Machine Learning HW5

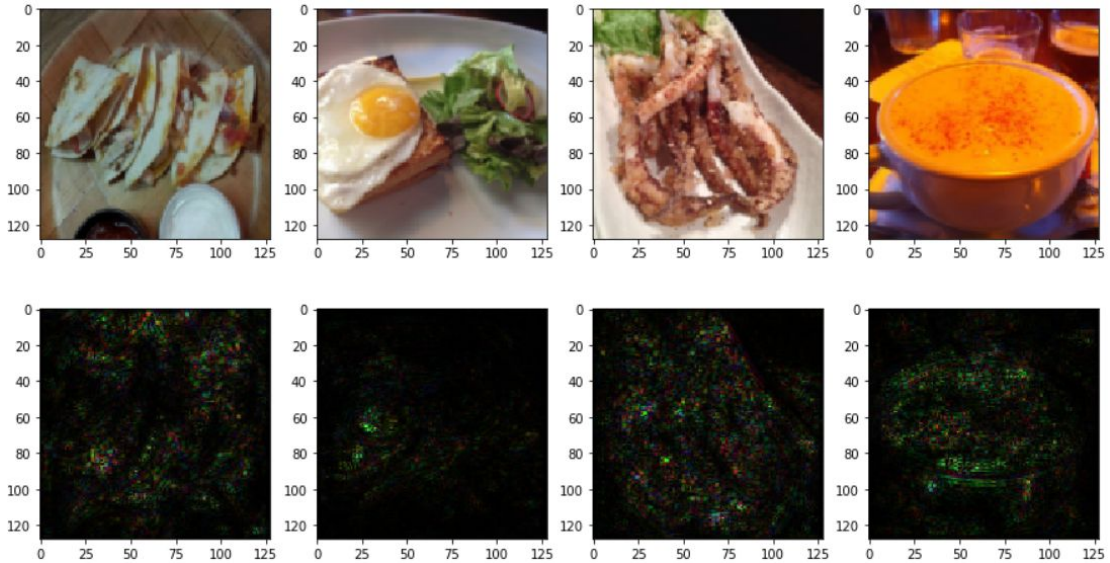
ML TAs
ntu-ml-2020spring-ta@googlegroups.com

Outline

- Task Introduction
- Task1 - Sailency Map
- Task2 - Filter Visualization
- Task3 - Lime
- Task4 - Any visualization/explaining method you like
- FAQ

Task1 - Saliency Map

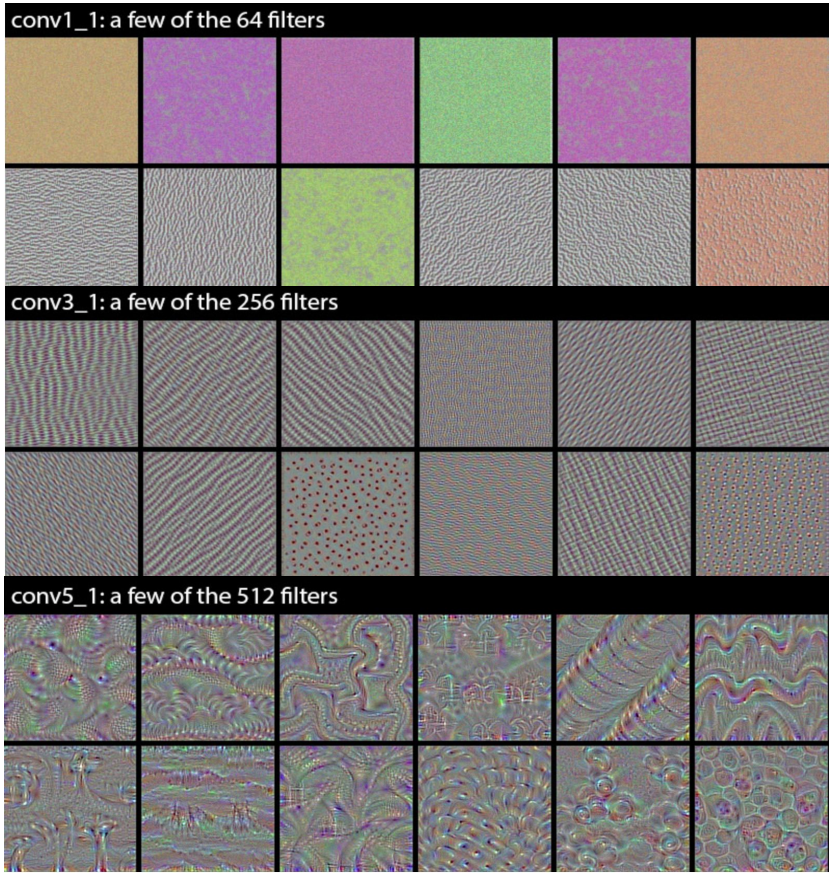
Compute the gradient of output category with respect to input image.



Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps: https://arxiv.org/pdf/1312.6034v2.pdf

Task2 - Filter Visualization

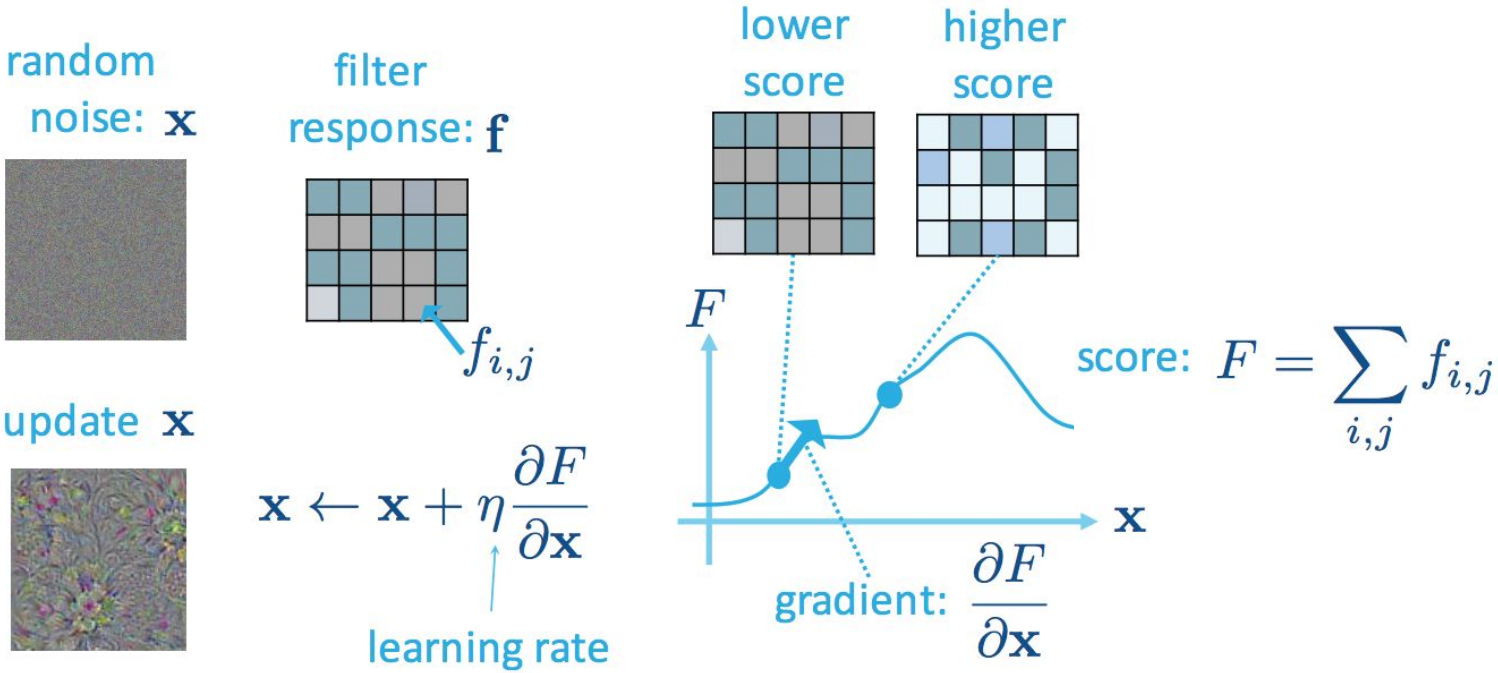
- Use **Gradient Ascent** method to find the image that activates the selected filter the most and plot them (start from white noise).



