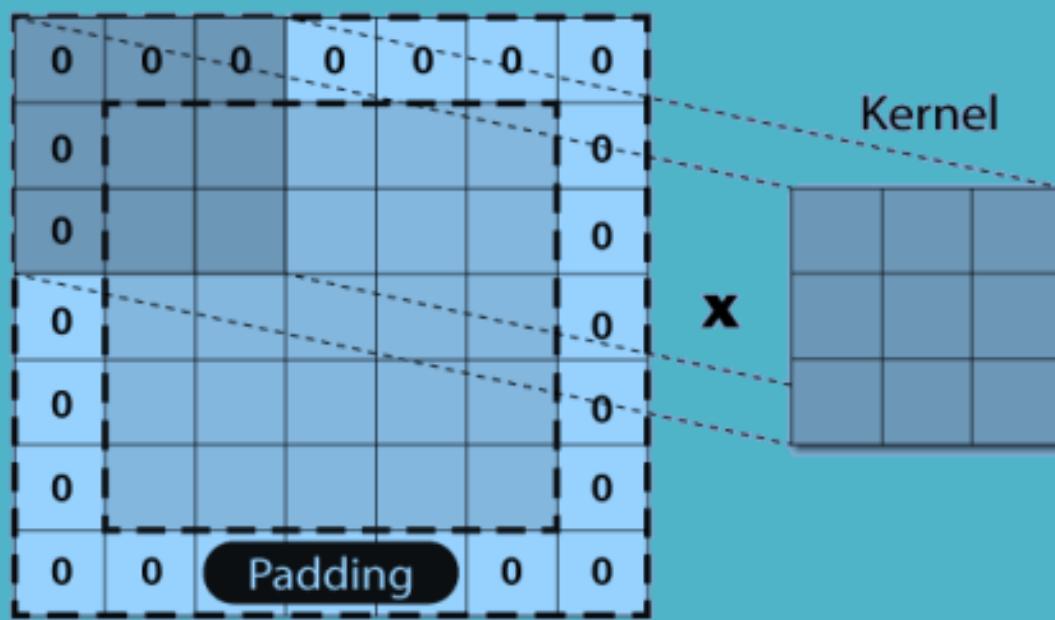


Padding

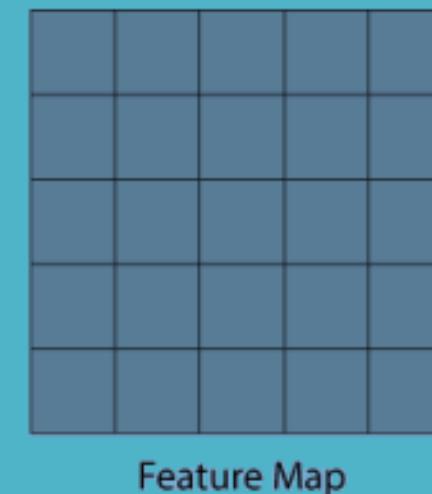
- Padding est une technique utilisée pour ajouter des pixels supplémentaires autour des bords de l'image d'entrée, afin d'éviter que la carte d'entités de sortie ne devienne plus petite que l'image d'entrée.
- Il existe deux types de rembourrage : le remplissage valide, où aucun remplissage n'est ajouté et où la carte des entités de sortie est plus petite que l'image d'entrée ; et le même remplissage, où le remplissage est ajouté afin que la carte des entités de sortie ait la même taille que l'image d'entrée.
- Le même remplissage est couramment utilisé dans les CNN pour s'assurer que les cartes d'entités de sortie ont les mêmes dimensions spatiales que les images d'entrée, ce qui facilite leur combinaison en couches entièrement connectées..

PADDING

Input



Output



Création de CNN

Création d'une architecture CNN

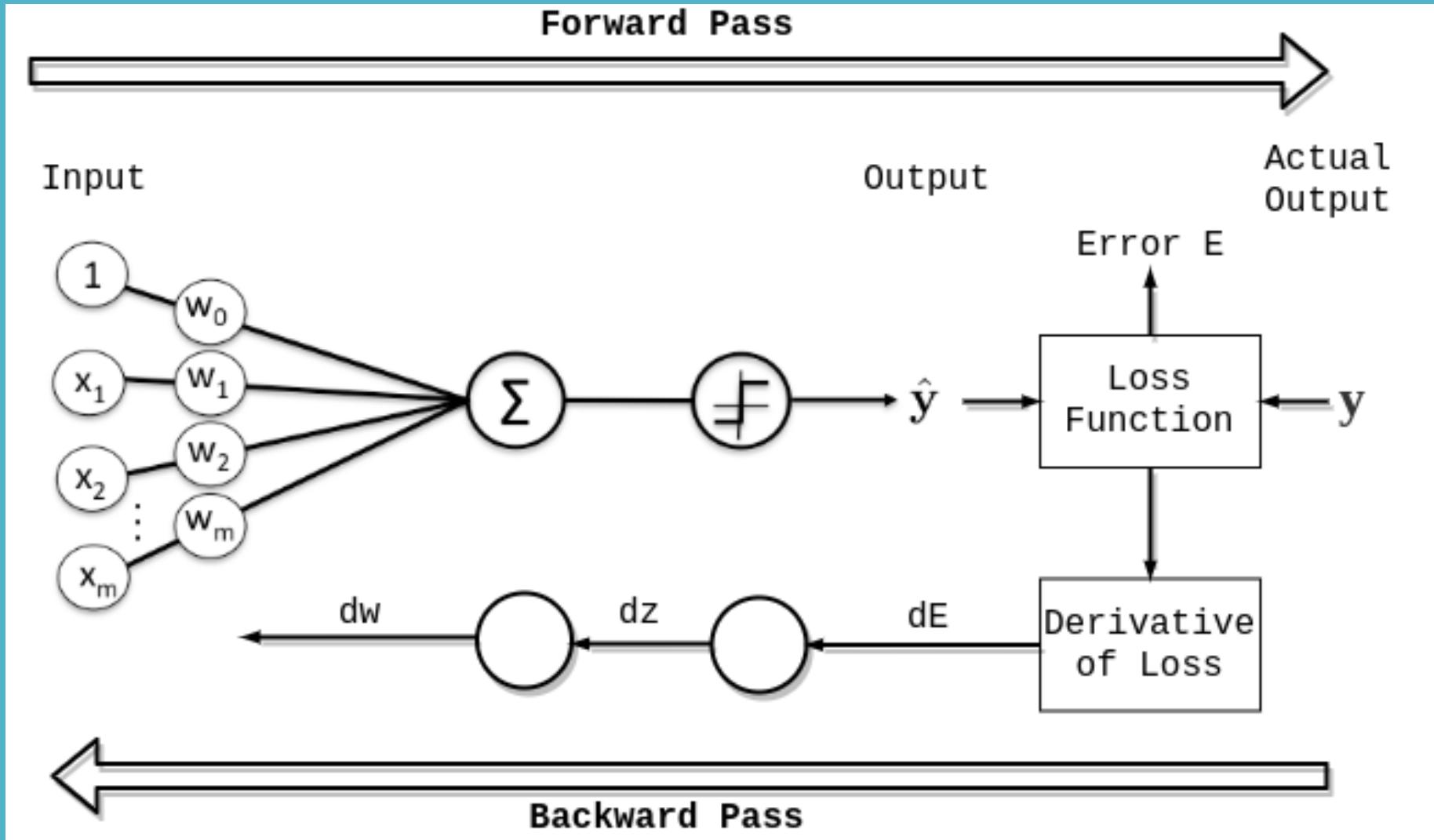
- Le processus de construction d'une architecture CNN implique de décider du nombre et du type de couches à inclure, de la taille des filtres, des fonctions d'activation, des opérations de mise en commun et du nombre de neurones dans les couches entièrement connectées.
- Les architectures CNN peuvent être personnalisées pour des tâches spécifiques, telles que la classification d'images ou la détection d'objets, en ajustant le nombre et le type de couches et d'autres hyperparamètres.

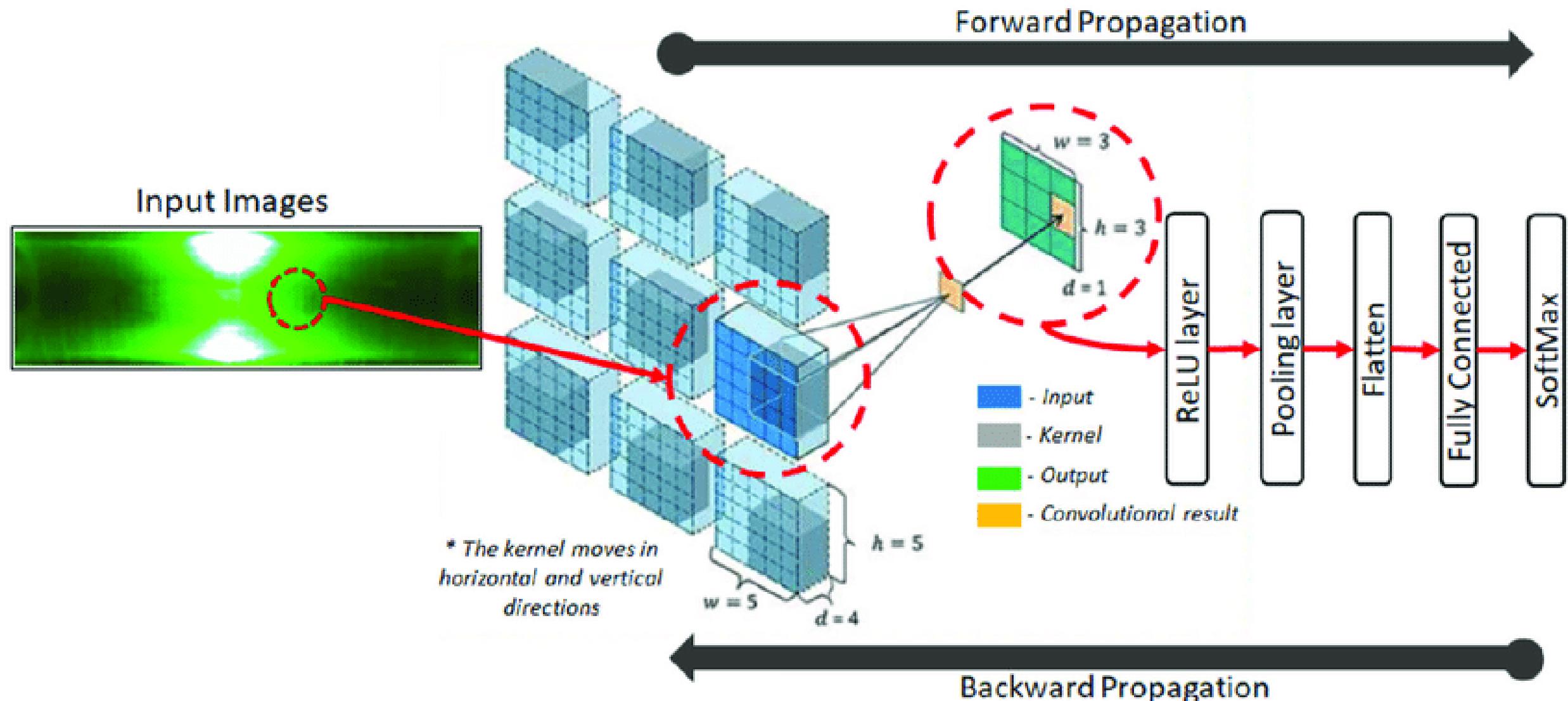
Initialisation du poids

- L'initialisation du poids est le processus de définition des valeurs initiales des poids dans le CNN.
- Une bonne initialisation du poids est importante pour s'assurer que le CNN converge rapidement et évite le surajustement.
- Les techniques d'initialisation de poids courantes incluent l'initialisation aléatoire, l'initialisation Xavier et l'initialisation He.

