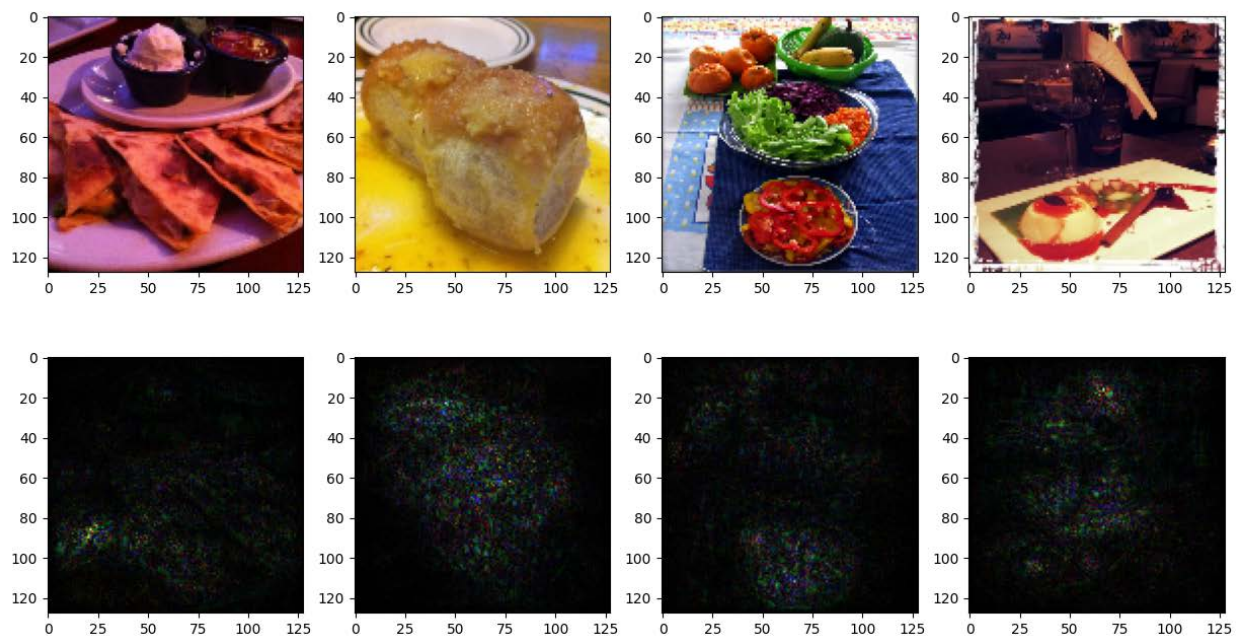


1. (2%) 從作業三可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？
(Collaborators:)

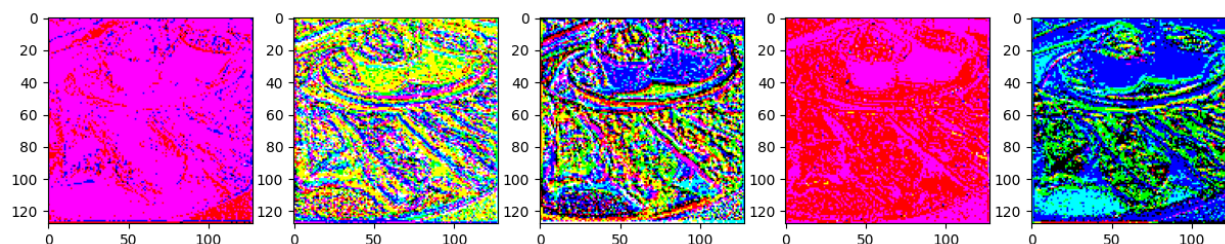
此處隨機挑選 4 張圖片，編號 [454,87,1543,3350]

可以看到在 saliency maps，model 的確是有關注食物出現的位置，在編號 87 以及 1543 可以看到 model 漂亮的點出食物所在的位置，而 454 可能由於食物疊在一起，所以雖然依舊可以指出食物位置，但 activation 的程度較為不一，而 3350 的圖片顏色對比鮮明，食物佔圖片比例又較少，可以看到 model 在 3350 的表現就比較差。

