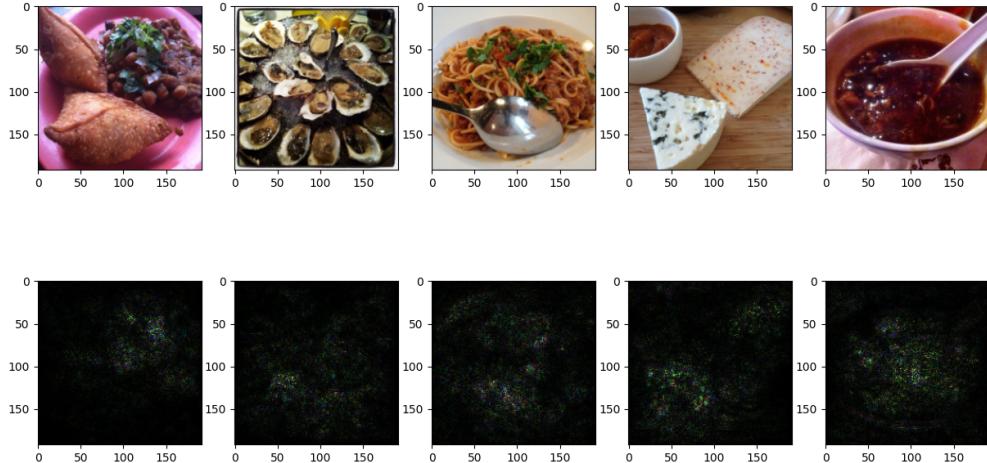


# Machine Learning Spring 2020 - HW5 Report

學號:B07902064 系級: 資工二 姓名: 蔡銘軒

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？

在這個report裡，都會使用相同的五張圖片做比較。這五張圖片是我random選擇出來的，從左邊到右邊分別屬於Fried food, Seafood, Noodles/Pasta, Dairy product, Soup



第一張圖片裡，影響model判斷比較多的區域似乎是右上角那一塊看似肉燥或是納豆(我也無法辨識)的東西，而不是左半部的炸物。我認為這對model而言還算合理，因為即使是人眼辨別，我也不知道左半部的食物是炸物（乍看之下以為是豆皮壽司或是肉粽），且右上角的部分確實也佔了這張圖片很大的比例。神奇的是model在這張圖片的預測是正確的。

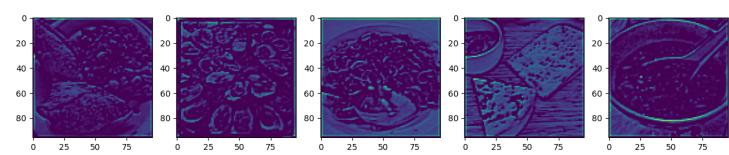
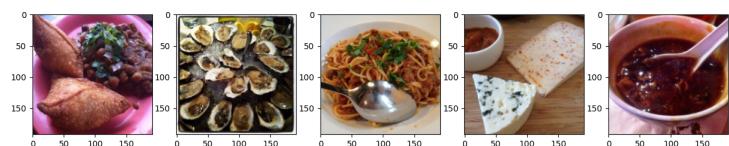
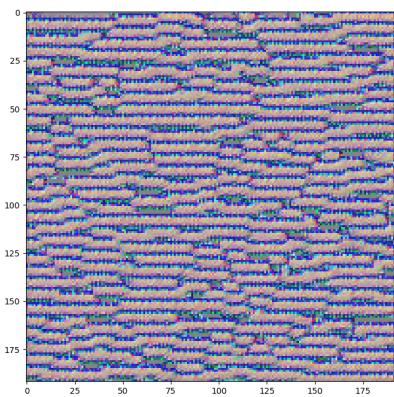
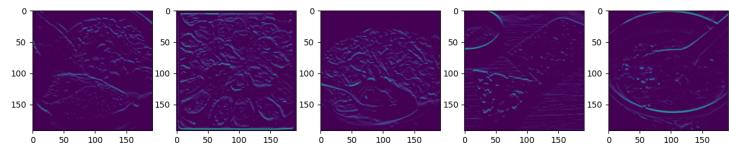
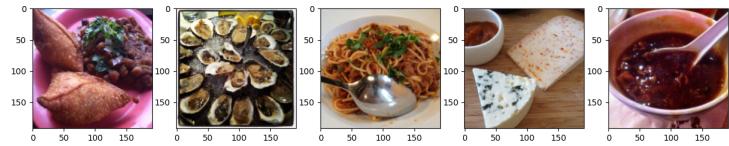
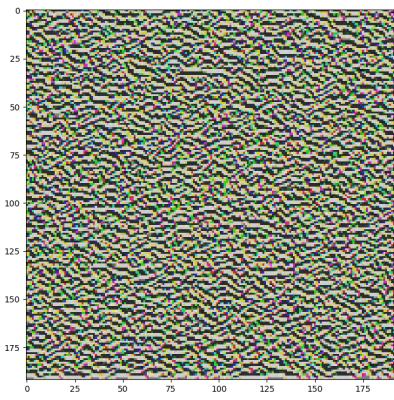
第二張圖片的亮點比較分散，符合圖片裡幾乎全都是海鮮的結果。且亮點確實分佈在海鮮上，而這張圖片也被model正確的分類。

