

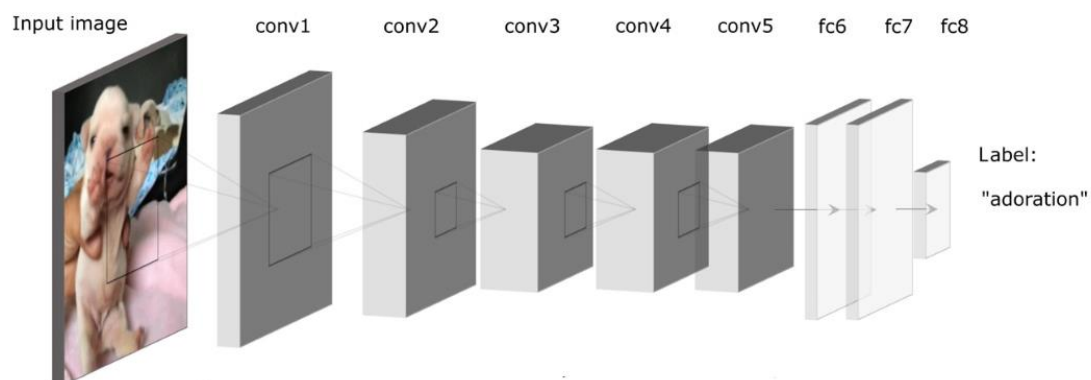
標題：從大腦活動看透情緒

關鍵字：情緒分類、AlexNet、IAPS、枕葉、fMRI



理論學家長久以來認為，當生物發覺生存環境有所改變，甚至打亂原本的平靜時，「情緒(emotion)」會是他們典型的反應之一。而近期的研究指出，由於生物利用各類感官與環境溝通、交流，感覺皮質(sensory cortex)中的表現或多或少能反映出情緒的狀態。舉視覺為例，當我們撞見大屠殺場景時，將會產生恐懼或反胃的情緒；看到美食佳餚時，則會讓人與食慾連結。以下我們將以情緒評價論(Appraisal Theory)中的情感圖式(emotion schemas)代表各種情緒，討論相關的計算模型，進而闡述「情感圖式究竟如何嵌入人類的視覺系統」。

EmoNet



(圖一、EmoNet 結構)

EmoNet 是知名深度神經網路 AlexNet 的變形款，雖然在結構上承襲了 AlexNet 的五個蜷曲層(convolutional layer)與三個全連接層(fully-connected layer)，但是功能已從 AlexNet 的物體辨識修正成情緒分類。研究學者們會根據不同需求，將特定的感官資料餵給 EmoNet 的輸入端，再設法根據輸出的情緒圖式驗證假說，解釋感官資訊與情緒分類之間的關聯。

就此，我們先著眼於視覺對於情緒的影響，提出三項假說：首先，模型可以單從圖片的特徵預測出各項情緒分類的比例，並可以正確分辨不同的情緒分類；第二，模型能夠透過大腦視覺皮質中的活動狀態表示；最後，視覺皮質必須成功破解不同的情緒種類。