Ref:
https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html

Filter Visualization

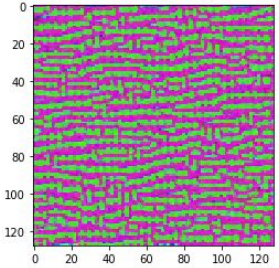
- Gradient Ascent : Magnify the filter response



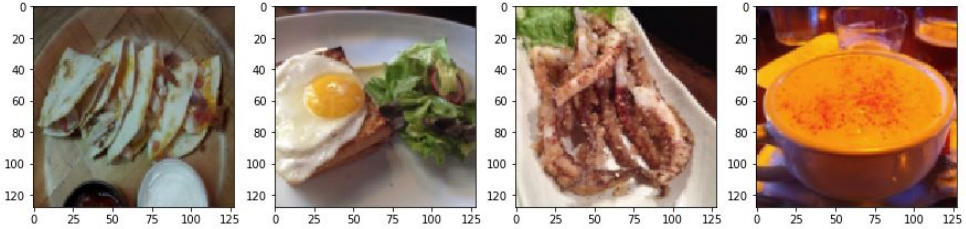
https://www.csie.ntu.edu.tw/~yvchen/f105-adl/doc/161103_ConvolutionalNN.pdf

Filter Visualization

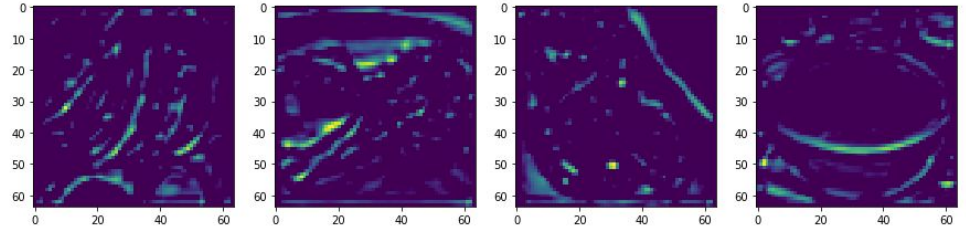
Filter visualization



Original image



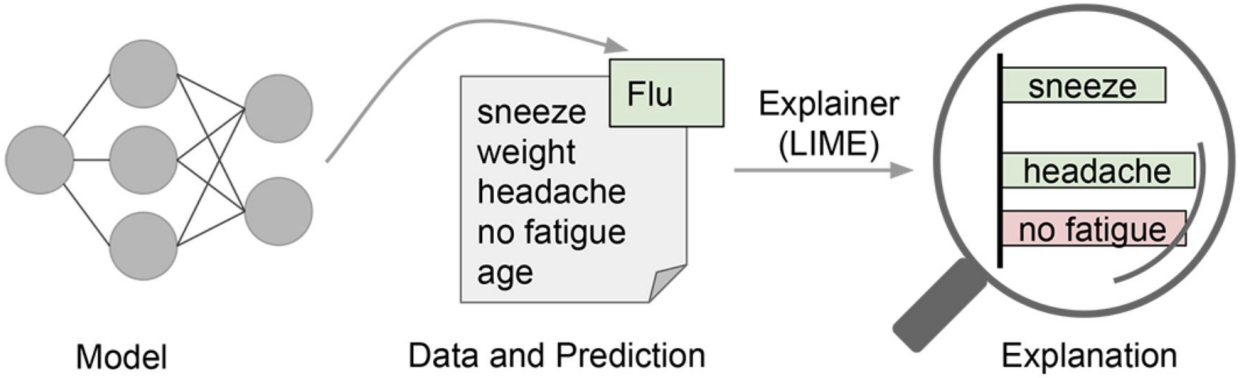
Activation map



Task-3 Lime