Propagation vers l'avant et vers l'arrière

- La propagation vers l'avant fait référence au processus de transmission des données d'entrée à travers le CNN pour produire une sortie prévue.
- La propagation en arrière, ou rétropropagation, fait référence au processus de calcul des gradients de la fonction de perte par rapport aux poids dans le CNN.
- La rétropropagation est utilisée pour mettre à jour les poids dans le CNN pendant l'entraînement, en utilisant un algorithme d'optimisation comme la descente de gradient stochastique (SGD).





Fonctions de perte

- Les fonctions de perte sont utilisées pour mesurer la différence entre la sortie prévue du CNN et la sortie réelle.
- Les fonctions de perte courantes utilisées dans les CNN comprennent l'erreur quadratique moyenne (MSE), l'entropie croisée binaire et l'entropie croisée catégorielle.
- Le choix de la fonction de perte dépend de la tâche spécifique et du type de sortie prévu.

Means Square Error

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where n is the number of data points, y_i is the actual output for the i^{th} input, and \hat{y}_i is the predicted output for the i^{th} input.

Cross Entropy

$$\text{Cross-entropy} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij})$$

where n is the number of data points, C is the number of classes,
 y_{ij} is the actual output for the i^{th} input and j^{th} class,
and \hat{y}_{ij} is the predicted probability of the input belonging to the j^{th} class.

Optimiseurs

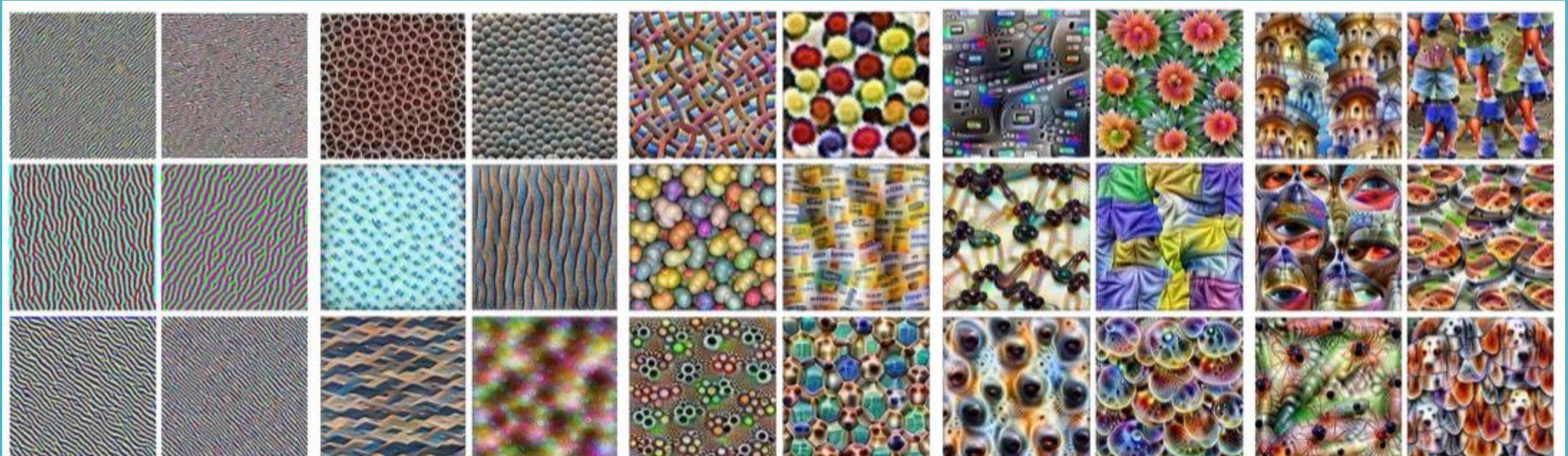
- Les optimiseurs sont utilisés pour mettre à jour les poids dans le CNN pendant l'entraînement, en fonction des gradients calculés lors de la rétropropagation.
- Les optimiseurs courants utilisés dans les CNN incluent SGD, Adam, Adagrad et RMSprop.
- Le choix de l'optimiseur peut avoir un impact significatif sur les performances du CNN et peut dépendre de facteurs tels que la taille du jeu de données et la complexité de l'architecture.

| Name | Update Rule |
|----------|---|
| SGD | $\Delta\theta_t = -\alpha g_t$ |
| Momentum | $m_t = \gamma m_{t-1} + (1 - \gamma)g_t,$ $\Delta\theta_t = -\alpha m_t$ |
| Adagrad | $G_t = G_{t-1} + g_t^2,$ $\Delta\theta_t = -\alpha g_t G_t^{-1/2}$ |
| Adadelta | $v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2,$ $\Delta\theta_t = -\alpha g_t v_t^{-1/2} D_{t-1}^{1/2},$ $D_t = \beta_1 D_{t-1} + (1 - \beta_1)(\Delta\theta_t/\alpha)^2$ |
| RMSprop | $v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2,$ $\Delta\theta_t = -\alpha g_t v_t^{-1/2}$ |
| Adam | $m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_t,$ $v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2,$ $\hat{m}_t = m_t / (1 - \beta_1^t),$ $\hat{v}_t = v_t / (1 - \beta_2^t),$ $\Delta\theta_t = -\alpha \hat{m}_t \hat{v}_t^{-1/2}$ |

Mise au point des CNN

Apprentissage par transfert

- L'apprentissage par transfert est le processus d'utilisation d'un CNN pré-formé comme point de départ pour une nouvelle tâche.
- En tirant parti des fonctionnalités apprises par le CNN pré-entraîné, il est possible d'atteindre une grande précision avec moins de données d'entraînement et de temps de calcul.
- L'apprentissage par transfert est particulièrement utile lorsque la nouvelle tâche est similaire à la tâche initiale pour laquelle le CNN a été formé.



Edges

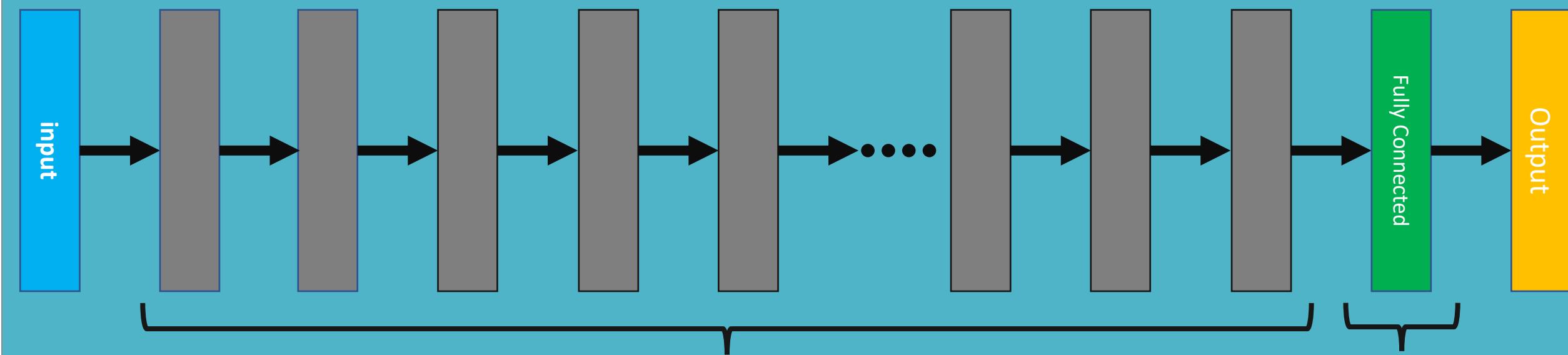
Textures

Patterns

Object Parts

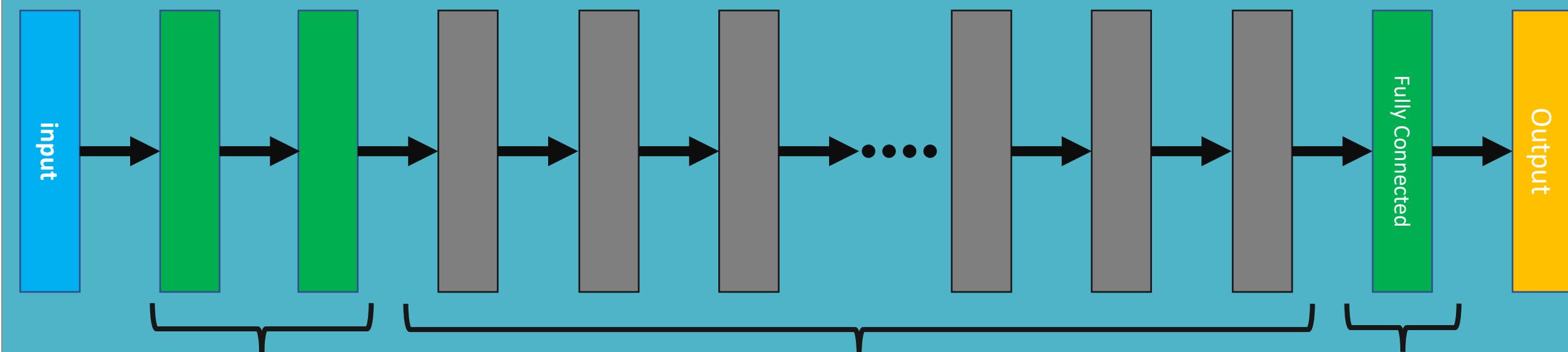
Objects

Scenario I: train FC layers only

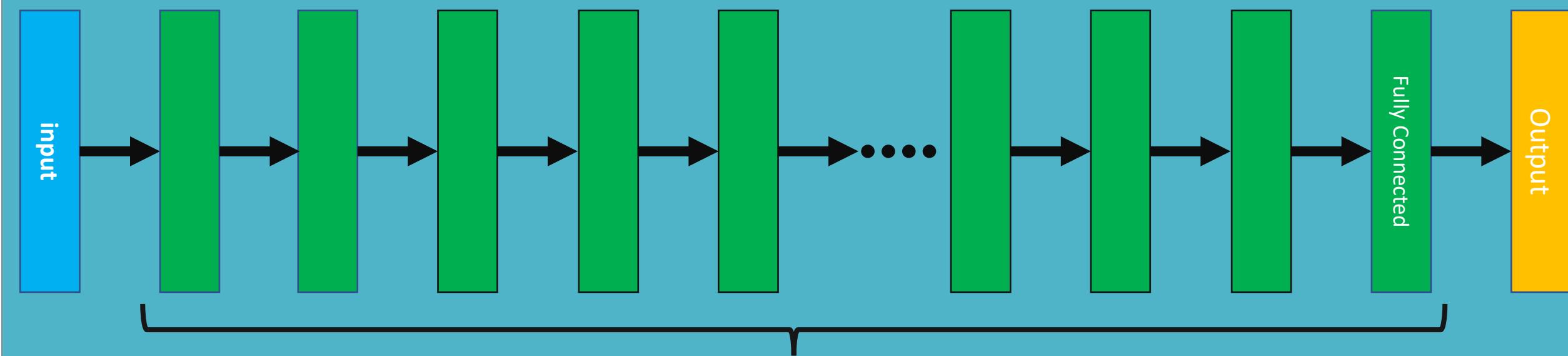


- Freeze these layers
- Weight from pre-trained model
- Feature extraction
- New data similar to original model
- Train these layers

Scenario II: train some layers



Scenario III: train all layers



- Train these layers
- Weights from pre-trained model
- Feature extraction
- New data not similar to original model

Modèles pré-entraînés

- Les modèles préentraînés sont des CNN qui ont été formés sur de grands ensembles de données pour des tâches spécifiques, telles que la classification d'images ou la détection d'objets.
- De nombreux modèles pré-entraînés sont accessibles au public et peuvent être téléchargés et utilisés pour l'apprentissage par transfert.
- Des exemples de modèles pré-entraînés populaires incluent VGG, ResNet et Inception.