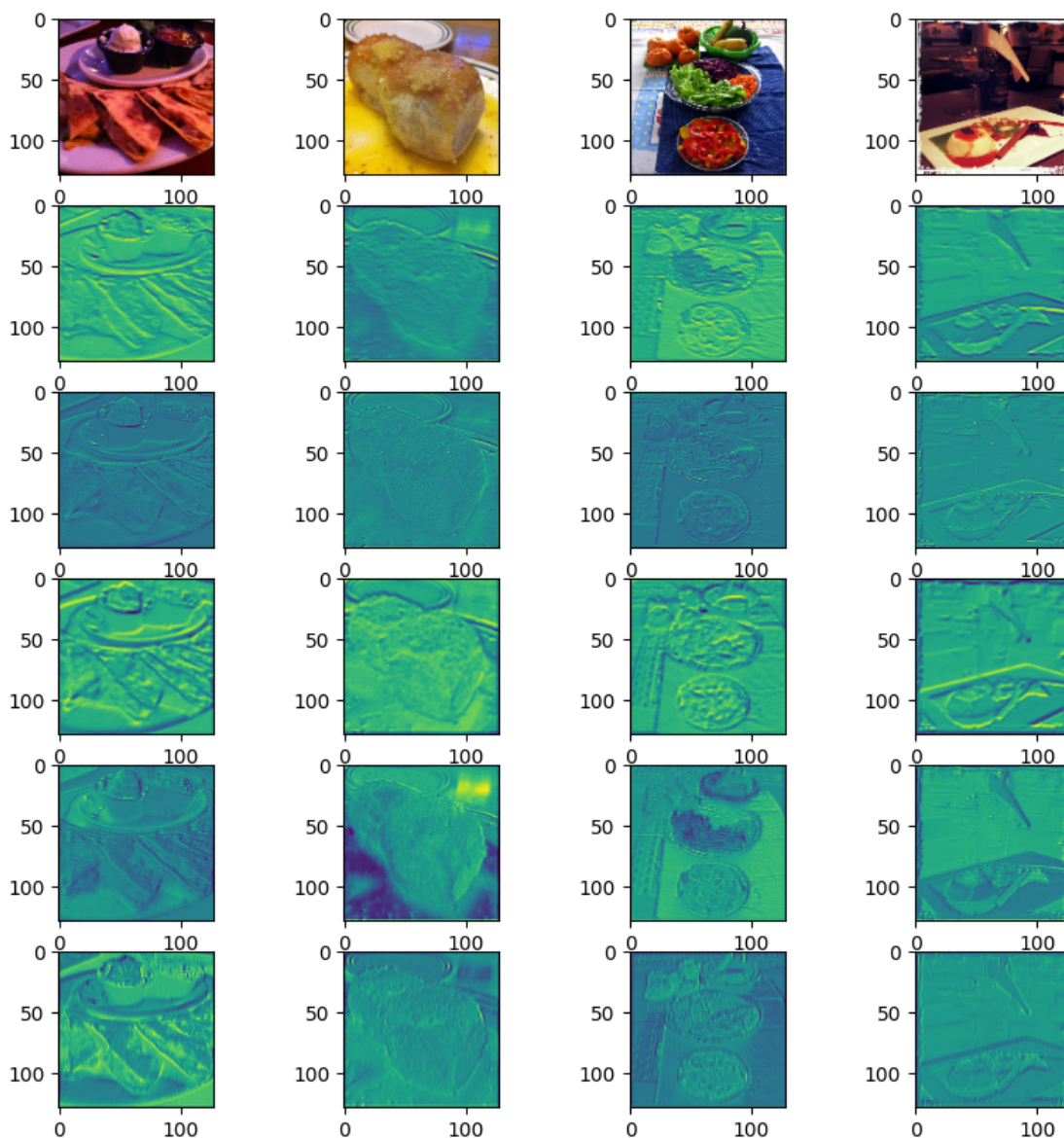


2. (3%) 承(1) 利用上課所提到的 **gradient ascent** 方法，觀察特定層的 **filter** 最容易被哪種圖片 **activate** 與觀察 **filter** 的 **output**。(Collaborators:)

