百川學士學位學程大一下專題探索

利用Detr建立辨識3D物體模型

孫奇霆

摘要

本專題將透過COCO數據集，利用DETR建立一個3D物體辨識模型。

前言

傳統的物體檢測流程是：(1)輸入圖像，並切分成不同的區域；(2)將各區域傳遞給卷積神網絡(CNN)，並將區域分類；(3)一旦將每個區域劃分為相對應的類別後，就可以組合這些區域，來檢測原始圖像。

但這種方法的缺點是：圖像中的對象具有不同的寬高比和空間位置，譬如在某些情況下，對象可能覆蓋了大部份的圖像，而其他對象可能只覆蓋圖像的一小部份。而且對象的形狀也可能不同。另外，劃分大量的區域也會花費大量的計算時間。

RCNN仍然延續傳統物體檢測的思想，先將物體分類，即先提取一系列的候選區，然後對候選區域進行分類。

後來Fast RCNN也是類似的作法，只是速度上加速非常多。

動機

由於先前的物體檢測模型，以卷積神經網路為主，缺點是耗時。而Facebook AI於2020年提出的DETR（Detection Transformer）則是一種基於Transformer架構的新穎物體檢測模型，它採用了一種全新的檢測機制，並利用了Transformer的自注意力機制來捕捉圖像中不同位置間的相互關係。

以下是DETR模型的一些主要特點和組件(引自iT邦幫忙)：

1. End-to-End訓練：DETR是一個端到端訓練的模型，它不依賴於傳統的區域提議網絡（RPN）和非極大值抑制（NMS）等手段，從而簡化了物體檢測流程。
2. Transformer架構：該架構最初被設計用於自然語言處理（NLP）任務，但在DETR中得到了成功的應用。Transformer幫助模型理解圖像中的上下文信息，並處理物體之間的關係。
3. Bipartite Matching Loss：DETR引入了一種名為雙邊匹配損失（Bipartite Matching Loss）的新型損失函數，用於在訓練過程中匹配預測邊界框和真實邊界框。這種方法解決了物體檢測中的賦值問題。
4. 物體和背景的區分：DETR使用一種稱為「無物體類別」（"no object" class）的機制來區分圖像中的物體和背景，這有助於模型更準確地進行物體檢測。
5. 多尺度特徵表示：DETR利用卷積神經網絡（CNN）提取多尺度的特徵表示，並將這些特徵融合到Transformer中，以捕捉不同大小和形狀的物體。
6. 全局理解：通過自注意力機制，DETR能夠全局地理解圖像，這使得模型能夠處理一些複雜的場景和遮擋情況。

DETR提供了一種新穎且統一的檢測架構。由於其獨特的設計和優異的性能，DETR已成為計算機視覺領域的重要研究對象。

目標

•

•

•

環境設定

•視覺設備：

•程式語言：

•測試物體：

•運動方式：

研究工具

本專題會運用COCO數據集，做為模型建立的測資。COCO數據集是一個由Microsoft、Facebook、CVDF 及 Mighty Ai等組織所提供的一個大型開源圖片數據集，可以用來作圖像識別、分割、標題的圖像數據集，官方網址：<http://cocodataset.org>

COCO數據集可以用於CV 領域的各類研究，如：Detection, Segmentation, Keypoints……，有33萬張以上的影像（其中超過20萬張影像已標記），還包含150萬個物件、並分成80個類別（for object detection）以及91類的stuff（for semantic scene labeling），也正因為它的強大，所以現在CV領域的主流研究幾乎都會採用這份資料集來測試及驗證他們的方法。

COCO數據集的官網同時也提供一組API(一般都稱之為COCO API)來協助加載，解析和可視化COCO中的標註(annotations)。API支持對象實例(object instance)，對象關鍵點(object keypoints)和圖像標題(image caption)標註(annotation)資訊的提取。

研究方法

本專題的運作流程與方法如下：

1.。

目前進度

了解 如何進行 ，以及理解 的原理。

遇到的困難

。

參考文獻

作者不詳。DETR(Detection with Transformer)。iIT邦幫忙。<https://ithelp.ithome.com.tw/m/articles/10327551>

Jason Chen（2020）。[從COCO Dataset中提取所需的類別資料](https://jason-chen-1992.weebly.com/home/coco-dataset)，https://jason-chen-1992.weebly.com/home/coco-dataset