Local Interpretable Model-Agnostic Explanations

To approximate a black-box model by a simple model locally



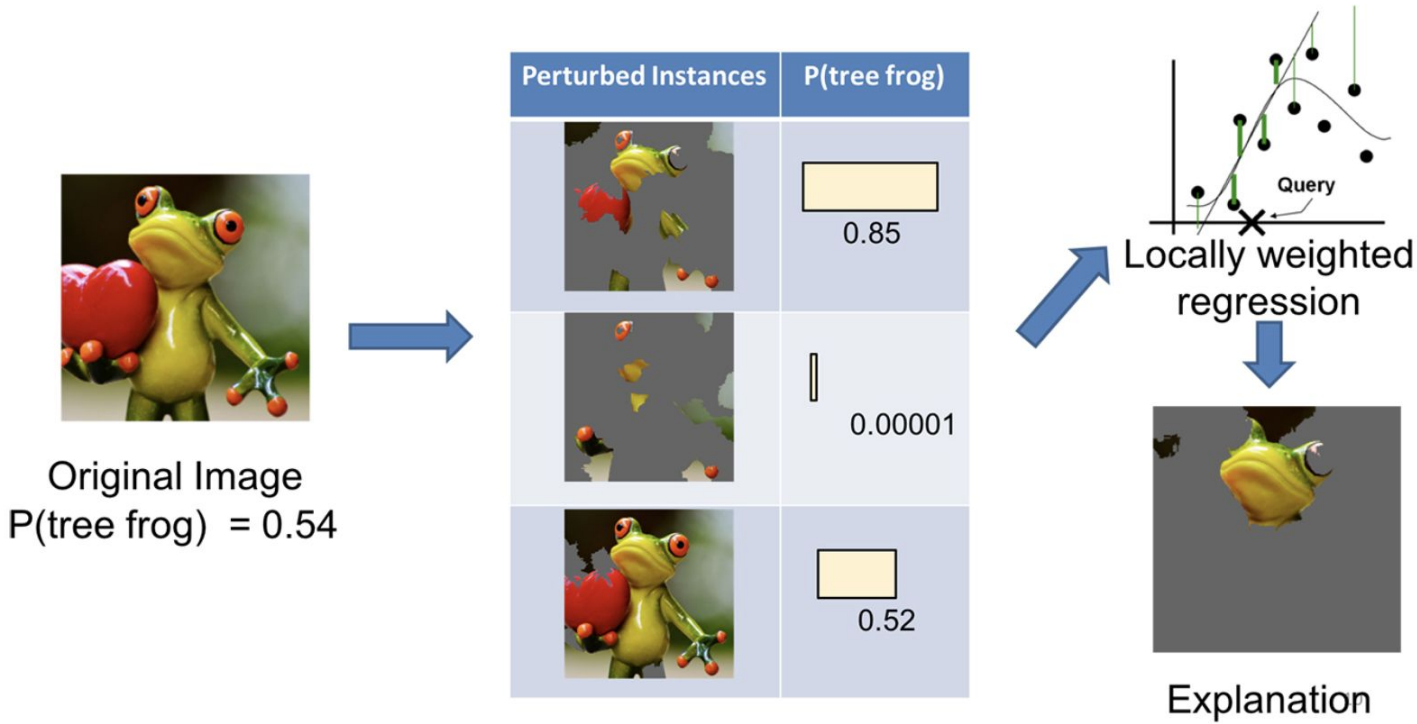
Ref: "Why Should I Trust You?": Explaining the Predictions of Any Classifier

Lime - 1/3

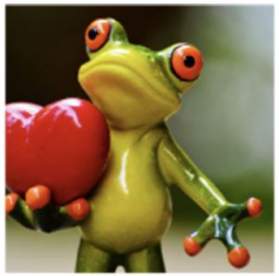
Original Image

Interpretable Components

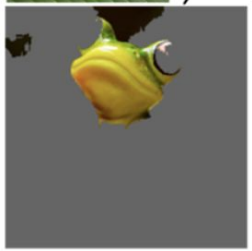
Lime - 2/3


Original Image
P(tree frog) = 0.54
Perturbed Instances
P(tree frog)
0.85
0.00001
0.52
Query
Locally weighted regression
Explanation