我觀察了 `cnn_id = 3`, `filter = 1,15,30,45,63`，filter activation map

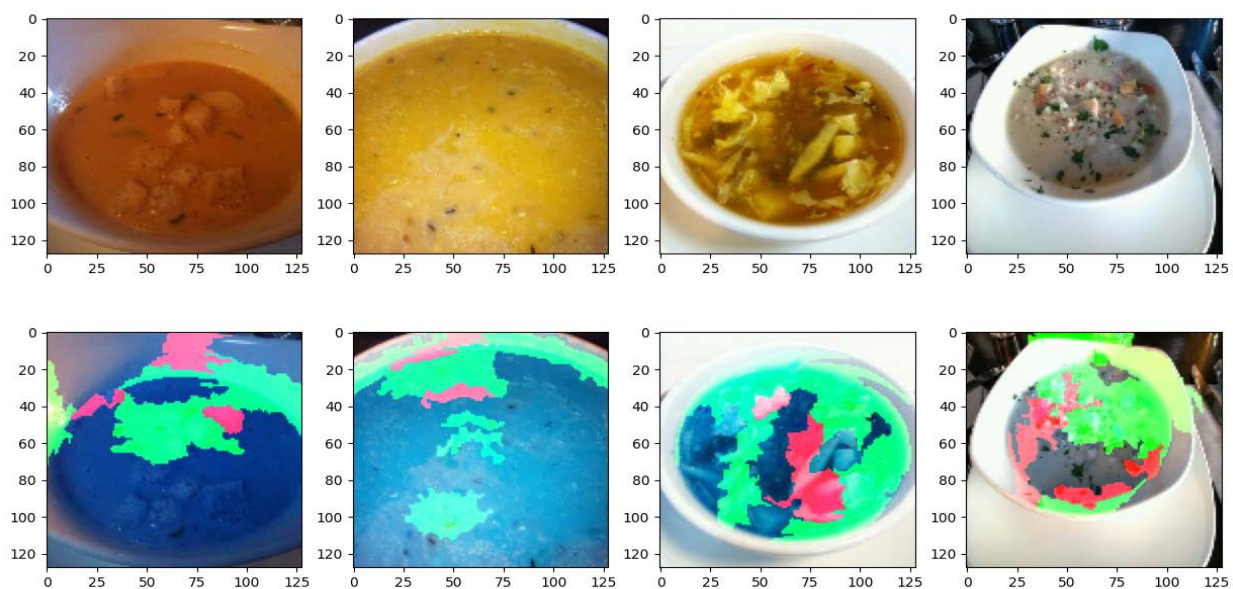


可以看到這五張圖片的「構圖線條」其實都差不多，但是各自的色調卻差非常多，所以我認為這層的主要是在分辨圖片中色調的差別。接著看 output

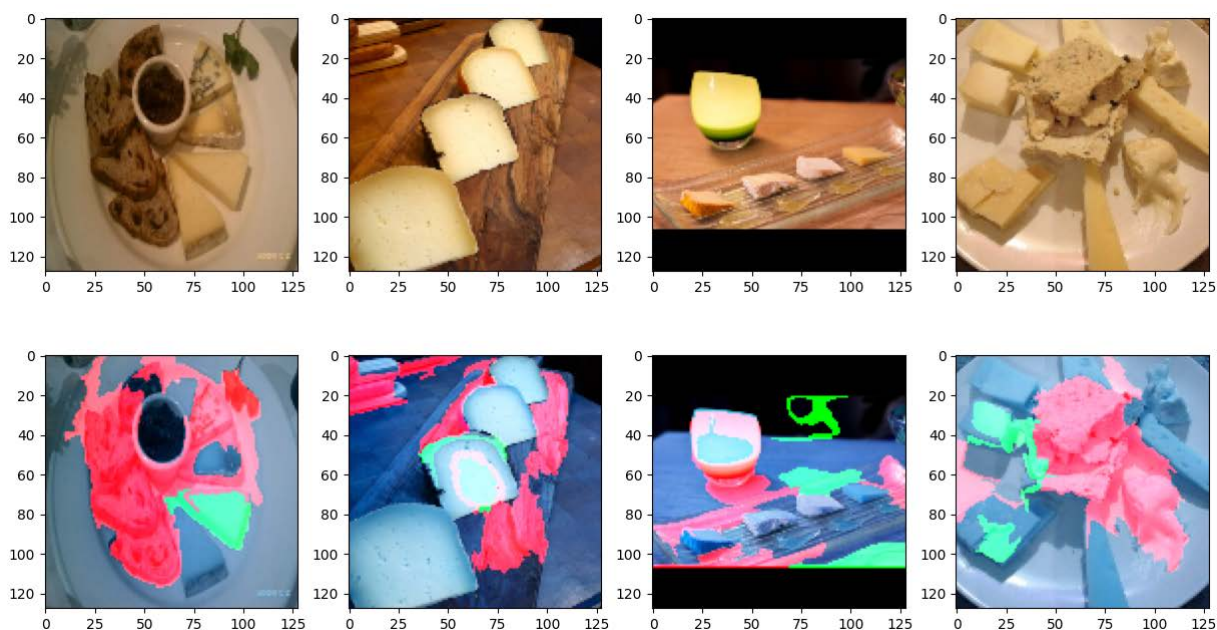


各列自上而下分別為 filter 1, 15, 30, 45, 63，可以看到各張圖針對同樣色系的區域，在不同的 filter 下有著鮮明對比的 output，由此可知此層的確是在辨別照片的色調差異。

3. (2%) 請使用 Lime 套件分析你的模型對於各種食物的判斷方式，並解釋為何你的模型在某些 label 表現得特別好 (可以搭配作業三的 Confusion Matrix)。



