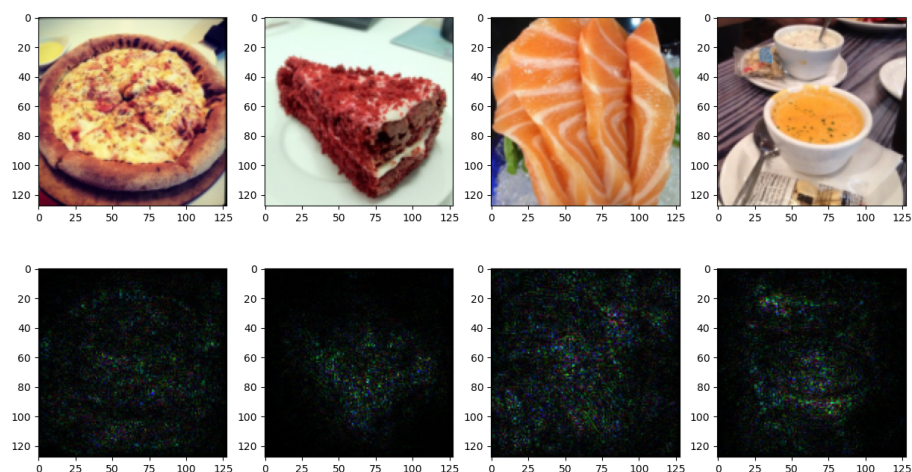


Machine Learning HW5

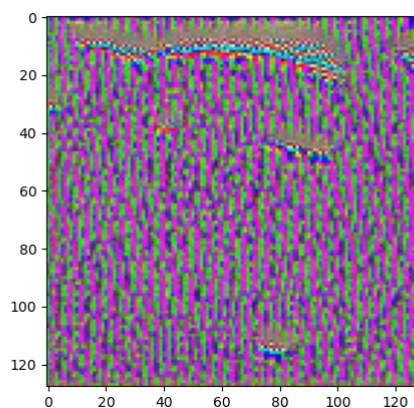
學號：B06902060 系級：資工三 姓名：鄒宗霖

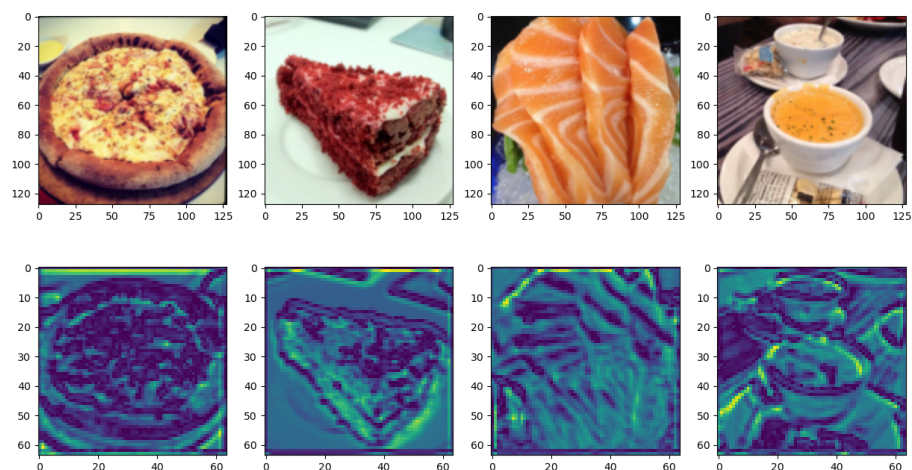
1. (2%) 從作業三可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？



從上面第一張到第三張圖可以看到模型抓出了食物大致的形狀，而第四張圖可以看到模型抓出了碗的形狀，因此模型在做 classification 時，確實有看到物體的輪廓而不是胡亂猜測。

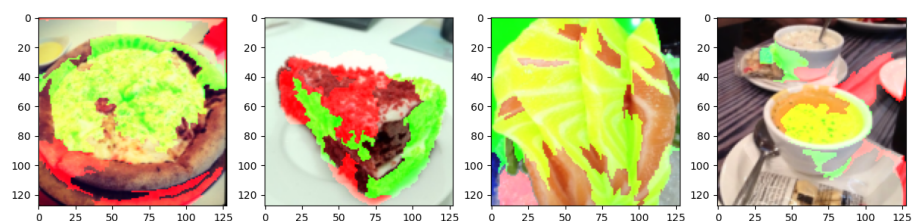
2. (3%) 承(1) 利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate 與觀察 filter 的 output。





第一張圖為可以最大化 activate filter 的圖片 (cnnid = 9 , filterid = 10)，我們可以推斷出這個 filter 主要的任務就是抓出垂直的線條；第二張圖為 filter activation 的結果，從圖中可以看到就算原本不是垂直的線條，這個 filter 也把他切成許多由垂直線條組成的直線。

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)。



上圖為 lime 套件分析的結果，在第一張圖中，lime 是以 pizza 圓形的輪廓以及配料的部分作為分類主要的依據，而且單看 pizza 底下的陰影，lime 容易做出錯誤的判斷；在第二張圖中，lime 是以蛋糕側邊的分層作為分類主要的依據，而且單看蛋糕上的配料，lime 容易做出錯誤的判斷；在第三張圖中，lime 是以生魚片的位置作為分類主要的依據；在第四張圖中，lime 是以碗的形狀作為分類主要的依據。在作業三中，pizza、湯為表現比較好的類別，推測是因為模型看到了圓形以及 pizza 的配料、切橫，就知道那是 pizza，而模型看到了圓形及碗的形狀，就知道那是湯；點心、海鮮為表現比較差的類別，推測是因為點心的類型太多了，有些看起來像麵包、有些看起來像乳製品，因此模型表現不佳，海鮮的類型多、顏色差異也大，像上面第三張圖中，生魚片後面黑色的背景竟然被 lime 套件標成綠色，推測可能被誤認為是黑色的牡蠣 (？因此模型表現不佳。

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。

[reference] : <https://github.com/utkuozbulak/pytorch-cnn-visualizations>



上面幾張圖為實作 deep dream 的結果，實作的方法為利用 gradient descent 更新 input image 的參數來最大化某個 (noodle/pasta) neuron output 的平均 (learning rate = 0.04, weight_decay = 1e-4)，我們讓模型誇張化它所看到的東西，而誇張化的對象就是模型 noodle/pasta 這個類別的 output，從上到下分別為調整圖片參數 iter = 200, 400, 600, 800, 1000 的結果，我們可以看到麵包的紋路慢慢地被一條一條的麵線所取代，這就是模型把 noodle/pasta 這個類別的 output 所看到的東西誇張化的結果，也代表模型心目中所認為的麵長這樣。