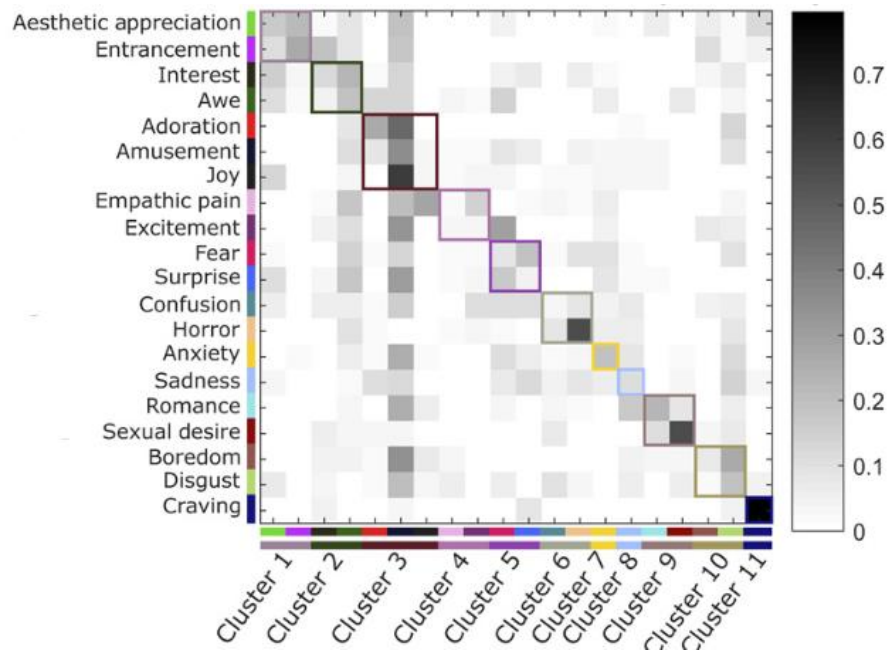
靜態影像的情緒分類



(圖二、分類與影像的範例)

為了測試第一項預測，說明模型可以抽取影像特徵進行 20 項情緒類別分類，EmoNet 會從影片中擷取 137482 張照片，搭配 853 位參與者的分類標記送入與 AlexNet 相同的前七層，並重新訓練最終的全連接層，以達到修改模型目的，完成情緒預測的任務。

在面對與訓練資料完全相異的 24634 份測試影像時，和人類事前討論出來的分類共識相比，EmoNet 能夠得到 62.6%的 top-5 準確率以及 23.09%的 top-1 準確率，而橫跨 20 項情緒類別的 ROC 曲線下方面積(AUC)大約 0.75(效應值 Cohen's $d = 0.945$)，結果優於隨機猜測，除了能夠代表 EmoNet 擁有二十類情緒分類的預測價值之外，更顯示出模型對於不同情緒的分辨能力。