Techniques de réglage fin

- Le réglage fin est le processus de modification du CNN pré-entraîné pour l'adapter à une nouvelle tâche, en réentraînant certaines ou toutes les couches avec de nouvelles données.
- Le réglage fin peut impliquer l'ajustement du taux d'apprentissage, la congélation ou le dégel de certaines couches et l'utilisation de différents algorithmes d'optimisation.
- Le réglage fin peut prendre beaucoup de temps et nécessite un réglage minutieux des hyperparamètres.

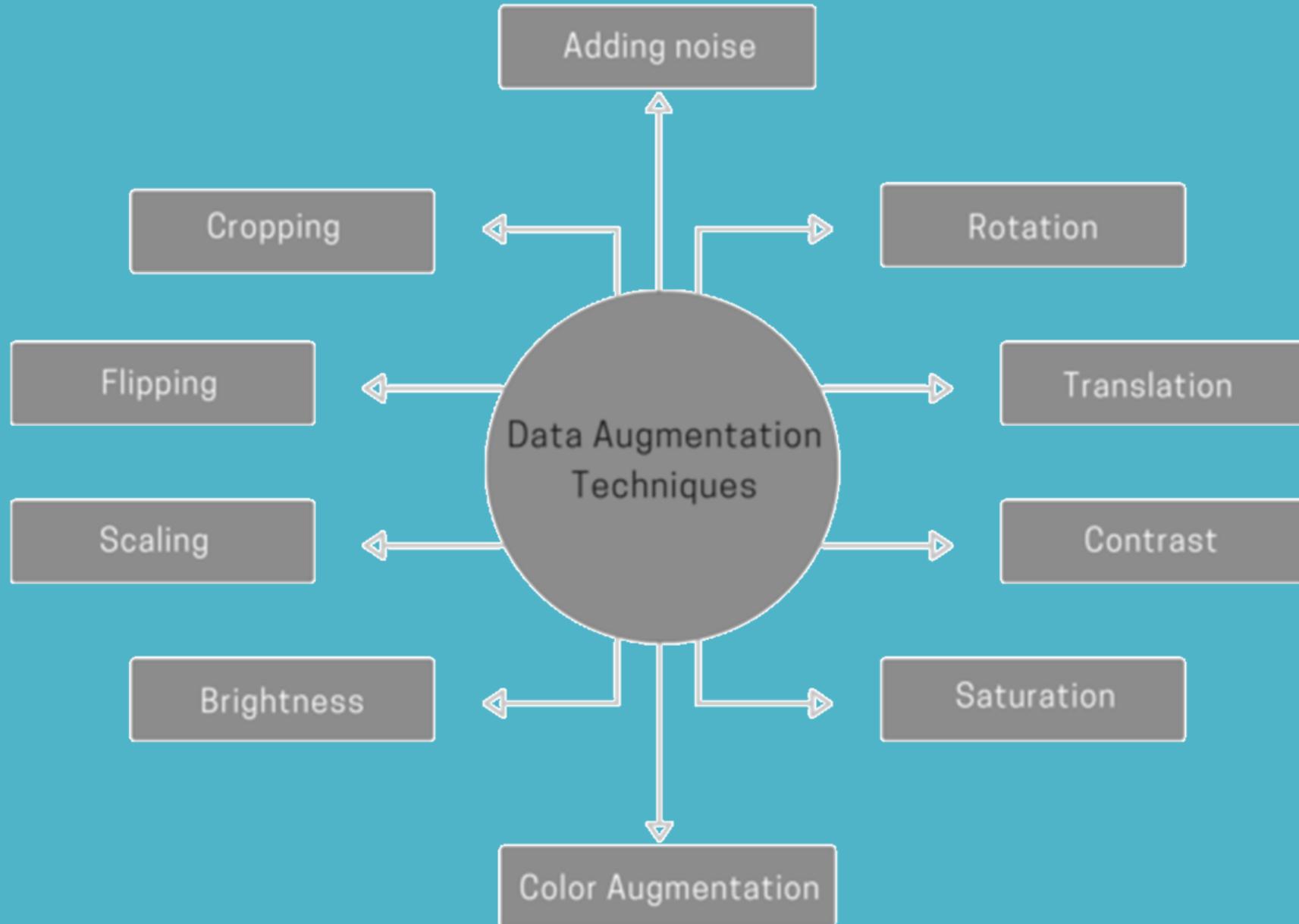
Régularisation

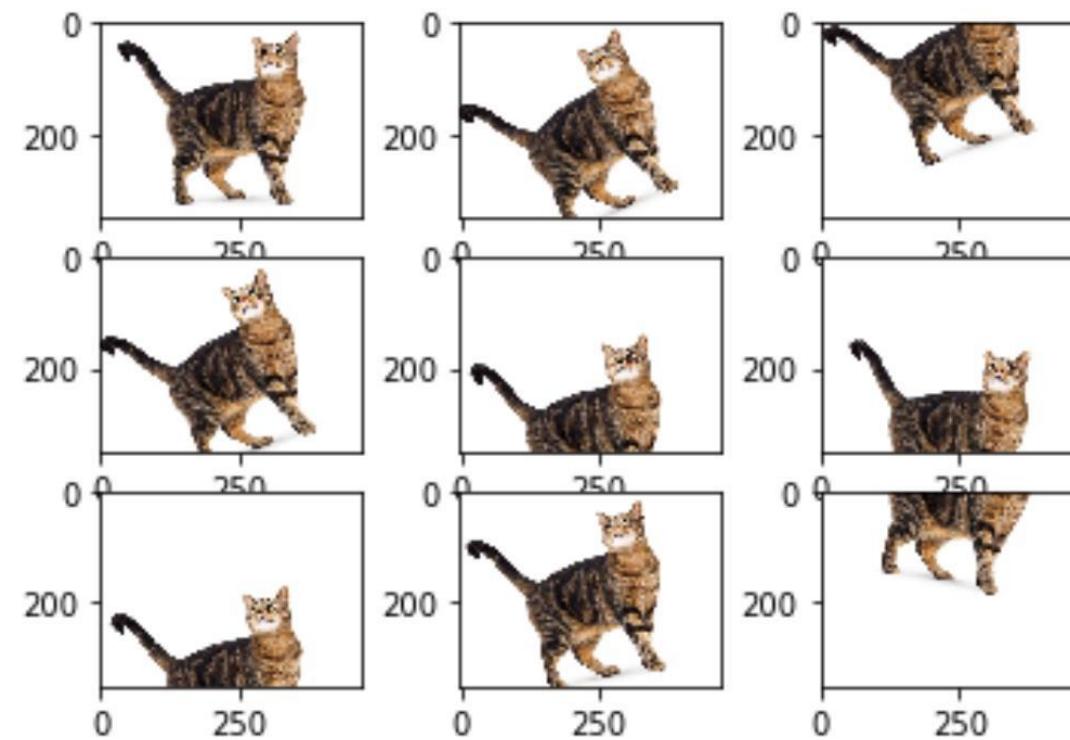
- Les techniques de régularisation peuvent être utilisées pour éviter le surajustement lors de la mise au point d'un CNN.
- Les techniques de régularisation courantes comprennent la régularisation L1 et L2, l'abandon et l'arrêt précoce.
- La régularisation peut aider à améliorer la performance de généralisation du CNN, en l'empêchant de mémoriser les données d'entraînement.

Amélioration des performances CNN

Augmentation des données

- L'augmentation des données est une technique utilisée pour augmenter la taille et la diversité du jeu de données d'apprentissage, en appliquant des transformations aléatoires aux images d'entrée.
- Les techniques courantes d'augmentation des données incluent le recadrage, la rotation, le retournement et le réglage de la luminosité aléatoires.
- L'augmentation des données peut aider à améliorer les performances de généralisation du CNN, en l'exposant à un plus large éventail de variations d'entrée.





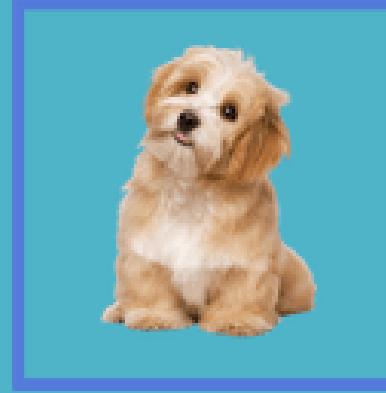
Original



Rotation



Flip



Scaling



Brightness





Flipping



Rotating



Cropping



Original

Color distortion



Sharpening



Random erasing



Réglage des hyperparamètres

- Le réglage des hyperparamètres consiste à trouver les meilleures valeurs pour les hyperparamètres du CNN, telles que le taux d'apprentissage, le nombre de filtres, la taille du noyau et le nombre de couches.
- Le réglage des hyperparamètres peut être effectué à l'aide de diverses méthodes, telles que la recherche par grille, la recherche aléatoire ou l'optimisation bayésienne.
- Un réglage correct des hyperparamètres est important pour obtenir les meilleures performances du CNN.

