

多媒體內容分析 HW1 賴廷璋 F44054045

程式執行環境

使用 Colab 環境執行，為 ipynb 檔

使用套件：

CV2

Matplotlib

Numpy

Sklearn

tensorflow

ffmpeg-python

Edge Change Ratio

Visual Features

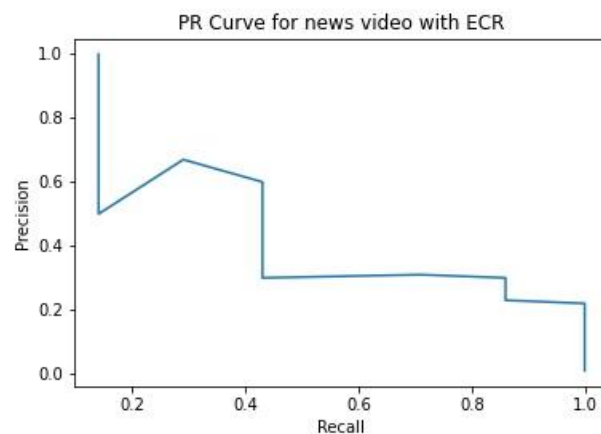
使用的 feature 為 frame 的邊緣，利用 cv2.Canny 並設定門檻值 0 與 200。其中利用 cv2.dilate 將 edge 放大膨脹，使變化較容易偵測。

演算法

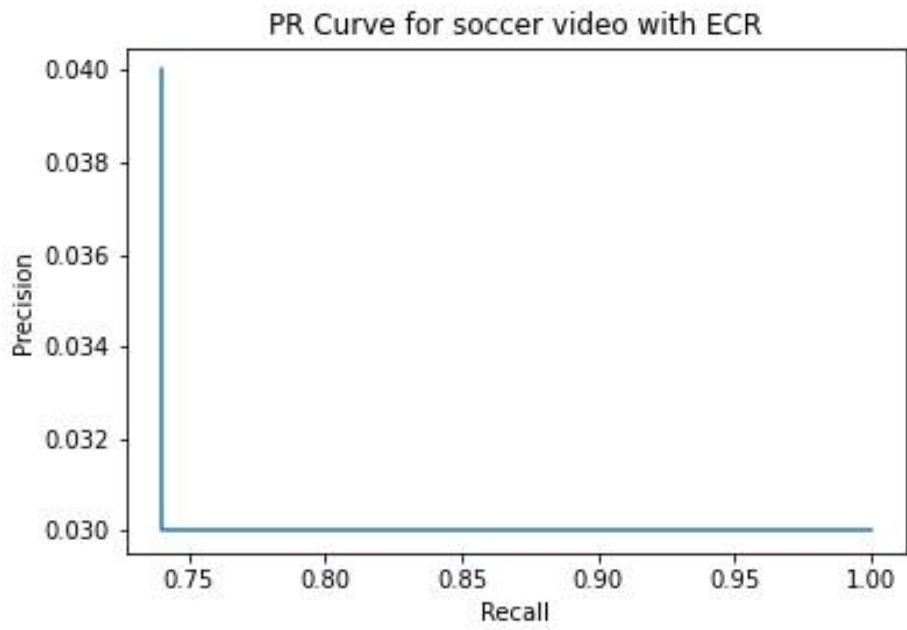
利用上下連續兩個 frame 的邊緣 pixel 的 entering edge 與 exiting edge 來決定是否畫面有顯著的變動。並設定 Threshold，若變動程度大於閾值則視為 shot change。

