**4. Partie III – Architectures avancées**

La troisième partie explore des modèles plus sophistiqués adaptés à des tâches complexes.

* **Réseaux convolutifs (CNN)** : adaptés au traitement d’images, ces réseaux exploitent la structure spatiale des données. Le livre détaille la façon dont les filtres explorent localement les pixels, comment le pooling réduit la résolution tout en préservant les caractéristiques essentielles, et comment des architectures telles que ResNet ou DenseNet permettent d’entraîner des réseaux très profonds.
* **Réseaux récurrents (RNN, LSTM, GRU)** : on s’intéresse aux données séquentielles comme le texte ou les signaux. Les mécanismes de mémoire et de portes (entrée, oubli, sortie) sont expliqués pour montrer comment le réseau peut retenir ou oublier des informations sur le long terme.
* **Mécanisme d’attention et Transformers** : cette section décrit la révolution introduite par l’attention, qui permet de modéliser directement les dépendances entre tous les éléments d’une séquence. L’architecture Transformer, basée exclusivement sur l’attention multi-têtes et sans connexions récurrentes, est décryptée, ainsi que les méthodes de pré-entraînement suivies d’un ajustement fin pour des tâches spécifiques.
* **Modèles génératifs** : le livre présente les autoencodeurs variationnels, qui apprennent à encoder et générer des données, puis les réseaux antagonistes génératifs (GAN), où deux réseaux s’affrontent pour produire des échantillons de plus en plus réalistes. Les modèles de diffusion récents, qui mènent à des performances de pointe en génération d’images, sont également abordés.