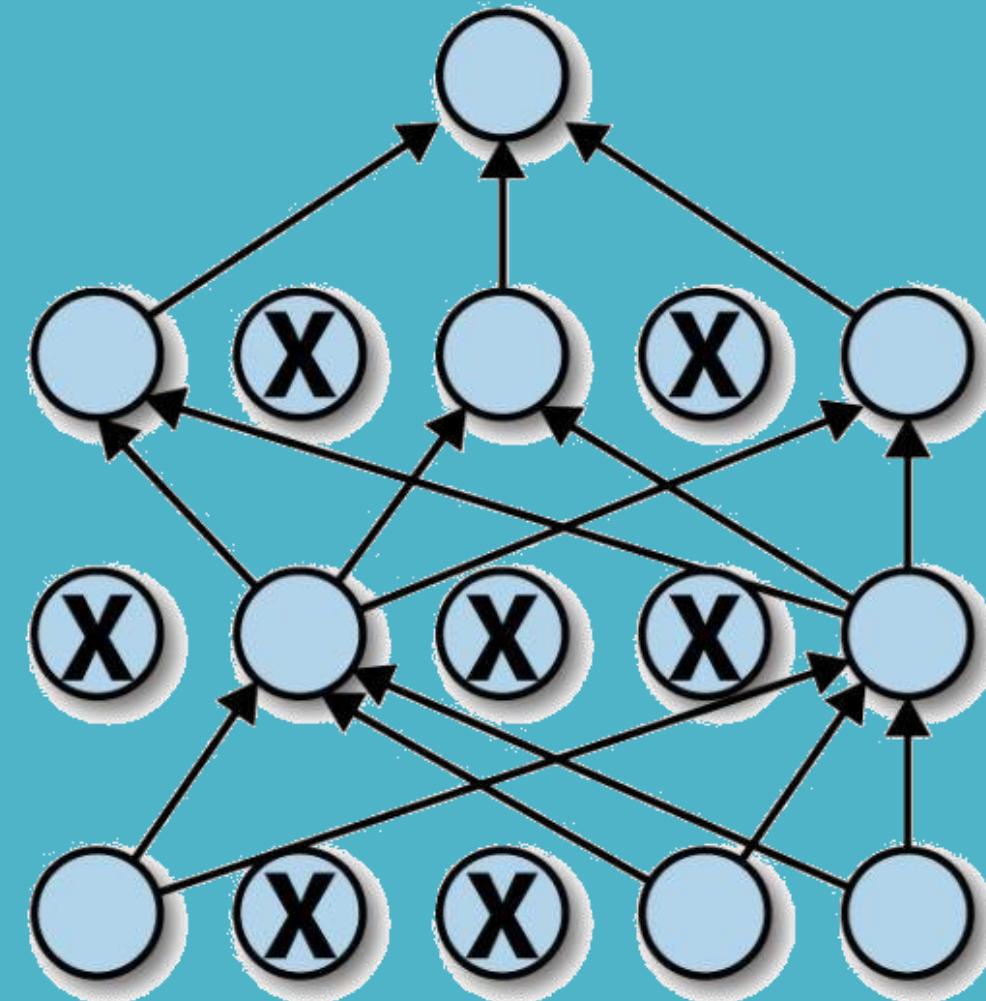
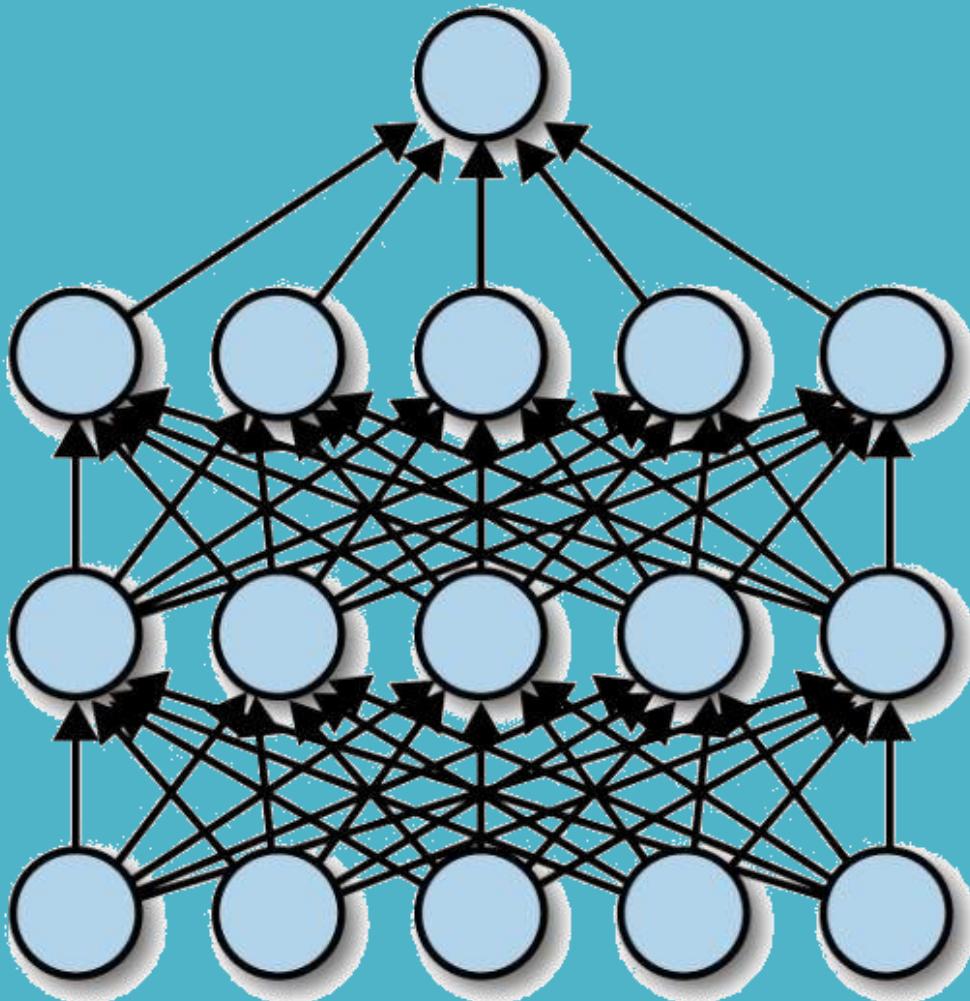
Normalisation des lots (Batch normalization)

- La normalisation par lots est une technique utilisée pour améliorer la stabilité et les performances du CNN, en normalisant les activations de chaque couche.
- La normalisation par lots réduit le décalage de covariable interne, ce qui peut ralentir le processus d'optimisation et le faire converger plus lentement.
- La normalisation par lots peut aider à améliorer la précision et la vitesse du CNN, en particulier pour les architectures plus profondes.

Dropout

- Dropout est une technique de régularisation utilisée pour prévenir le surajustement dans le CNN, en abandonnant au hasard certains neurones pendant l'entraînement.
- Dropout peut aider à empêcher le CNN de mémoriser les données d'apprentissage et peut améliorer ses performances de généralisation.
- Dropout peut être appliqué à l'entrée ou aux couches cachées du CNN, et le taux d'abandon peut être ajusté pour contrôler la force de la régularisation.

Dropout

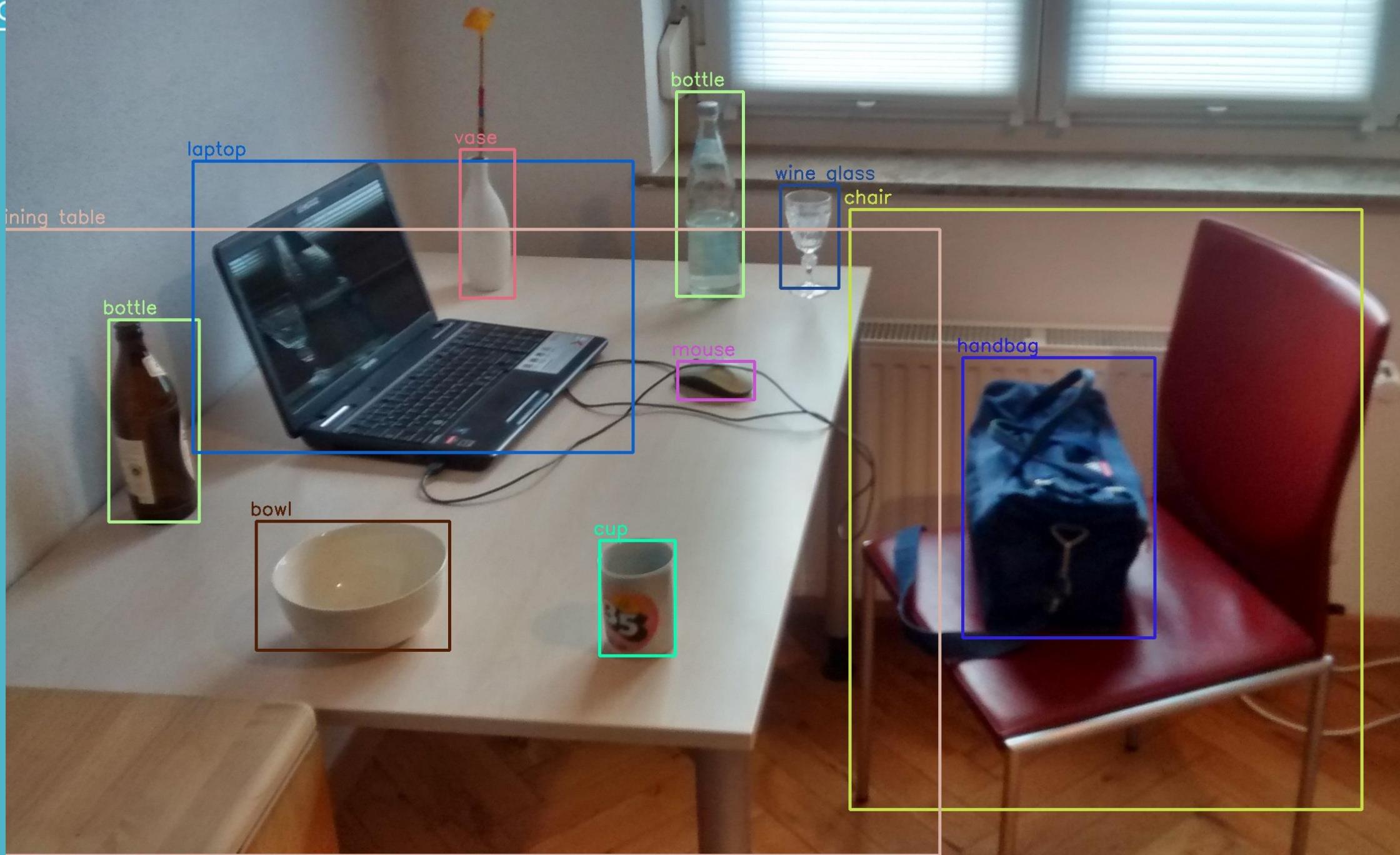


These are Digital Bridge Original and Proprietary Materials - Do
not use without explicit authorization

Études de cas et applications

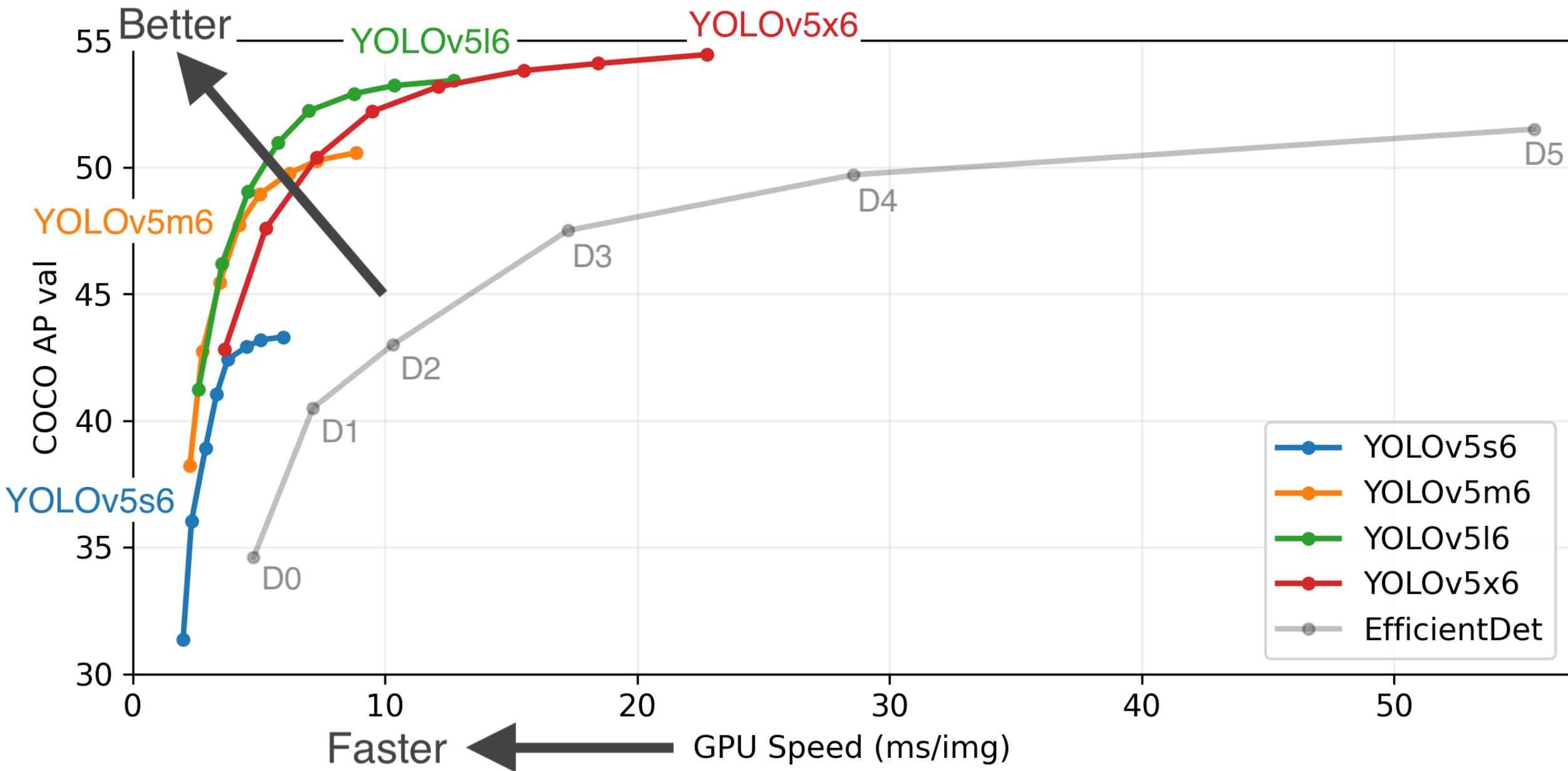
Détection d'objets

- La détection d'objets est la tâche de détection et de localisation d'objets dans une image ou une vidéo.
- Les CNN sont couramment utilisés pour la détection d'objets, en utilisant des techniques telles que les CNN régionaux (R-CNN), You Only Look Once (YOLO) et Single Shot Detector (SSD).
- La détection d'objets a de nombreuses applications, telles que les voitures autonomes, la surveillance et la robotique.