根據作業三的 confusion matrix 中可以發現，我的模型在 label 9 表現最好，有 9 成的命中率，但 label 1 卻只有 5 成 5，此處隨機挑此兩 label 的圖片來比較。



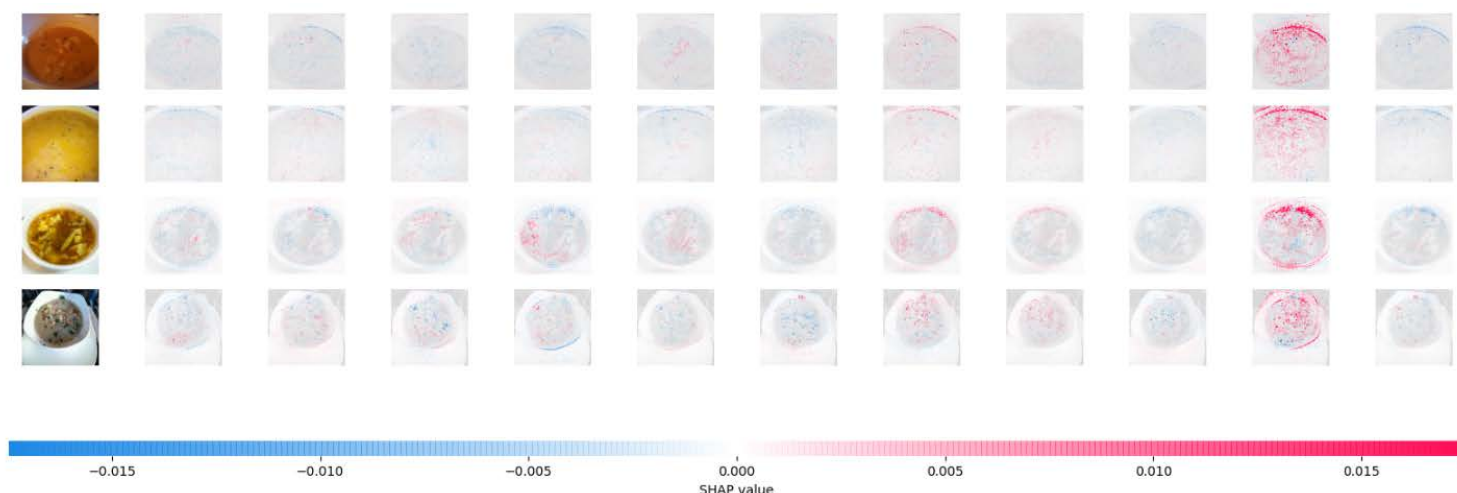
上圖為挑選 id = 9008,9009,9010,9011 四張圖片，可以看到當湯內有料時，model 可以準確鎖定碗內食物，並判斷他是「湯」，而當湯內料較少時，model 則改為專注於碗與食物的邊緣猜測，而這也很合理，因為湯是液體，與碗會有接合的地方。

上圖為挑選 id = 2001,2002,2003,2004 四張圖片，可以發現圖片的組成十分不一致，儘管人看得出是麵包，但由於擺盤，多種拍攝角度以及光源等等問題，使得圖片的辨識度不高，甚至在應該是麵包的地方 model 卻給出了負相關的輸出。