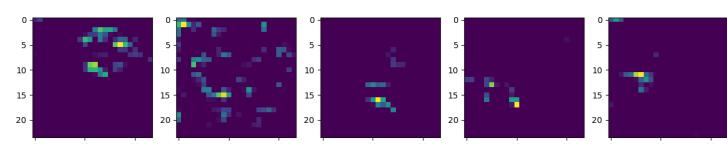
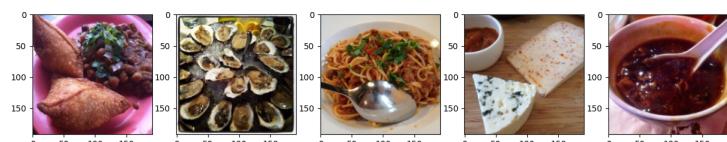
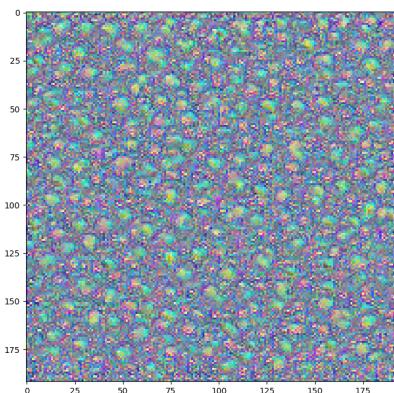
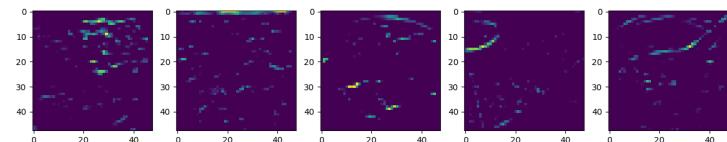
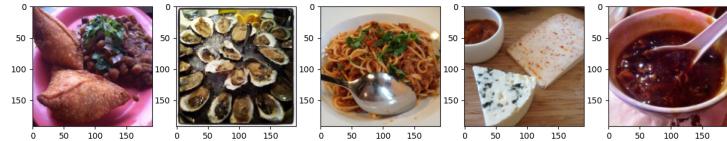
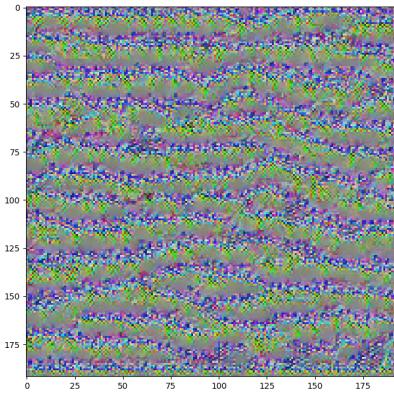
第三張圖片裡湯匙的影響是滿大的，在這邊還看不太出來湯匙是提升還是降低分類成Noodles/Pasta的機率，但因為model也有捕捉到麵的部分，這張圖也被正確的分類。