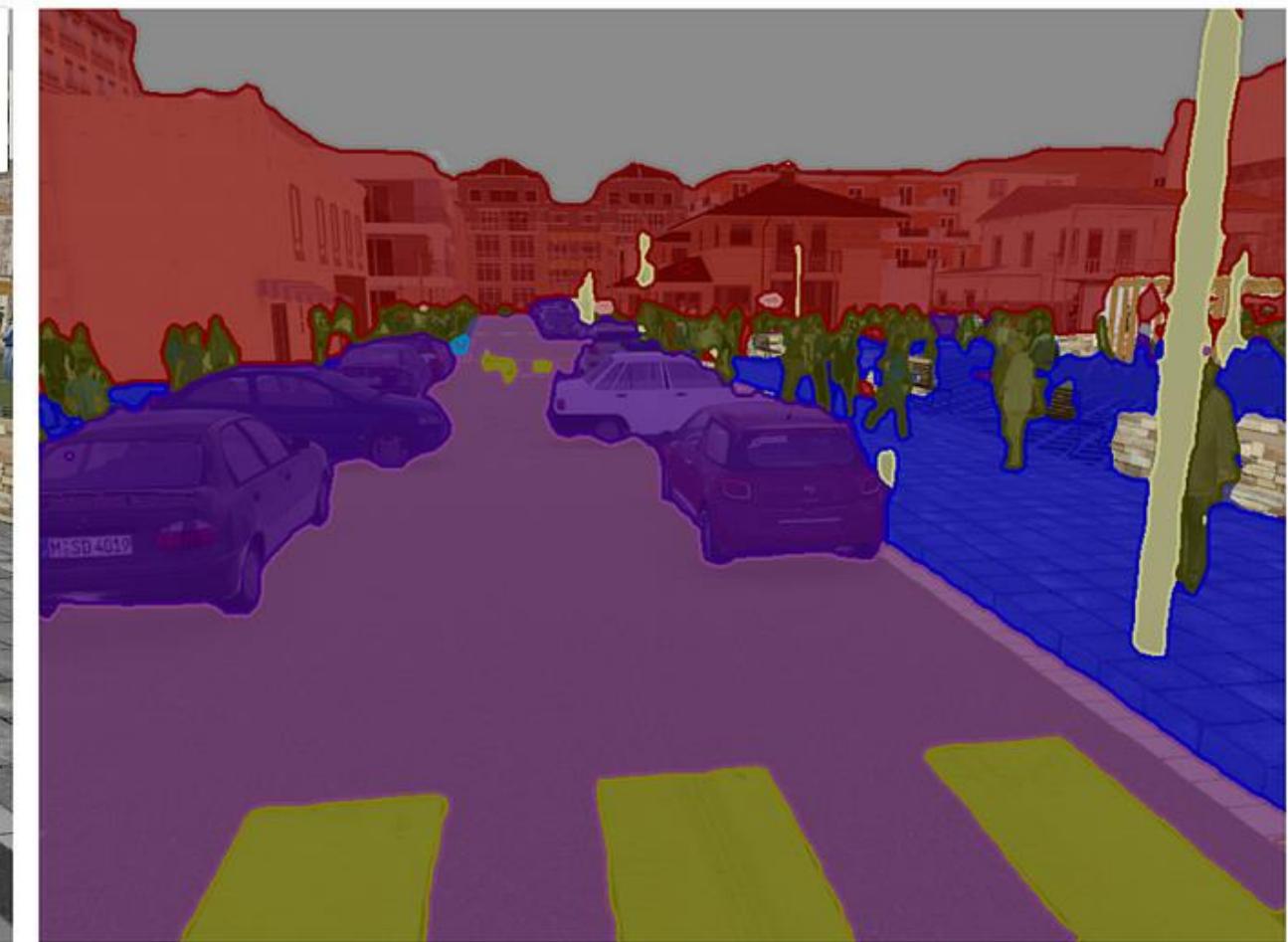
YOLO: YOU ONLY LOOK ONCE





Segmentation d'image

- La segmentation d'image est la tâche de diviser une image en plusieurs segments ou régions, chacun correspondant à un objet ou une partie de l'image différente.
- Les CNN sont couramment utilisés pour la segmentation d'images, en utilisant des techniques telles que les réseaux entièrement convolutifs (FCN), U-Net et Mask R-CNN.
- La segmentation d'images a de nombreuses applications, telles que l'analyse d'images médicales, l'imagerie par satellite et la conduite autonome.

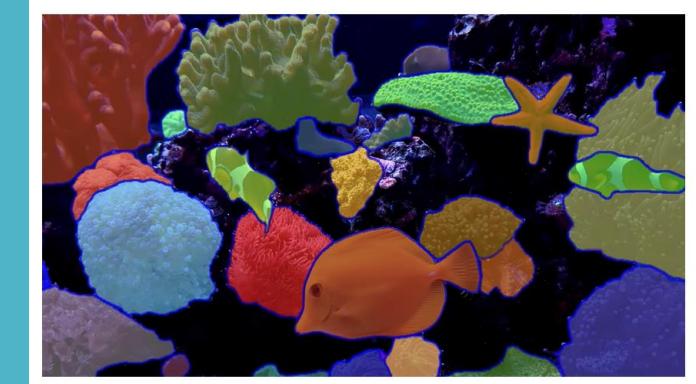
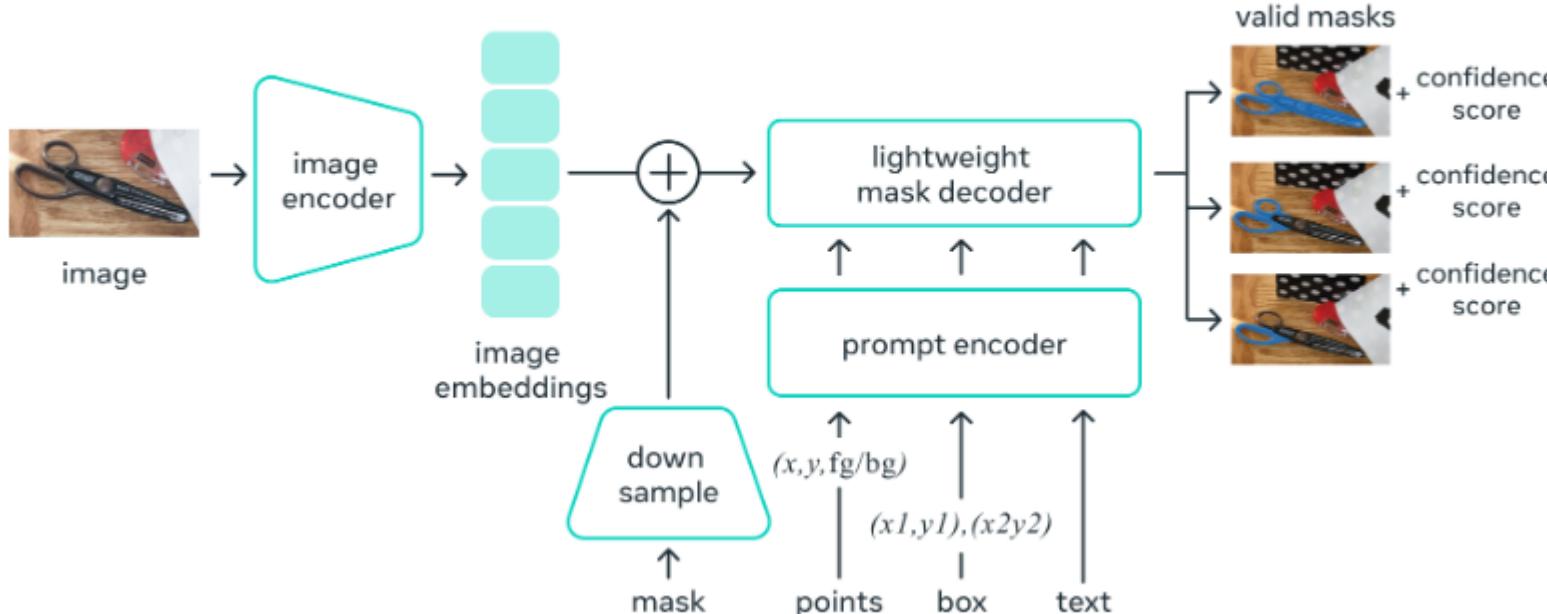


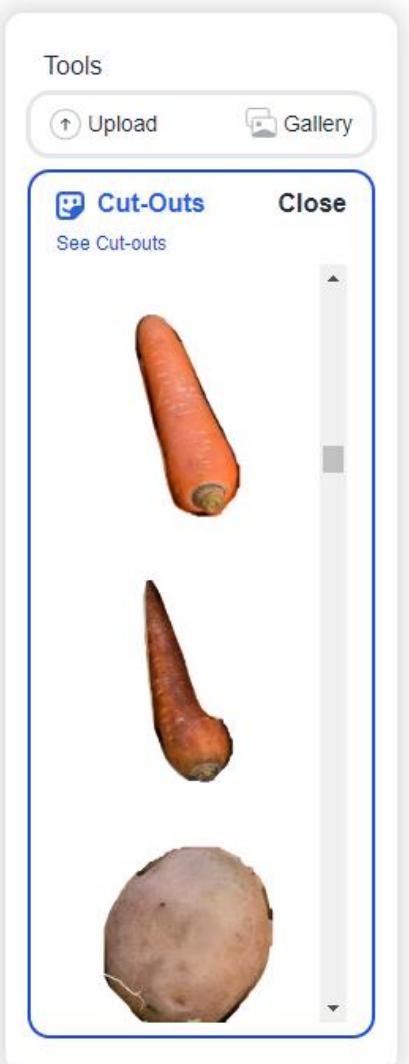
■ Sky ■ Building ■ Road ■ Sidewalk ■ Fence ■ Vegetation ■ Pole ■ Car ■ Sign ■ Pedestrian ■ Cyclist



<https://ai.facebook.com/blog/segment-anything-foundation-model-image-segmentation/>

Universal segmentation model





<https://segment-anything.com/demo#>

These are Digital Bridge Original and Proprietary Materials - Do
not use without explicit authorization

Traitements du langage naturel

- Natural language processing (NLP) is the task of processing and understanding human language, such as text or speech.
- CNNs are commonly used for NLP, using techniques such as text classification, sentiment analysis, and language translation.
- NLP has many applications, such as virtual assistants, chatbots, and automated customer service.

Analyse de séries chronologiques

- L'analyse de séries chronologiques est la tâche d'analyser et de prédire les tendances et les modèles dans les données variant dans le temps.
- Les CNN sont couramment utilisés pour l'analyse de séries chronologiques, en utilisant des techniques telles que les réseaux neuronaux convolutifs 1D et les modèles basés sur LSTM.
- L'analyse de séries chronologiques a de nombreuses applications, telles que les prévisions financières, l'analyse des données des capteurs et la surveillance de l'environnement.