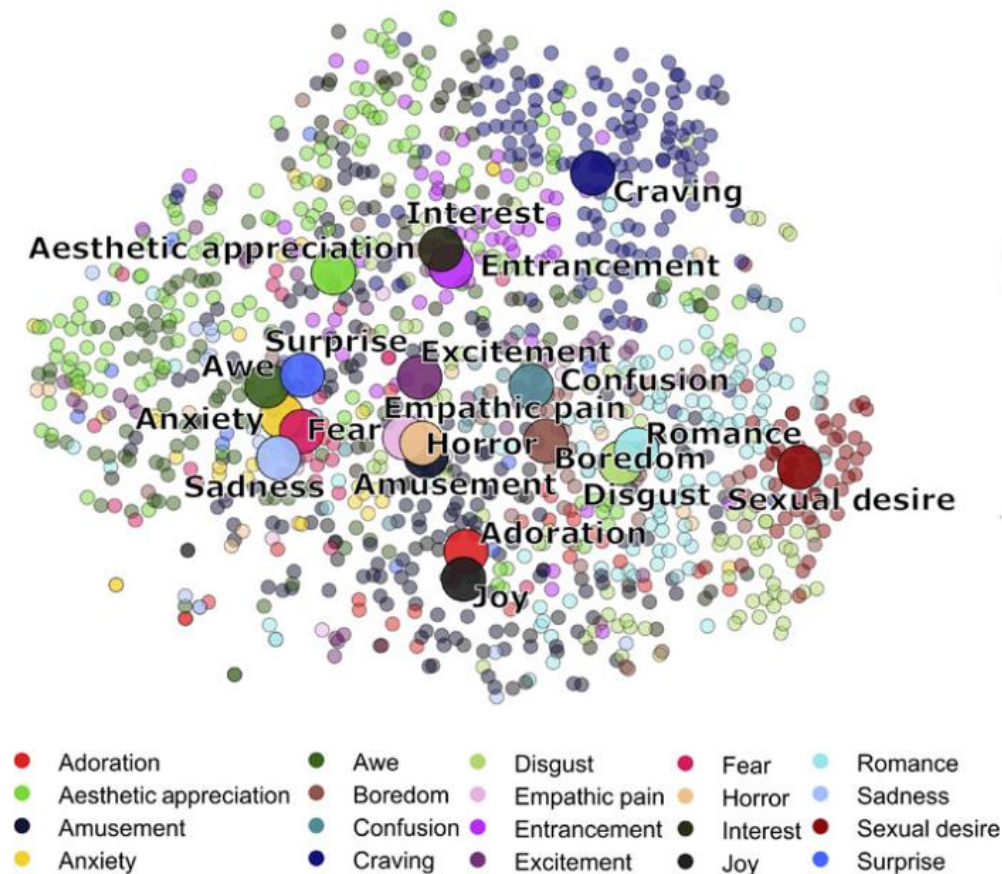


(圖三、混淆矩陣)

若我們以預測結果為行座標、正確分類為列坐標，畫出一張測試實驗的標準化混淆矩陣(normalized confusion matrix)，將會發現二十項情緒分類大約集結成十一組群集(如圖三)，說明測試資料中某些情緒之間的界線，其實不夠清晰，導致機器難以預測正確。舉例來說，「有趣(amusement)」、「愛慕(adoration)」與「喜悅(joy)」在意義上不盡相同，但所對應到的視覺特徵極為相似，導致機器的辨識產生誤差。此外，在矩陣的對角上，存在一些較大的數值，其所對應到的情緒(像是「情慾(sexual desire)」和「渴望(craving)」)可能擁有相當完整的視覺表現，比起那些抽象或短暫的情緒(例如「困惑(confusion)」或「驚喜(surprise)」)，更適合透過 EmoNet 預測。

情緒分類的推廣

研究指出，視覺對人類產生的正面程度和刺激感，是影像情緒特徵的兩大要素，好比「反胃(disgust)」、「恐懼(horror)」及「悲傷(sadness)」的情緒都與高刺激性的負面經驗息息相關，而人類在事前為素材分類情緒時，可能也是依循上述的模式。因此，所有關於情緒的變數，都可以利用二維的正面程度(valence)-刺激感(arousal)來表示，讓預測的推導能更加明確，模型的應用也更加普遍。



(圖四、圖中的小點代表某張照片在利用正向程度-刺激感預測之後，比例最高的情緒結果；大點則代表各種情緒的平均位置)

這項任務的實踐需要在 EmoNet 最終全連接層中的二十個特徵單位，建構「偏最小平方迴歸模型(PLS regression)」，並輸入由 IAPS(International Affective Picture System)提供的影像素材，其中的每張照片通常都包含一種以上的情緒，好讓實驗更符合現實情況。經由十折交叉驗證的樣本外測試，模型在正向程度方面能達到 88%的正確率(均方根誤差 RMSE=0.9849)，而刺激感方面則是 85%(RSME=0.5843)；也就是說，當 EmoNet 拜訪 IAPS 時，能夠明確分辨出不同正向程度與刺激感等級的差別，也能夠進一步與二十項情緒分類配對(如圖四)。

然而，這樣的發現並不代表二維平面足夠成為輸出結果的編碼，根據研究人員推估，大約需要十七維表示才算完整。由於模型的訓練被侷限於視覺特徵，而非全方位的感受，才導致模型預測與人類的共識產生分歧。舉例來說，「令人敬畏(awe)」的照片通常會被人類評為正向，「令人驚嘆的(awe-inspiring)」則偏向刺激(像是極限運動類型等等)，但他們其實都歸類在「敬畏(awe)」之中，代表正向程度-刺激感的二維平面是有些盲點的。