第四張圖對model影響最大的是左下角的起司，右上角的起司切片也有被捕捉到。左上角用碗裝的醬則不太有影響，model有捕捉到這張圖的重點且正確的分類。

第五張圖影響較大的是碗裡裝的東西，一樣這邊還無法分辨是正影響還是負影響。仔細觀察還會發現碗的輪廓（尤其是下緣）也是有被捕捉到的。這張圖的分類也是正確的。

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。





我檢視了model裡前四層ReLU，按照上到下的順序。

從左邊的圖看起來，第一層在捕捉圖片裡基本的線條。而從右邊的圖片看起來，model確實捕捉到了各個圖片裡的基本輪廓，例如第五張圖（Soup）有看到了碗跟湯匙的外型，但湯的內容還不是很明顯。

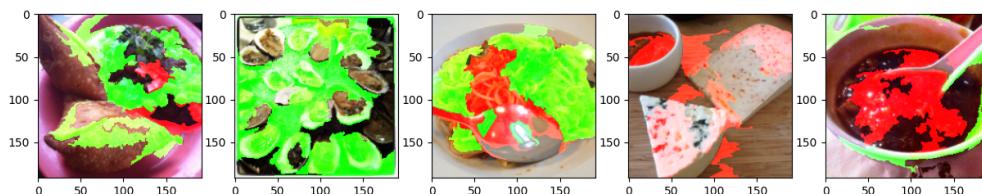
第二層從左邊的圖片看起來是在捕捉圖片裡的紋路，而右邊的五張圖片跟第一層比起來，也掌握了更多細節。例如第四張圖片（Dairy product）左下角的起司上，紋路比第一層的明顯許多，而那些紋路滿符合左邊圖片的模式，感覺是正好是這一層要抓的東西。第二張圖片（Seafood）也捕捉到海鮮更多的細節，相較於第一層看起來只捕捉到一些橢圓的外框又細緻了不少。第一張圖片（Fried food）裡也在第二層出現了炸物上的紋路，在第一層時只有捕捉到炸物的外型而已。整體來說五張圖片在這一層都顯得更細緻了。

第三層感覺捕捉到的有弧度的線段。例如右邊的第四張圖（Dairy product）很明顯地捕捉到了碗的外緣，第三張（Noodles/Pasta）跟五張圖（Soup）也抓到了湯匙跟碗的部分。

第四層從左邊的圖看起來，在捕捉一些圓形或是顆粒狀的部分。右邊的圖片因為經過前面幾層layer被降低了畫素，已經難以辨認原本的圖片，但還是可以看到model捕捉的部位。第一張圖（Fried food）明顯看到右上角很多顆粒狀的（肉燥或是納豆）東西被model看到，第四張圖片（Dairy product）主要被看到的部分也是左下角那塊起司上的顆粒特徵。第五張圖（Soup）也捕捉到湯裡顆粒狀的內容物。

後面的layer因為圖片的像素已經低到無法辨識，因此沒有繼續討論。但從前四層可以發現model在每一層都捕捉不同的特徵，且捕捉的特徵有越來越複雜的趨勢。從第一層捕捉簡單的外型、第二層細緻的紋路，到最後一層顆粒狀的特徵。

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)。



Lime的結果跟Saliency Map的結果大致是呼應的。第一張圖確實右上角的肉燥/納豆對model的影響是大的，意外的是居然是正面的影響，可能model誤以為那是炸物。在Saliency Map裡沒有顯示出來的是左下角的炸物，在Lime的結果來看對預測也是有正面影響的。

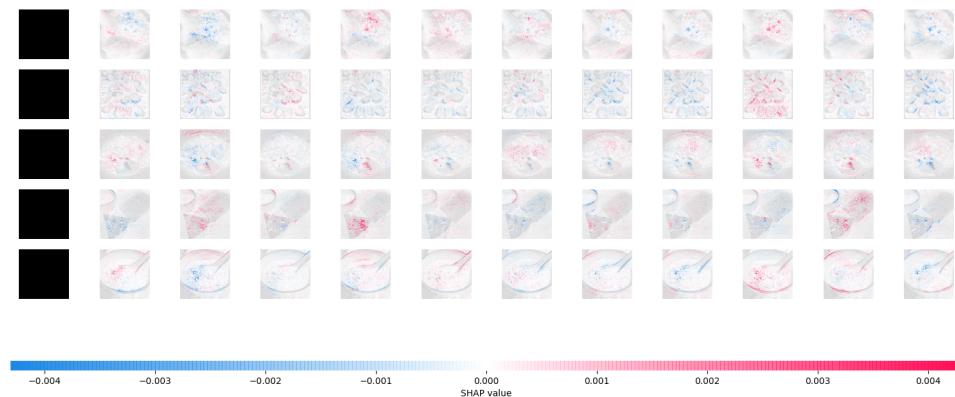
第二張圖因為幾乎整張圖都是海鮮，因此Lime的結果還算合理，塗色的範圍也大致跟Saliency Map呼應。