Orientations futures des CNN

CNN profonds

- Les CNN profonds sont des CNN avec de nombreuses couches et il a été démontré qu'ils atteignent des performances de pointe sur de nombreuses tâches de vision par ordinateur.
- La tendance vers des CNN plus profonds devrait se poursuivre, avec le développement d'architectures et de techniques d'optimisation encore plus complexes.

CNN récurrents

- Les CNN récurrents (RCNN) sont une combinaison de CNN et de réseaux neuronaux récurrents (RNN) et sont conçus pour gérer des données séquentielles, telles que la vidéo ou l'audio.
- Il a été démontré que les RCNN atteignent des performances de pointe sur de nombreuses tâches d'analyse vidéo, telles que la reconnaissance d'action et le sous-titrage vidéo.

Mécanismes d'attention

- Les mécanismes d'attention sont une technique utilisée pour concentrer l'attention de CNN sur les parties les plus pertinentes des données d'entrée, plutôt que de traiter l'ensemble des données d'entrée en même temps.
- Il a été démontré que les mécanismes d'attention améliorent les performances des CNN sur de nombreuses tâches, telles que le sous-titrage d'images, la traduction automatique et la reconnaissance vocale.

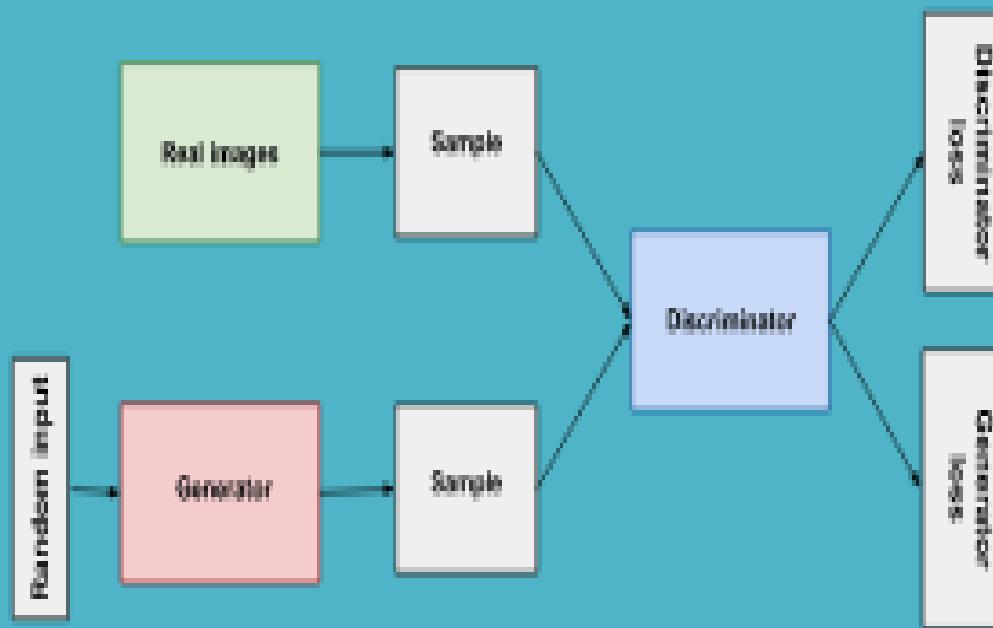
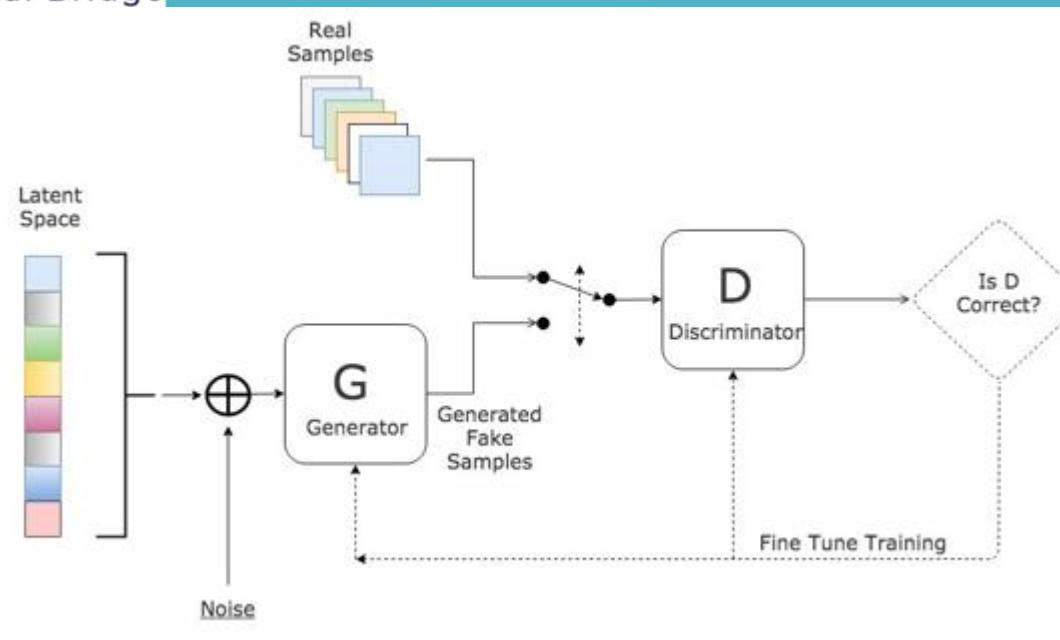
Modèles hybrides

- Les modèles hybrides sont des combinaisons de CNN avec d'autres types de réseaux neuronaux, tels que les réseaux antagonistes génératifs (GAN) ou les autoencodeurs.
- Il a été démontré que les modèles hybrides atteignent des performances de pointe sur de nombreuses tâches, telles que la génération d'images, le transfert de style et l'intégration d'images.
- On s'attend à ce que les modèles hybrides continuent d'être un domaine de recherche actif à l'avenir.

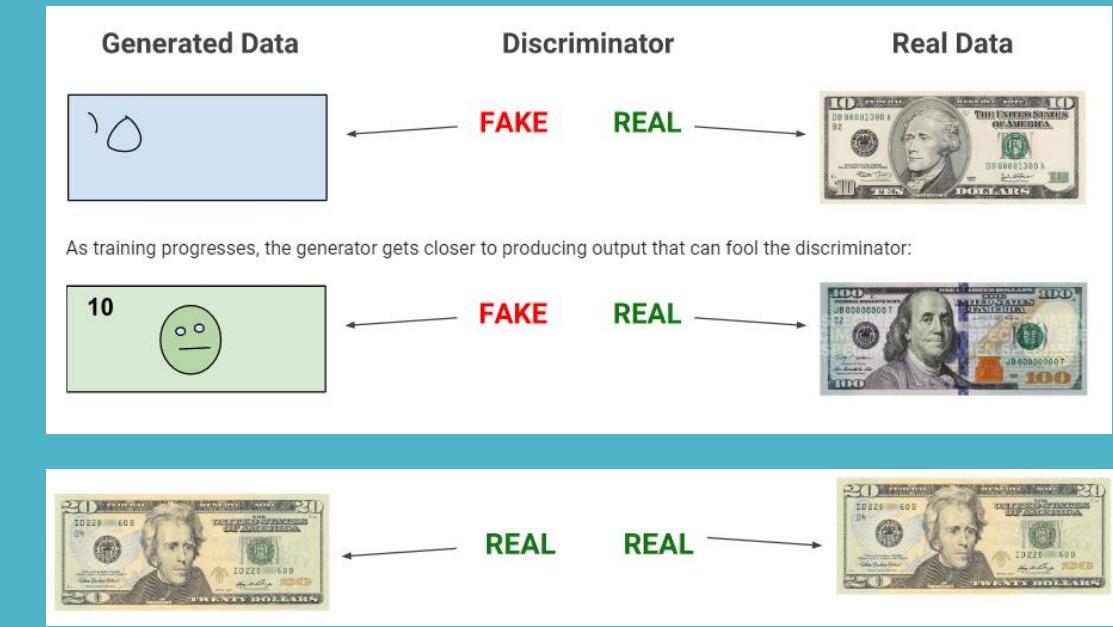
<https://paperswithcode.com/method/gan>

Generative Adversarial Networks (GAN)

- Le générateur crée de nouvelles données (par exemple, des images, de l'audio, du texte) qui imitent un ensemble de données d'entraînement particulier.
- Le discriminateur tente de faire la distinction entre les données générées et les données réelles de l'ensemble de données d'apprentissage.
- Le générateur et le discriminateur sont formés ensemble, le générateur essayant de créer des données qui peuvent tromper le discriminateur, et le discriminateur essayant d'identifier correctement si les données sont réelles ou générées.
- Grâce à ce processus itératif, le générateur apprend à créer des données de plus en plus réalistes, tandis que le discriminateur devient meilleur pour identifier les fausses données.
- Les GAN ont été utilisés pour une variété d'applications, telles que la synthèse d'images et de vidéos, la génération de texte et la génération de musique.
- Cependant, la formation des GAN peut être difficile, et ils peuvent être sujets à l'effondrement du mode (où le générateur produit une variation limitée de sa sortie) et à d'autres problèmes.



These are Digital Bridge Original and Proprietary Materials - Do not use without explicit authorization



Applications de GANs

- Synthèse d'images et de vidéos : les GAN peuvent générer des images et des vidéos de haute qualité qui semblent réalistes, ce qui les rend utiles pour des applications telles que le développement de jeux vidéo, la réalité virtuelle et l'infographie.
- Augmentation des données : les GAN peuvent être utilisés pour générer des données d'apprentissage supplémentaires pour les modèles d'apprentissage automatique, ce qui contribue à améliorer leur précision et leur robustesse.
- Détection d'anomalies : les GAN peuvent être utilisés pour apprendre la distribution normale d'un ensemble de données, puis pour détecter des anomalies ou des valeurs aberrantes.
- Génération de texte : les GAN peuvent être utilisés pour générer du texte en langage naturel qui imite un style ou un ton d'écriture particulier, ce qui les rend utiles pour des applications telles que les chatbots, la traduction linguistique et la génération de contenu.
- Transfert de style: Les GAN peuvent être utilisés pour transférer le style d'une image à une autre, en créant de nouvelles images avec un style similaire à l'image originale.
- Imagerie médicale : les GAN peuvent être utilisés pour générer des images médicales synthétiques qui peuvent être utilisées pour former et tester des modèles d'apprentissage automatique, ou pour générer des images avec des caractéristiques spécifiques.
- Mode et design: Les GAN peuvent être utilisés pour générer de nouveaux designs pour les vêtements, les meubles et autres produits, permettant aux concepteurs d'explorer de nouvelles idées et concepts.
- Génération de musique : Les GAN peuvent être utilisés pour générer de la nouvelle musique qui imite un genre ou un style particulier, ce qui les rend utiles pour des applications telles que la composition musicale et la conception sonore.