4. **(3%)** 【自由發揮】請同學自行搜尋或參考上課曾提及的內容，實作任一種方式來觀察 CNN 模型的訓練，並說明你的實作方法及呈現 **visualization** 的結果。

此題使用 shap 套件來延伸解釋第 3 題。由於已經發現 label 9 以及 label 1 在該 label prediction 下模型注意的地方，不如再看看模型其他 label output 對 label 9 以及 label 1 的關注點又在哪裡

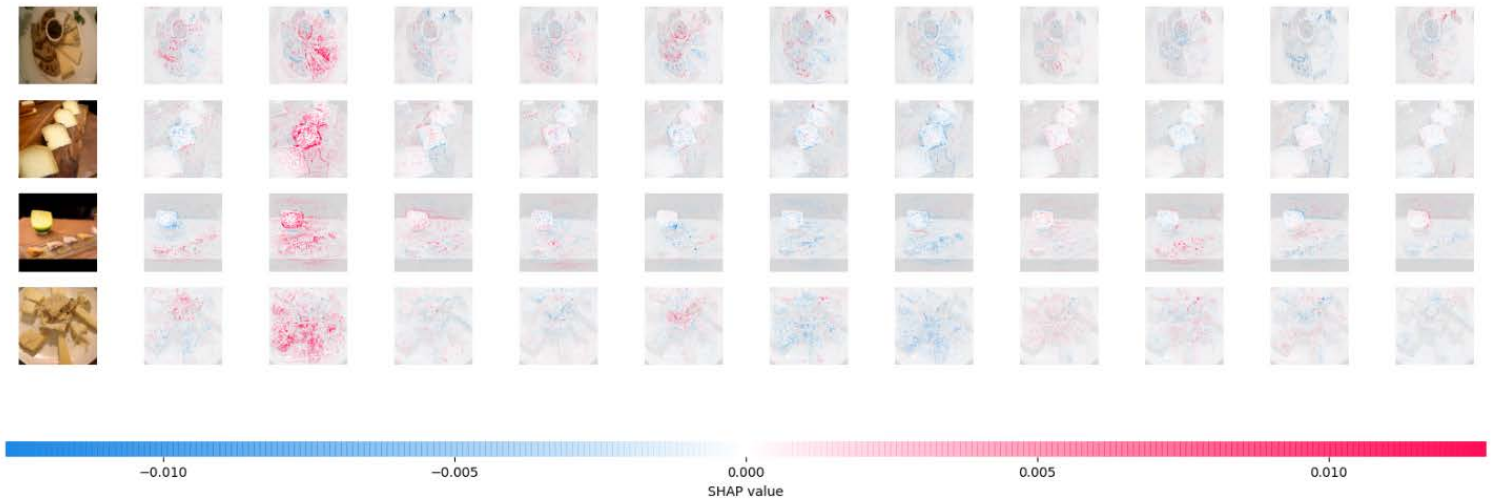
下圖的 label 順序為 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10



Label 9

可以看到在 label 9 的部分 SHAP value 高的區域相對比起其他 label 都還要多很多，跟 LIME 的結果一樣，同時其他 label 的區域就相較 label 9 淺上許多，惟要注意的是許多其他 label 也有專注在碗的邊緣 (label 6)，要是湯裡頭的料太不明顯，導致辨別需要依靠碗的邊緣時，該圖就容易辨識錯誤。

Label 1



從 label 1 的 SHAP value 來看，可以看到預測成果比較差的 label 1，在其他 label 的 output 中同時也佔有不少的正負相關，由於此類別的圖片內容擺放雜亂，以及常語其他類別食品共同出現（如圖片 1，麵包(label0)與起司(label1)放在一起），故使得其他 label 也容易剛好遇到相似的特徵。故可以看到其他圖片中的紅色散佈比例相較於 label 9 的散佈比例，就高很多。

但剛好可以從這張圖的圖一中發現，SHAP value 都關注在起司，麵包的 value 是 0，故我們可以認定 label 0 以及 label 1 可以分得很開，再透過 confusion matrix 發現誤判比例， $\text{true0, pred1} = 0.006$ ， $\text{true1, pred0} = 0.062$ ，都十分的低。所以我們也可以透過 SHAP value 來判斷，到底 label 1 跟誰（其他 label 內容物）同時出現時，容易混淆。