**作業三：視訊處理與顯著性偵測實習**

0611031 資工10 謝至恆

1. **原理介紹**

**接續作業二的原理(我也附上作業二的報告)**

**這次要做的是關於影片中連續影像作顯著性偵測的判斷，**

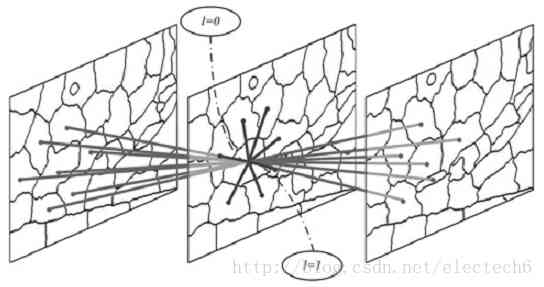
**主要重點應該是在影片中不同的網格中如何快速並準確地找到顯著物，以及因為影片中顯著物應該會有某種相似的行為，我們可以根據這一行為來減少演算的複雜度，並增加準確度。**

**新的東西**

**Supervoxel:**

接續著superpixel而來的運用**，**在視頻前景分割中用，因為相比像素，超像素處理速度會快幾十倍、幾百倍甚至更高，這對實時性要求較高的視頻分割比較重要，還有最近提出的supervoxel概念（可以認為是三維的superpixel）

應用在RGBD相機獲取的點雲數據中，通過使用基於三維空間的播種的方法和使用顏色和幾何特徵的約束來實現點雲的局部約束與聚類。**文章還清楚的說明了點雲的超體分割與“超體素”沒有關係，超體素是二維算法簡單的擴展到三維，這種方法是不能應用在三維的無序的空間中的，只能在有規則的體素中起作用。**

****

**概念圖**

****

**參考論文**

**(CVPR19) See More, Know More: Unsupervised Video Object Segmentation with Co-Attention Siamese Networks**

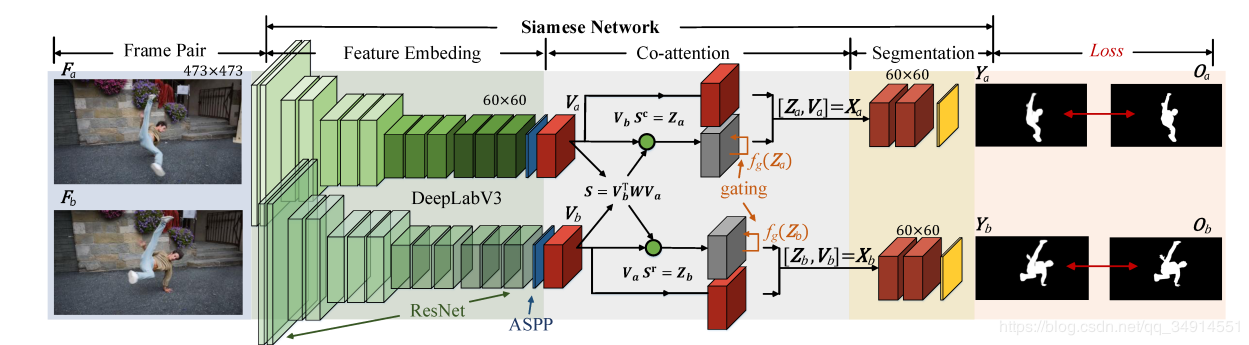
# idea

作者提出一種co-attention，基於一個視頻序列全局角度，來提升UVOS的精度。（確實領先目前的很多模型，davis官網的數據）。以往的一些方法，有通過顯著性檢測得到所要分割的目標，或者通過有限幀之間計算出的光流信息。COSNet則從整個視頻序列中考慮哪個目標是需要分割的。在測試階段，COSNet會綜合所有前面的幀得到的信息，推理出當前幀中哪個目標是顯著的同時還是經常出現的。Co-attention模塊挖掘了視頻幀之間豐富的上下文信息。基於co-attention，作者提出了COSNet（co attention Siamese）來從一個全局視角建模UVOS 。現在可能讀者還是不能理解這個全局視角是什麼，在method部分會解釋。