Generative Adversarial Networks (GANs)

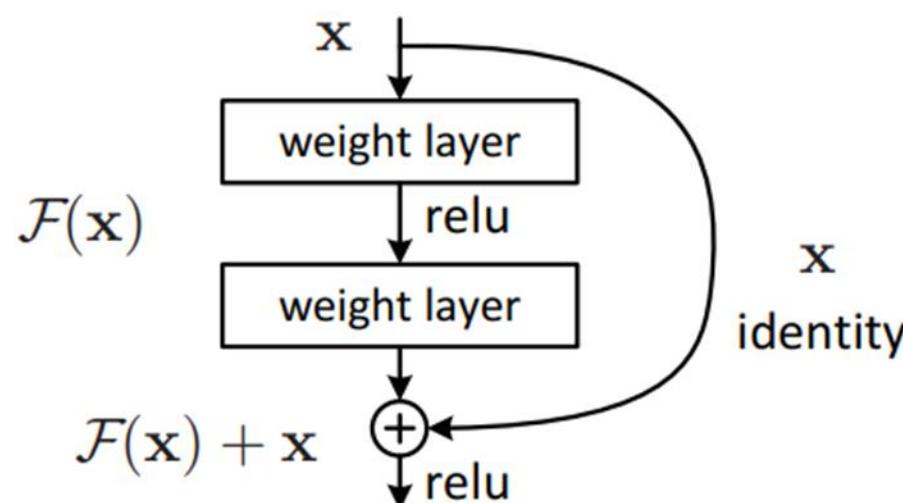


Ces personnes n'existent pas

ResNet

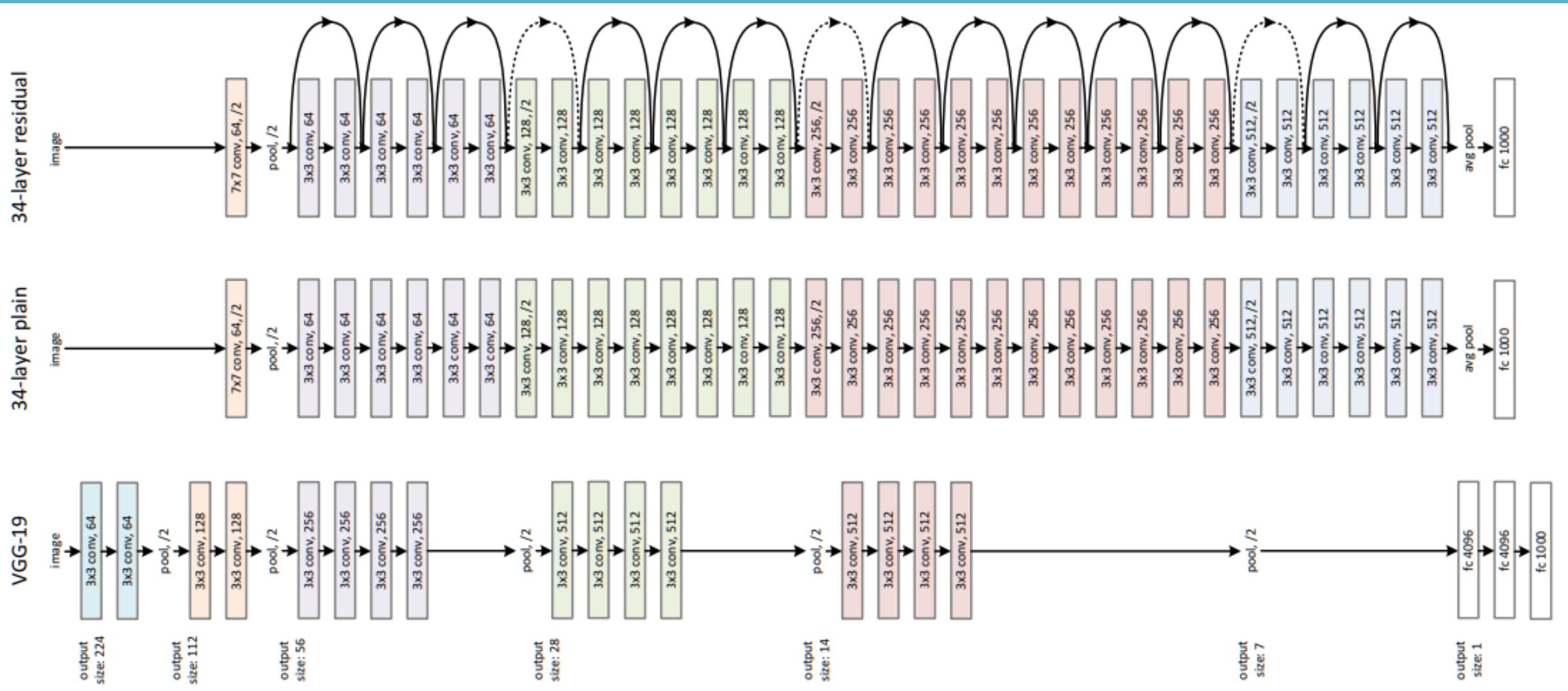
- ResNet, abréviation de Residual Network, est une architecture de réseau neuronal profond qui a été introduite pour la première fois par des chercheurs de Microsoft en 2015.
- Il s'agit d'un type de réseau neuronal convolutif (CNN) qui utilise des connexions résiduelles pour surmonter le problème des gradients de fuite dans les réseaux profonds.[**Vanishing gradients**]
- L'innovation clé de ResNet est l'utilisation de blocs résiduels, qui contiennent des connexions de saut qui permettent au réseau d'apprendre des fonctions résiduelles au lieu de fonctions complètes.
- Les connexions résiduelles permettent au réseau de contourner certaines couches et de propager directement l'information d'une couche à l'autre, ce qui contribue à réduire la profondeur du réseau et à faciliter la formation.[**Skip connections**]
- Les architectures ResNet ont établi de nombreux records dans les benchmarks de classification d'images, tels que l'ImageNet Large Scale Visual Recognition Challenge (ILSVRC), et ont été largement adoptées dans diverses applications de vision par ordinateur.
- Le document original de ResNet a présenté plusieurs variantes de l'architecture, notamment ResNet-18, ResNet-34, ResNet-50, ResNet-101 et ResNet-152, qui diffèrent par le nombre de couches et la complexité.
- ResNet a également inspiré le développement de nombreuses autres architectures de réseaux neuronaux profonds qui utilisent des connexions résiduelles ou des idées similaires, telles que DenseNet, ResNeXt et Wide ResNet.
- Bien que ResNet ait réussi dans de nombreuses tâches de vision par ordinateur, il s'agit toujours d'un modèle complexe et coûteux en calcul qui nécessite de grandes quantités de données et de ressources informatiques pour s'entraîner efficacement.
- Des recherches récentes se sont concentrées sur l'amélioration de l'efficacité et de l'évolutivité de ResNet, par exemple en utilisant des techniques d'élagage réseau, de quantification ou de distillation des connaissances.

ResNet



$$f(x) + x = h(x)$$

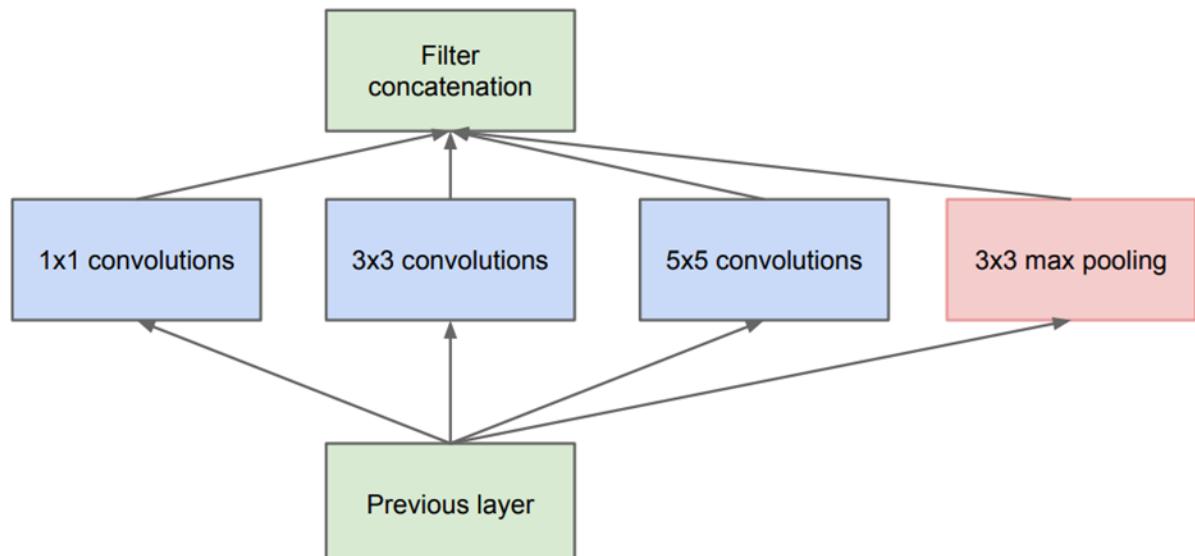
ResNet



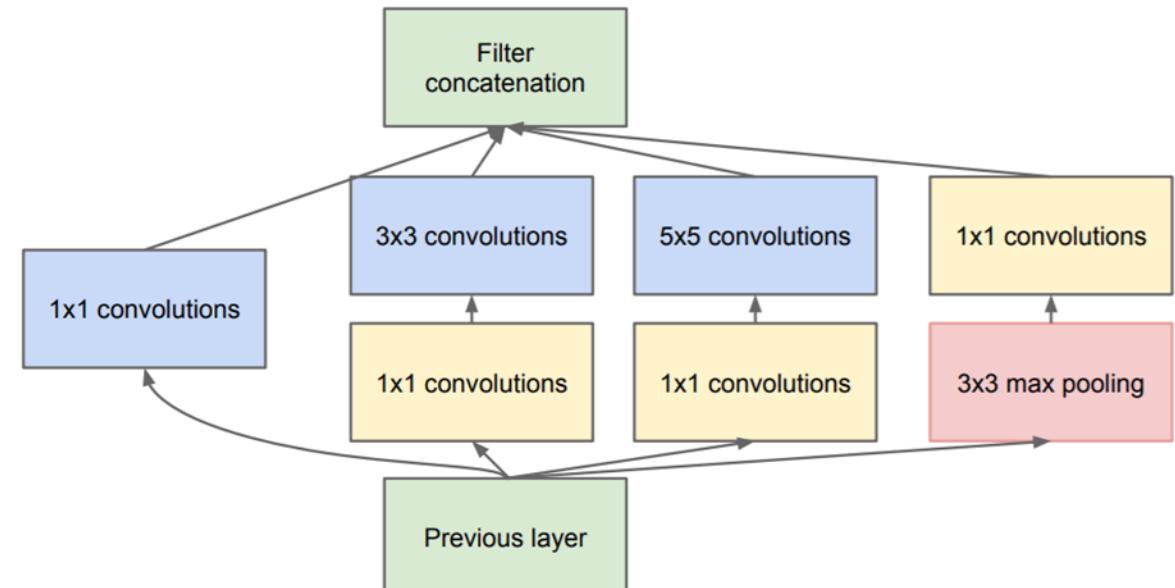
Inception

- Inception est une famille d'architectures de réseaux neuronaux introduite en 2014
- Inception utilise une combinaison de couches convolutives avec différentes tailles de noyau et des opérations de regroupement pour extraire des entités à plusieurs échelles
- Inception peut capturer des caractéristiques fines et grossières dans une image, utiles pour la reconnaissance et la détection d'objets
- Inception utilise des couches de goulot d'étranglement pour réduire les canaux d'entrée avant d'appliquer des circonvolutions plus coûteuses en calcul
- ResNet et Inception sont deux puissantes architectures de réseaux neuronaux utilisées dans la recherche en vision par ordinateur
- ResNet est mieux adapté aux réseaux très profonds
- Inception est meilleur pour capturer des fonctionnalités à plusieurs échelles
- Le choix entre les deux dépend des exigences spécifiques de la tâche à accomplir.

