CR2 Projet Image 16

BELOT Mathieu, DELVIGNE Brian
Master 2 Imagine
Université de Montpellier

October 27, 2024

1 Introduction

Le but de ce second compte rendu est de continuer l'état de l'art sur notre sujet
Le prétraitement des séquences d'images est une étape cruciale dans divers domaines comme la vision par ordinateur, l'analyse vidéo, et les systèmes d'intelligence artificielle. Il vise à améliorer la qualité des données et à préparer les séquences pour une analyse plus efficace. Voici un état de l'art des techniques de prétraitement des séquences d'images :

2 Méthodes basées sur les réseaux de neurones convolutifs

2.1 Les transformeurs

Un transformer est une architecture d'apprentissage profond principalement utilisé dans l'apprentissage de la langue mais peuvent aussi servir à traiter d'autres modalités comme les images. Ils marchent selon deux grands principes.

L'attention : ce mécanisme sert à attribuer une pertinance ou un poids aux pixels de l'image. Autrement dit cela permet de localiser les secteurs qui vont nous intéresser dans la suite du traitement de l'image et de ce concentrer sur les parties les plus importantes de la séquence.

Le self attention (ou auto attention) : celui-là mets les éléments en relation les uns avec les autres. L'objectif est de calculer pour chaque éléments une représentations pondérée des autres éléments. Ce principe est pertinent car les pixels sont souvents en relation et dépendants de leurs voisins.

le Vision Transformer (ViT) est un transformer particulier qui fonctionne en divisant une image en patches, puis en traitant ces patches comme des séquences de données à l'image d'une phrase et de mots comme pour un transformer classique. Chaque patch est représenté par un vecteur, puis les paires de vecteurs sont évaluées pour leurs relations grâce au mécanisme de self attention.

2.2 Le système YOLO

YOLO (You Only Look Once) est une famille de modèles spécialisée dans la détection d'objets en temps réel, conçue pour identifier rapidement et avec précision plusieurs objets dans des images ou des vidéos. A l'origine il a été crée pour la détection, mais se montre efficace pour la suivie d'objets dans des séquences vidéo lorsqu'il est associé aux bons algorithmes. Ce système CNN divise l'image en une grille de cellules au lieu de régions d'intérêts qui vont servir à prédire des scores d'appartenance en fonction de classes d'objets pré-déterminés. YOLO est appliqué en temps réel et très rapide, il est donc efficace et qualifié dans le domaine de la reconnaissance sur des séquences d'images, dû au fait qu'il génère toutes les prédictions en simultané.

3 Bibliographie

- https://fr.wikipedia.org/wiki/Transformeur
- https://larevueia.fr/introduction-aux-reseaux-de-neurones-transformers/
- https://datascientest.com/you-only-look-once-tout-savoir