效能

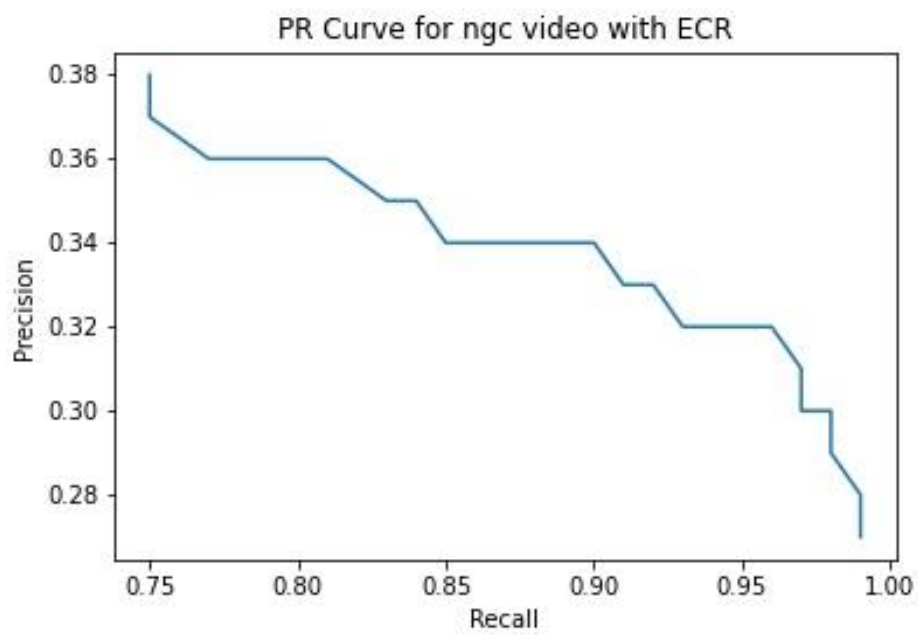
News



Soccer



NGC



Twin Comparison Approach based on Histogram

Visual Features

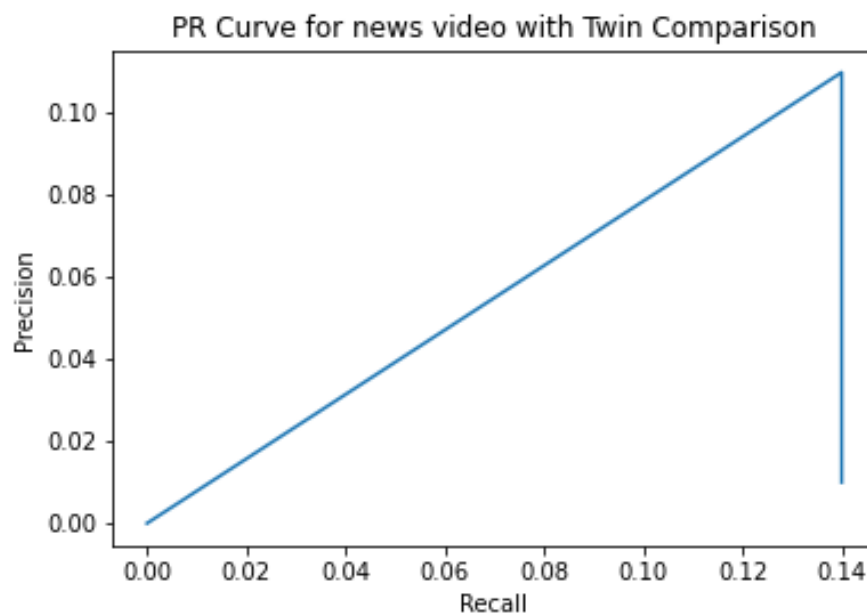
Feature 爲 frame 的 HSV Color Space，一個 frame 的特徵共有 1×256 維，分別由 H、S、V 三個維度加權(各爲 1×256 維)而成，權重分別爲 0.5, 0.3, 0.2。