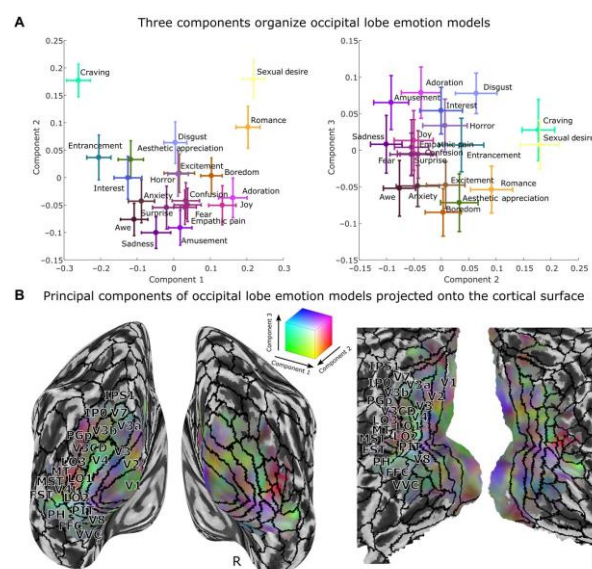
動態影像的分類

在驗證第一項假說時，主要的照片皆來自影片截圖，那麼我們也許能利用假說一的成功，來討論電影預告片的情緒分類—EmoNet 將預測所有各個畫格中的情緒，並在 EmoNet 最終全連接層中，利用 PLS 回歸模型與一對全部(one-versus-all)分類法，推出平均預測結果最符合的影片分類，將靜態影像推廣至動態影像的應用。

在十折交叉驗證之中，EmoNet 將電影預告分為愛情喜劇、恐怖片與動作片的準確率約 70%，ROC 曲線下方面積(AUC)大約 0.855(效應值 Cohen's $d = 1.497$)，結果一樣優於隨機分類，其中絕大部分的誤差都來自恐怖片與動作片，愛情喜劇則不容易被搞混。

EmoNet 與大腦活動

接著回到第二項假說，既然情緒圖式可以由視覺場景推得，那麼 EmoNet 中的所有情緒也許能透過人類視覺系統中的活動表示。為了測試這項假設，我們提供十八位受測者 112 張包含不同情緒的照片，並利用功能性磁振造影(fMRI)提取腦部活動的狀況，之後再經由 PLS 回歸模型找出 EmoNet 與 fMRI 結果之間的回歸關聯。



(圖五、A 表示各種情緒的三大元素組成；B 表示各種情緒影響枕葉的位置)

目前我們只關心視覺處理的部分，因此僅聚焦於人類的視覺系統—枕葉(occipital lobe)。主成分分析(PCA)顯示，枕葉中包含三大情緒相關的特徵，若由 7214 個

體素(voxels)表示枕葉，各種情緒的特徵組成將如圖五所示。由此可知，不同的情緒類型對應到不同的 fMRI 結果，也就是我們能夠合理利用枕葉的活動情形預測 EmoNet 中不同情緒分類單位的異同。若我們透過樣本外留一交叉驗證(leave-one-subject-out cross-validation)，並將結果畫作混淆矩陣，將會發現這樣的設計能夠在二十種情緒中，成功辨識出十五種，成效相當不錯。

為了提供更多的證據說明枕葉活動與情緒分類之間的關係，我們必須驗證假說三是否為真，於是將前述的訓練模型套用在動態影像分類上。在這實驗中，受測者將觀看七種電影剪輯，分別包含「滿足(contentment)」、「悲傷(sadness)」、「有趣(amusement)」、「驚喜(surprise)」、「恐懼(fear)」、「憤怒(anger)」、「中性(neutral)」，一部影片中不會出現混和情緒，和之前提及的 IAPS 資料庫有所差異。經過樣本外的八折交叉驗證之後，情緒分類結果主要可以分為五組群集，模型較難分辨出「驚喜(surprise)」與「恐懼(fear)」以及「悲傷(sadness)」與「憤怒(anger)」之間的差別，而五項情緒類別分類的準確率大約坐落於 40% 左右，完整的七項情緒類別分類則只能達到 30% 的準確率。

這樣看來，三項假說都得到了相當不錯的解釋，儘管目前僅討論視覺相關的情緒特徵，但那些經常影響情緒的因素，包括其他感官的資訊、關於各種經驗的記憶、面對動態變化的處理以及各種心理意象，在未來也能搭上 EmoNet 的順風車，得到更多解釋情緒的證據。

參考資料：

1. Philip A. Kragel, Marianne C. Reddan, Kevin S. LaBar, Tor D. Wager, "[Emotion schemas are embedded in the human visual system](#)", ScienceAdvances, 03 Jul 2019