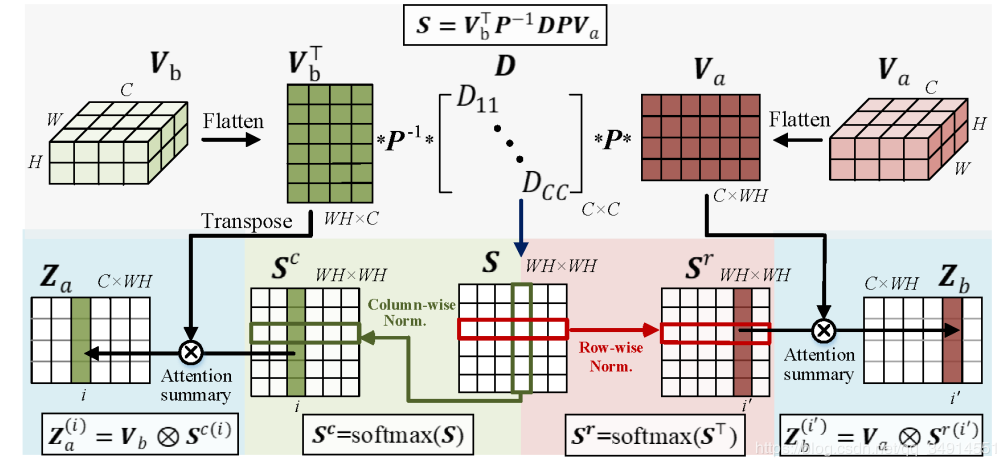
**main contribution**

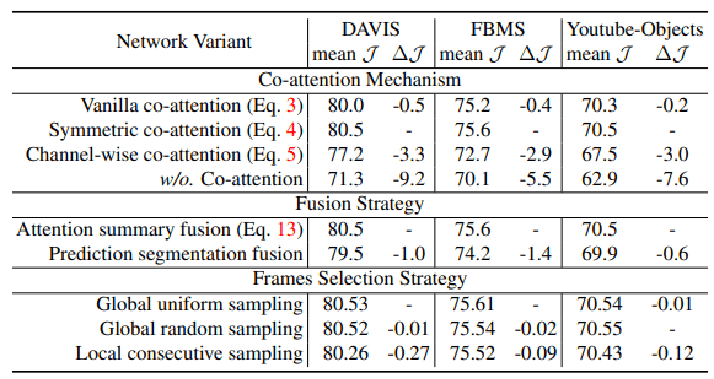
* COSNet採用的訓練方式是考慮一個pair，包含相同視頻中的任意兩幀，所以說極大的增加了數據量，不需要考慮時序關係，依次送入數據，而是可以打亂數據，隨機組合。
* 顯示建模幀和幀的聯繫，不依賴光流
* 它對於**增加訓練數據**非常有效。它允許在同一視頻內使用大量的任意成對幀進行訓練
* 在訓練過程中，可以看作是在基於同一視頻中任意幀對之間的相關性進行學習。這也可以認為是一種**上下文信息的學習**
* 在測試過程中，網絡以全局視圖推理主要目標，即**利用了測試幀和多個參考幀之間的co-attention information**

## training stage



**介紹作者的co-attention.**

****

**實際效果**

這部分對比試驗反映了幾點關鍵內容：

* 相似矩陣的三種不同設定的效果差異，可以看到，其中Symmetric co-attention效果最好
* 特徵層面上的平均策略（Attention summary fusion）更有效
* 全局採樣確實有著性能的提升

內容其實我不是完全都能看懂，部分數學運算的地方就不是很了解，所以我就沒有做。

**但我發現論文中將影片中任意兩幀當作一個集合，不同與其他只使用相鄰兩張為資料，在一影片中有著相同顯著物，其實每張變化不算太多，如果把全部顯著物任兩張拿來做一組資料來輸入，大大增加了測資的訓練量，或許可以讓訓練更為精準。這部分好像很重要，所以我這次是利用這個概念來修改我的神經網路架構。並且將原本的架構增加了兩層，希望可以得到更精準的模型。**

**參考文件:**

<https://github.com/carrierlxk/COSNet>

<https://www.twblogs.net/a/5d838fc5bd9eee532700004c>

<https://www.codenong.com/cs105859057/>

**(二)實作方法**

**-labelme**

**-深度學習網路**

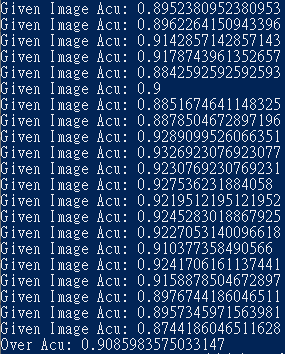
**與hw2大致相同**

**先做superpixel**

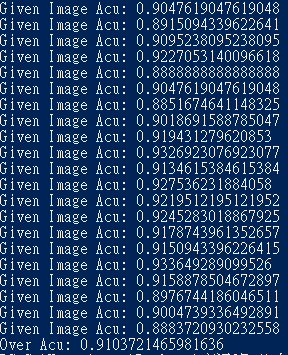
**將hw2的superCNN修改後套入TRAINING 和 TESTING**

1. **深度學習模型訓練/評估結果**

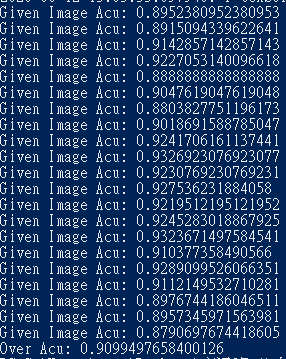
Epoch 10



Epoch 15



Epoch 20



準確度大概在91%上下浮動，應該是因為Epoch還沒到達收斂，或許有時間可以讓epoch跑多一點觀察結果。

不過相對於HW2準確度有所提升，85%變成91%

推測是因為影片中顯著物比較類似，而且有利用新的概念。所以準確度有提高一些。

**心得:**

這次作業對我來說非常困難，不知道其他人怎麼樣，因為hw2神經網路的構造跟這次論文的構造相差非常多，上次看得懂是因為助教給的程式碼算是寫得相當直白，慢慢看下去就會看得懂，但這次的實際去看論文，也查了許多有中文的文獻(英文不太好)，但還是有許多觀念不懂，只能大概了解是每個演算法的特點和造成結果不同的部分，但程式的部分就真的沒辦法了，看了論文的code發現裡面很多都是直接引用我不知道是什麼的參數或名稱，所以我看起來就不懂他在幹什麼，最後只能稍微修改hw2的架構來完成這次作業，之後有時間應該會去問其他同學或是再繼續找資料，如果有完成更好的架構再補交。

**總字數1592**