演算法

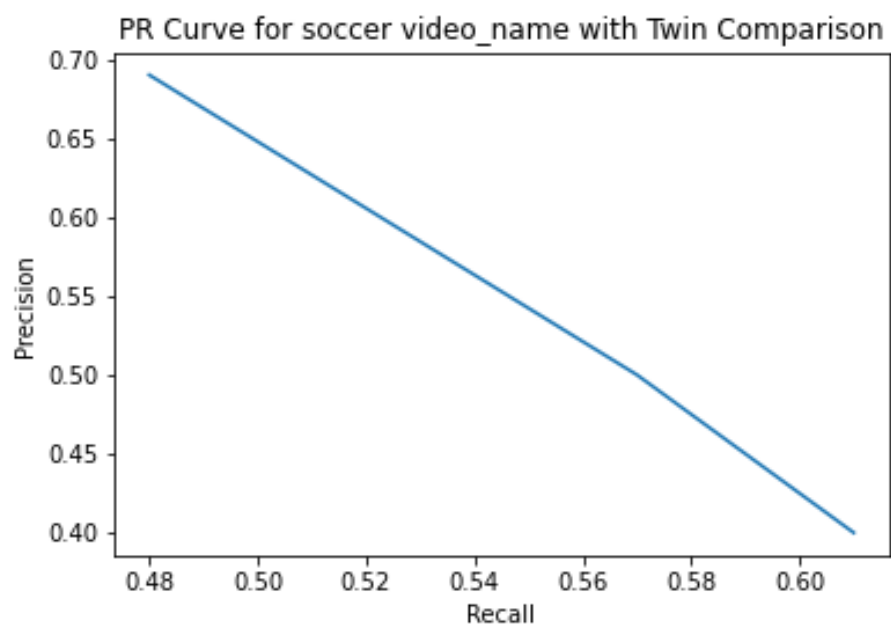
設定兩閾值分別爲 T_b 與 T_s ($T_b > T_s$)，並計算上下兩個 frame 之間在 256 維之間差的總和，若上下兩 frame 的差超過 T_b ，則視爲 hard cut，而若連續差超過 T_s 的 frames 的差總和超過 T_b ，則視爲 Dissolve。

效能

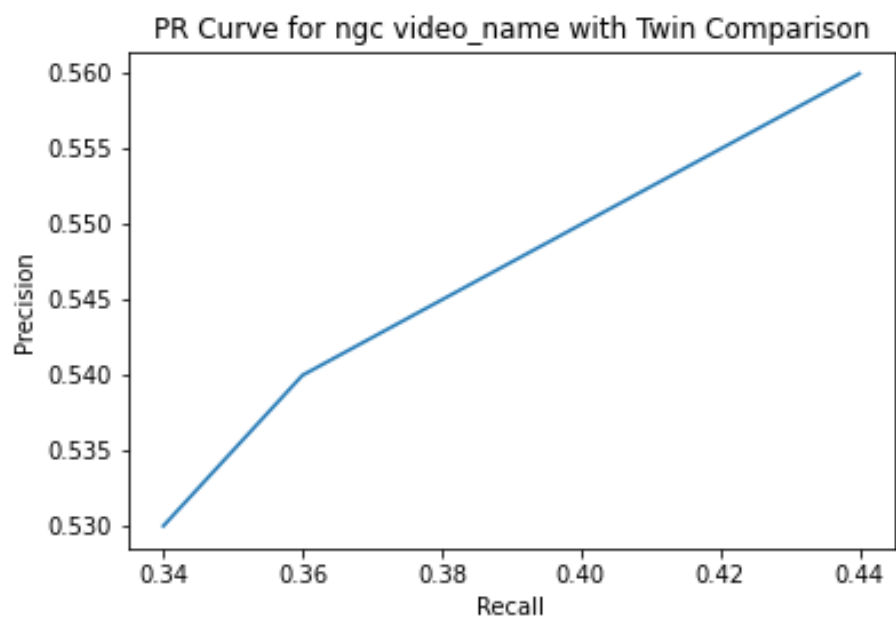
News



Soccer



Ngc



TransNet V2

Visual Features

此深度模型參考論文：[TransNet V2: An effective deep network architecture for fast shot transition detection](#)

此模型為 TransNet 的改進，TransNet 利用 CNN 卷積層的架構萃取圖片特徵，透過 4 個 Dilated CNN，隨之將其 concat 後，在透過 Max Pooling，即為 frame 的 feature。而 TransNet V2 的部分，加入了 RGB color histogram 和卷積層輸出的特徵作為特徵。