Lime - 3/3



P() = 0.54

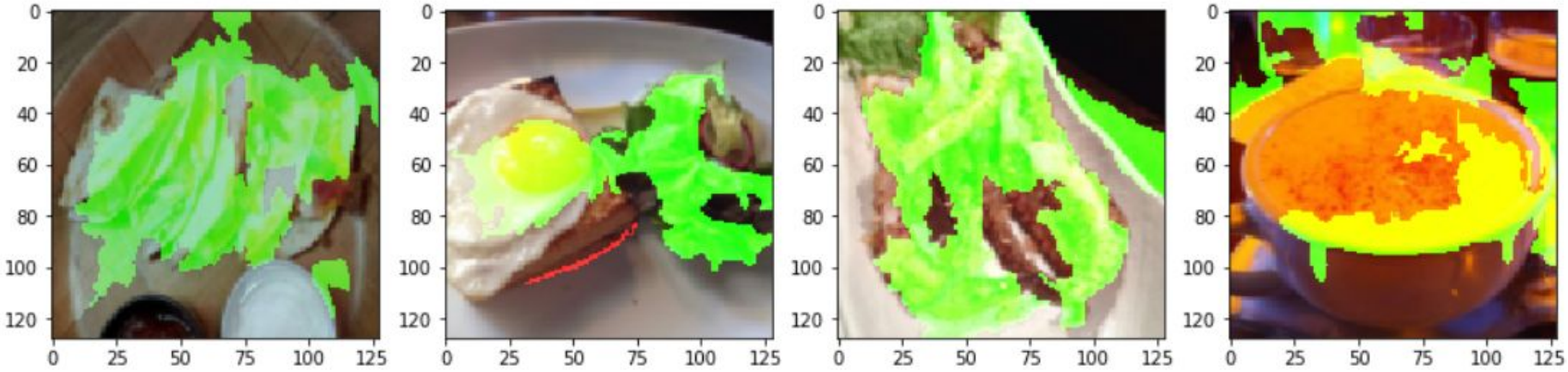


P() = 0.07

P() = 0.05

Lime

> pip install lime



GitHub Repo: https://github.com/marcotcr/lime

Ref: https://goo.gl/anaxvD

Requirements (1/2)

- 使用 HW3 的 Food Dataset 且不能使用額外 dataset
- 使用你 HW3 train 好的 CNN model, **不能使用 colab 範例的 checkpoint 和圖**
- 請用 template 寫一份 report.pdf 回答以下問題 (需用程式畫圖說明)
  - **(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？**
  - **(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。**
  - **(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)。**
  - **(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。(請附上 reference)**
    - eg., Deep Dream, Shap, others

Requirements (2/2)

- 請寫一份 hw5.sh 滿足以下內容
  - Usage: `bash hw5.sh [Food dataset directory] [Output images directory]`
  - The script should NOT require ANY USER INTERACTION
  - Download your checkpoint
    - How to upload & download your pretrained model
  - Model definition & load checkpoint
  - Draw and save **ALL** images appearing in your report
  - **嚴重:任何一張 report 的圖片若沒有經由 hw5.sh 畫出, 則該圖與其相關敘述不予計分**
  - hw5.sh 需在 **10 分鐘**內跑完

Submission

- GitHub 上 hw5-<account> 請至少包含：
  - report.pdf
  - hw5.sh
  - **嚴重：請不要上傳 dataset 到 GitHub，也不要在 hw5.sh 中下載 dataset。違者直接扣此作業總分 3 分，且沒有補救機會！**

Policy

- 遲交及補救規定請參考期初公告
- 每次批改完後會公告分數與算分依據
- 僅開放以下兩種情形可以補救，且只能改code
    - script 無法執行
    - report 有圖片無法被重現

Reminder

- 使用 Python 的 os 套件來列出資料夾中所有 images 時，套件並不保證列出來的順序。若沒有先將檔名們 sort 過，到了助教電腦上就可能會讀到錯誤的圖片，畫出來的也就不是你 report 上的圖片
- 記得善用套件的 fix random seed 功能
    - How to fix Lime’s random seed

FAQ

- 若有其他問題，請寄信至助教信箱，**請勿直接私訊助教**。
- 寄信時務必以 [HW5] 作為主旨開頭，否則不會被點開閱讀或回覆。
- 請不要心存僥倖，任何不確定會不會被扣分的細節，都歡迎寄信詢問。死線前詢問絕對每封都回覆，死線後是否違規完全由助教認定。
- 助教信箱：ntu-ml-2020spring-ta@googlegroups.com

Links

- Tutorial：colab / Python script
- checkpoint 上傳與下載
- Report template