Inception



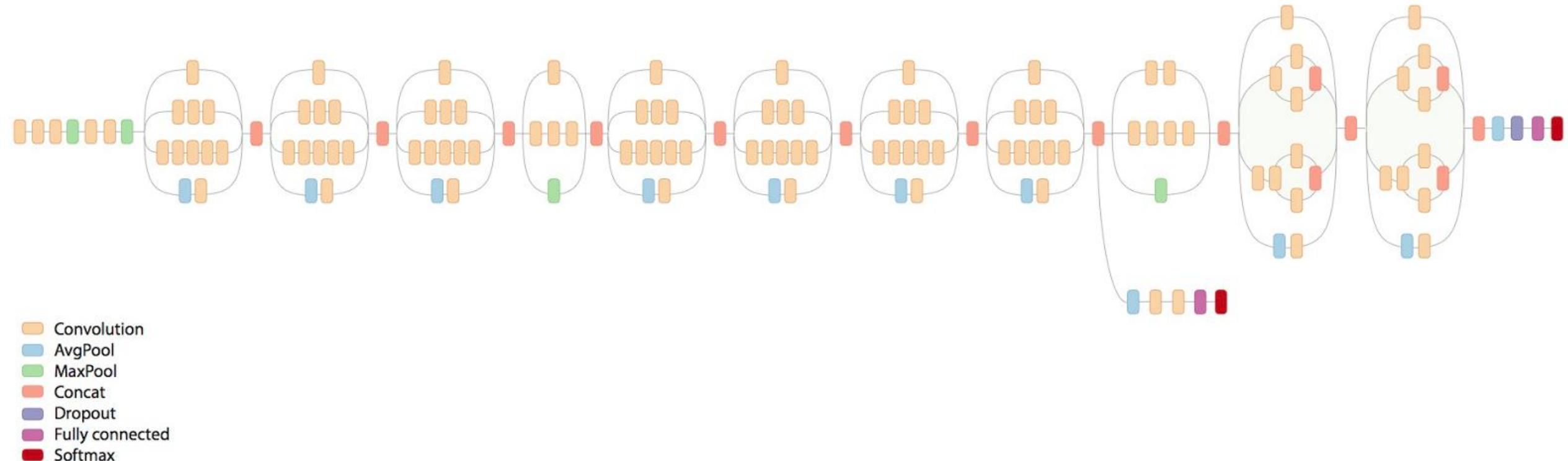
(a) Inception module, naïve version



(b) Inception module with dimension reductions

Inception module

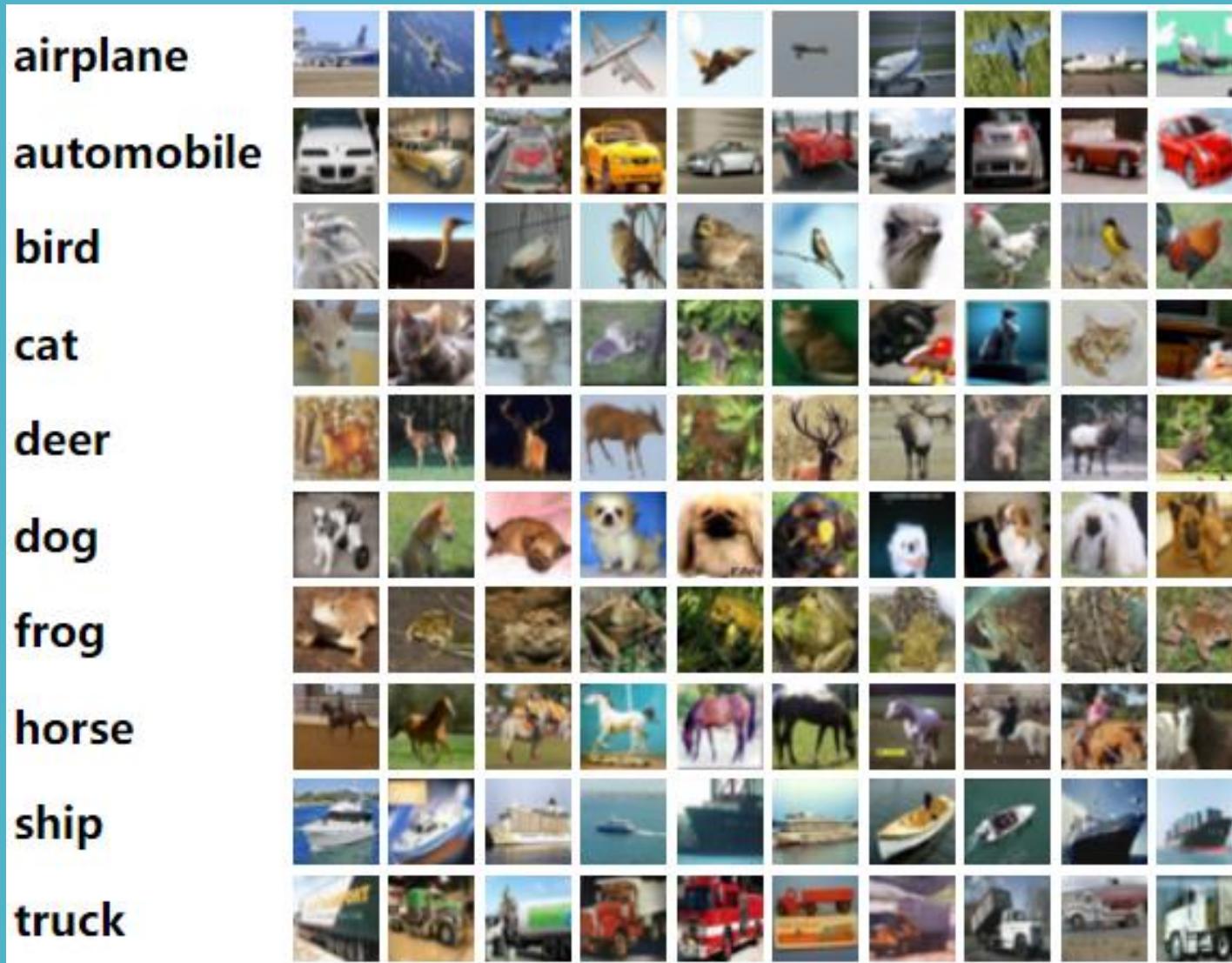
Inception



Ensemble de données CIFAR10

- CIFAR10 est un ensemble de données de 60 000 images couleur 32x32 réparties en 10 classes, à raison de 6 000 images par classe.
- L'ensemble de données est divisé en 50 000 images d'apprentissage et 10 000 images de test.
- Les classes comprennent: avion, automobile, oiseau, chat, cerf, chien, grenouille, cheval, bateau et camion.
- Les images de CIFAR10 sont de résolution relativement faible et ont été étiquetées manuellement.
- CIFAR10 est un ensemble de données de référence populaire pour les tâches de classification d'images dans la recherche en apprentissage automatique.
- En raison de sa taille relativement petite et de sa faible résolution, CIFAR10 est un bon ensemble de données pour expérimenter différentes architectures de réseaux neuronaux et techniques de formation.
- De nombreux modèles d'apprentissage profond à la fine pointe de la technologie, comme ResNet et DenseNet, ont été formés sur CIFAR10 et ont atteint une grande précision sur l'ensemble de tests.

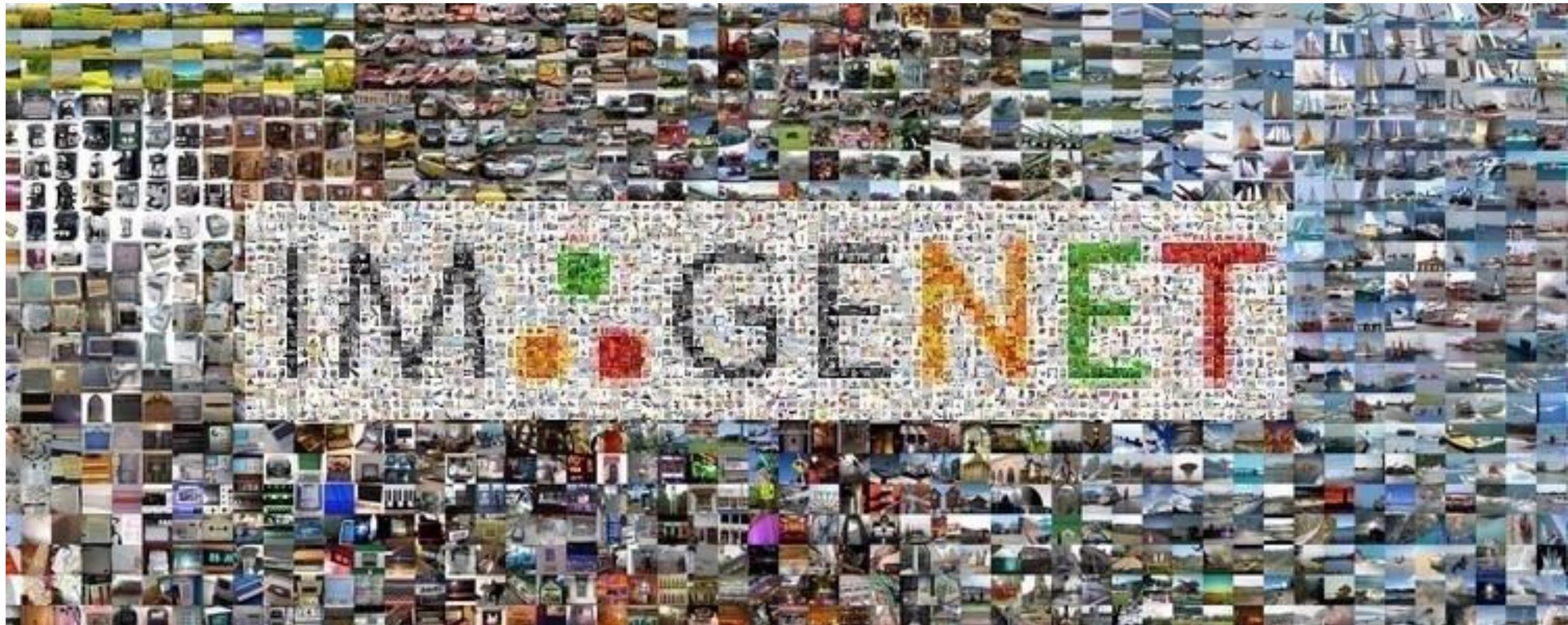
CIFAR10 classes



- CIFAR signifie Institut canadien de recherches avancées.
- Le CIFAR est un organisme de recherche mondial basé au Canada qui réunit des chercheurs interdisciplinaires pour travailler sur des défis scientifiques et sociaux complexes.
- L'ensemble de données CIFAR-10 a été créé par des chercheurs de l'ICRA pour des tâches de reconnaissance d'images.

ImageNet

- <https://image-net.org/>
- ImageNet est un défi de reconnaissance visuelle à grande échelle créé en 2009.
- L'ensemble de données comprend plus de 14 millions d'images, chacune annotée avec une ou plusieurs des plus de 20.000 catégories.
- Les catégories comprennent une grande variété d'objets, d'animaux et de scènes, ainsi que des concepts abstraits comme « amour » et « haine ».
- ImageNet a été créé pour promouvoir la recherche en vision par ordinateur et pour faire progresser l'état de l'art en matière de reconnaissance d'images.
- En particulier, ImageNet était l'ensemble de données utilisé dans le cadre du ImageNet Large Scale Visual Recognition Challenge (ILSVRC), qui était un concours visant à créer le meilleur modèle de reconnaissance d'images.
- Le concours ILSVRC a contribué à populariser l'apprentissage profond et les réseaux de neurones convolutifs, et a conduit à de nombreuses percées dans la reconnaissance d'images.
- Les meilleurs modèles du concours ILSVRC, tels que AlexNet, VGG et ResNet, ont pu atteindre une précision de reconnaissance supérieure à la performance humaine dans certains cas.
- ImageNet est encore utilisé aujourd'hui comme un ensemble de données de référence pour évaluer les performances de nouveaux modèles de reconnaissance d'images, et continue de stimuler les progrès dans le domaine de la vision par ordinateur.



These are Digital Bridge Original and Proprietary Materials - Do
not use without explicit authorization



not use without explicit authorization