演算法

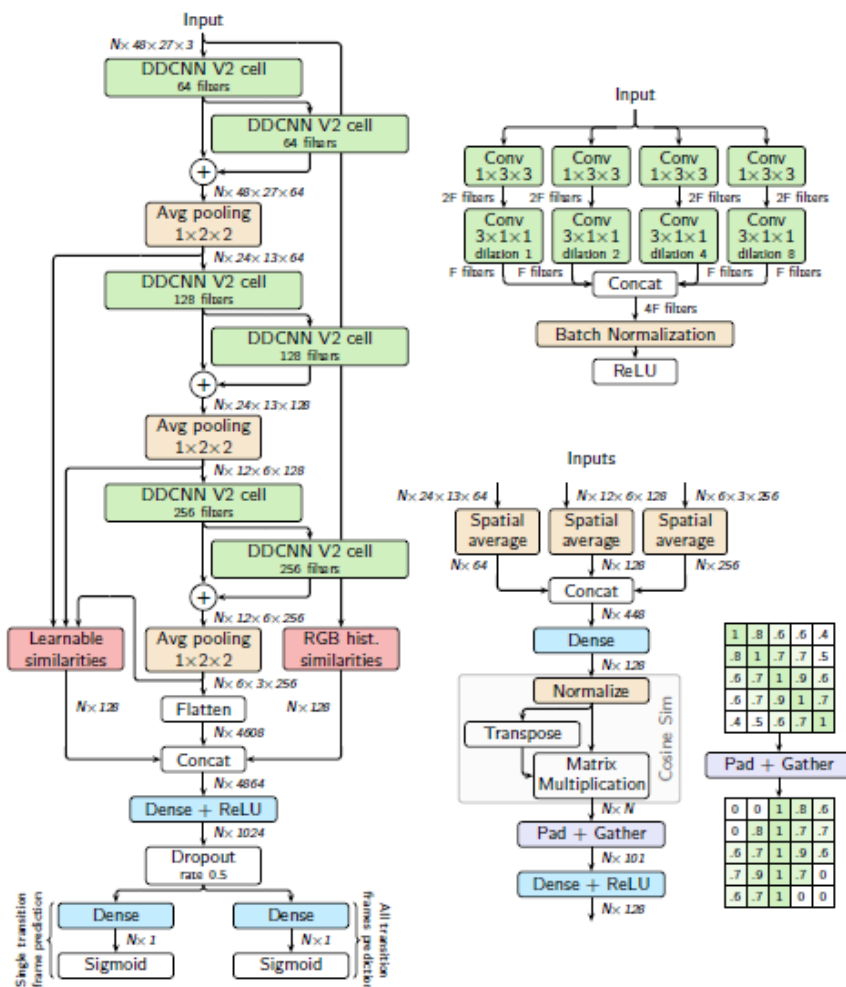
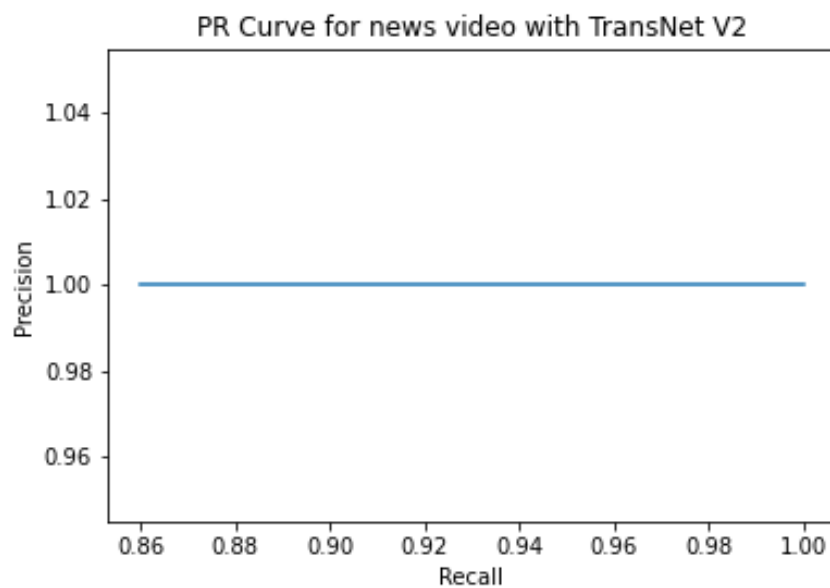


