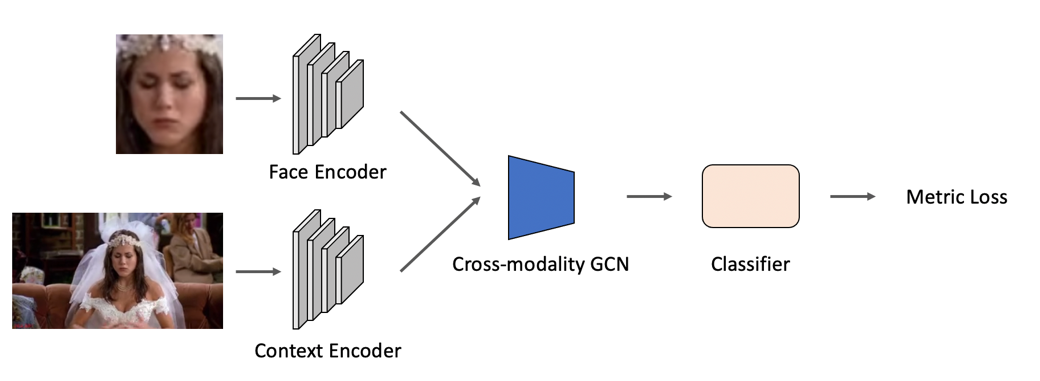
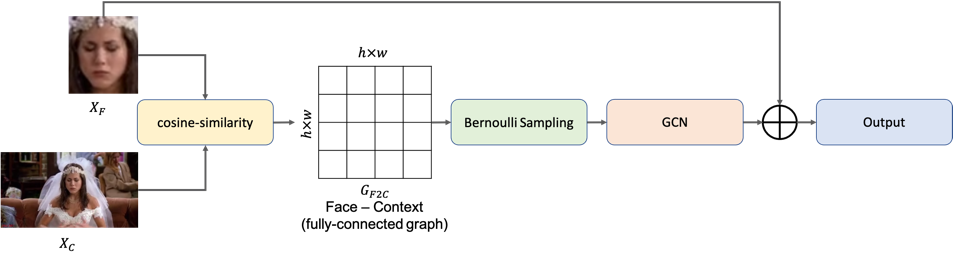
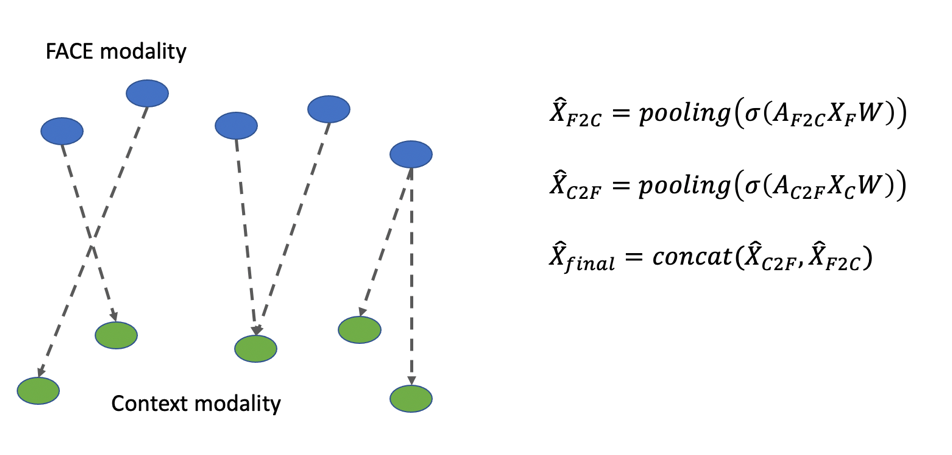
Emotion recognition:



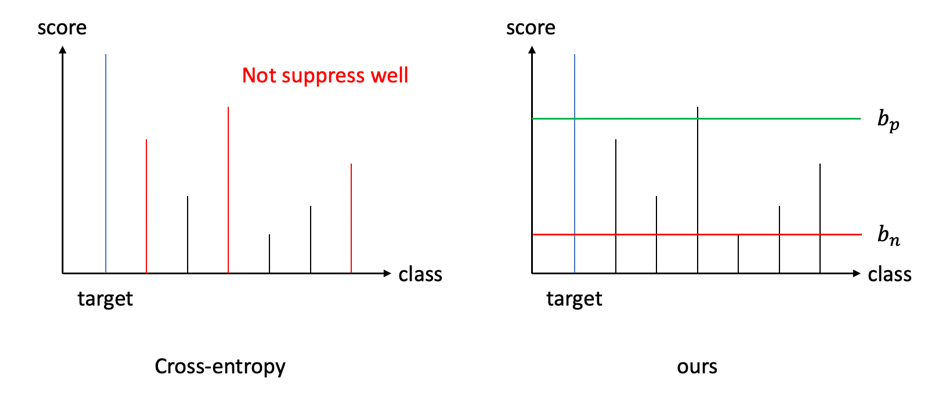
* Cross-modality graph convolutional network
  + 目的：有效的整合不同種類的特徵（face and context）
  + 方法：

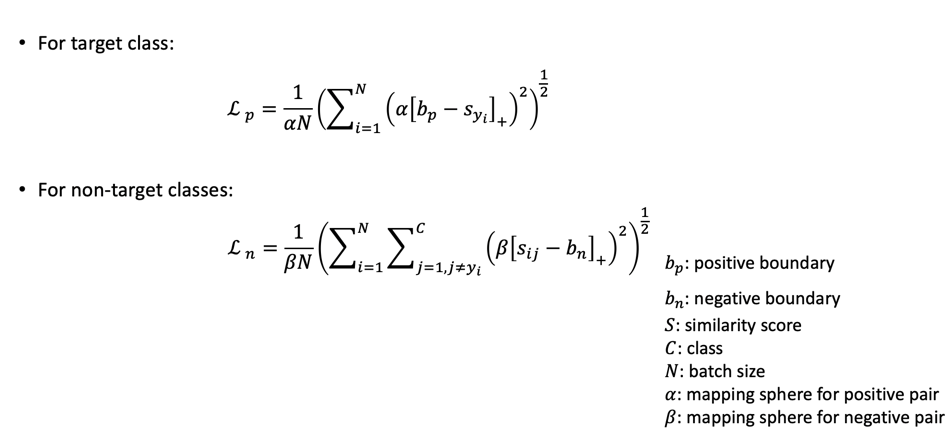




在這裏我們首先利用兩個embedding將原本cnn抽取出來的特徵圖（feature map）進行降維，接著去計算彼此之間的相似度（cosine-similarity），即cross-modality graph 。我們發現到face與context相互有關聯的位置只有少部分的區域，有許多的區域都是低相關的。即低相關性的元素的數量會遠大於高相關性的元素，因此我們利用bernoulli sampling的機制讓原本全聯接的圖 轉為稀疏的圖以更好地進行學習。由於bernoulli sampling可以有效保留高相似度的元素，並大幅捨棄低相似度的部分，所以能夠更好的建構稀疏的圖來進行學習。最後我們透過gcn將生成圖特徵並與原本的特徵進行整合並輸出。

* Metric Loss
  + 目的：更好的規範特徵的分佈





* + 方法：透過預定義的boundary去抑制non-target的相似度。我們在這裡使用具有對應正交關係的boundary（p\_m = 1 – n\_m），將可以同時規範類內緊湊性與類間離散性從而達到更好的結果。