第三張圖可以看到在Saliency Map裡影響很大的湯匙，在Lime的結果顯示是影響model判斷的，反而是麵的部分增加了model的信心，看來model是知道應該如何判斷麵條的。在我所有類別的預測裡，Noodles/Pasta的準確率是前三名。

第四張圖感覺使model困惑了。幾乎整張圖都是負面的影響。在所有類別裡，我的model對Dairy product的準確率是偏低的（約65%左右）。我認為這個現象有出現在這張圖片裡。例如左上角的碗，或許會讓model以為是Soup的碗，起司上圓弧的部分可能也讓model誤判。

第五張圖則觀察到湯的內容物是負面影響model判斷的，反而model是看到碗的形狀認為這是一碗湯。我認為這是合理的，因為這張圖片湯的內容物並不是清湯，而是有許多東西，因此model如果看內容物的話可能會分類成其他東西，例如Meat等（湯裡面確實看起來是肉）。而許多Soup的圖片都有一個很大的碗，因此model把碗當成Soup的特徵也是無可厚非。

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。



我使用Python裡的Shap套件來分析model。

除了最左邊的黑色圖片，從左到右共11個column分別代表11個類別（按照Bread, Dairy product...的順序）。

第一張圖 (Fried food) 大致符合前面Saliency Map跟Lime的結果，右上角的部分影響了model的判斷且是正面的影響，增加model想要分類給Fried food的機會（雖然不知道為什麼model覺得那是炸物）。比較值得注意的是同樣的區域也增加讓model分類成Egg的機會，可能是因為訓練資料裡有一些散蛋也是呈現類似的外型。但我在檢查model的正確率時，發現Fried food其實不會很常跟Egg搞混，這張圖可能剛好遇到這樣的情況。

第二張圖感覺對model來講是容易的，很清楚的看到這張圖在Seafood類別很明顯的一片紅色，且區域符合Saliency Map跟Lime標色的區域。這張圖對model來講非常明顯是海鮮，但在Dessert類別其實也呈現較輕的正面反應，可能是Dessert的圖片太五花八門了，很多東西都可能是Dessert。但除此之外這張圖片並沒有會被誤認為其他類別的跡象，大多是藍色。

第三張圖雖然在很多類別都有紅色的反應，但仔細看的話還是會發現在Noodles/Pasta的反應是比較強的，因此還是分類給了Noodles/Pasta。而且跟Lime的結果很符合，model確實是看到麵的部分才覺得這是Noodles/Pasta，而看到湯匙反而是讓model混淆。再仔細看的話會發現湯匙其實增加了分類給Egg, Seafood, Soup的機會，我認為這也是合理的。單看湯匙用來盛物的部分，確實是個橢圓形有點像Egg，而Seafood也常常出現貝類等橢圓的物體，例如這次取樣的第二張圖 (Seafood)，整張圖片都是橢圓形的海鮮。而Soup則是類似上一題的分析，因為model比較難從內容物判斷是不是Soup，所以他比較容易從碗或是湯匙這些常跟Soup一起出現的物體判斷。

第四張圖對model來說就遇到困難了。Shap在左下角的起司上呈現正面影響，不過觀察這張圖在其他類別裡的反應，發現同樣的區域其實在Egg, Soup裡也都很強的正面反應，甚至比Dairy product更強。也許在Lime裡這塊反應呈現紅色（負面）是因為model在這邊感到很困惑。我懷疑model是看到起司圓弧的邊緣所以以為是Egg或是Soup。

第五張圖符合前面Saliency Map跟Lime的觀察。model確實是看到碗跟湯匙所以把這張圖分類給Soup，內容物其實是妨礙判斷的。從Shap來看model看到內容物是會想要分類給Seafood的。幸好model有捕捉到整個碗的圓框，最後還是分類給Soup。