圖 1 論文提出的模型架構

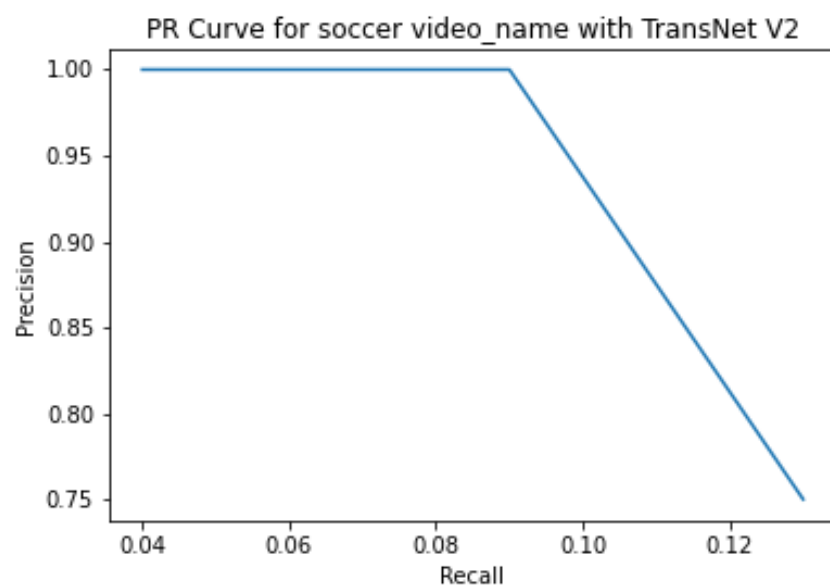
將 DDCNN V2 cell 定義為四個 3x3x3 的 CNN，分別的 dilation rate 為 1, 2, 4, 8，並加入 Batch Normalization 增加穩定性。與 RGB color space feature 結合後，再利用雙頭的連結層輸出結果，其中一層的 head 僅用於訓練時調整參數，以便學習 shot change 以及 shot change 的時間影響。

效能

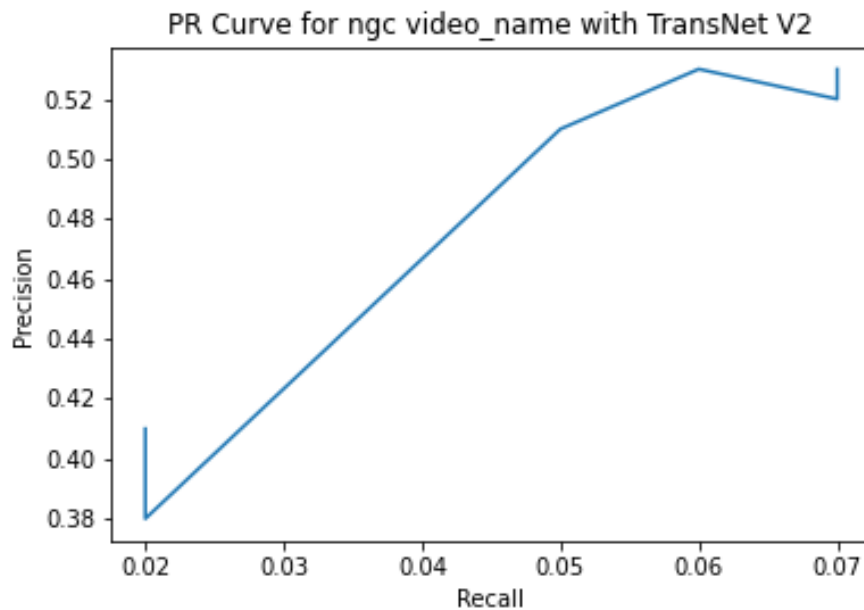
News



Soccer



Ngc



總結

對於 News 影片，利用預訓練之 TransNet V2 可以達到完美的 shot change，使得 recall 和 precision 皆為 1，其次 ECR 有著平均的表現，可以看到 Twin Comparison 的表現堪憂。

在 Soccer 的影片，因其 shot change 片段皆為 dissolve，故利用 Twin Comparison 顯然有最好的表現，而 ECR 和 TransNet V2 表現則不佳。

而最難的 NGC 影片，其參雜了 hard cut 和 dissolve 的畫面切換，利用 Twin Comparison 有較平均的表現(Recall 和 Precision 接近 0.5)，而 ECR 和 TransNet V2 則有較極端的表現(Recall 和 Precision 其中很高另外一個很低)。