

---

---

# Machine Learning HW3

MLTAs

mlta2019fall@gmail.com

---

---

# Outline

- HW3 - Image Sentiment Classification
  - Dataset and Tasks Description
  - Sample Submission
- Kaggle
- Grading / Assignment Regulation

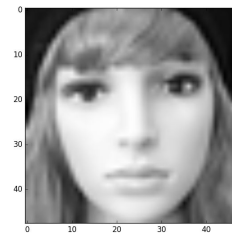
# Task - Image Sentiment Classification



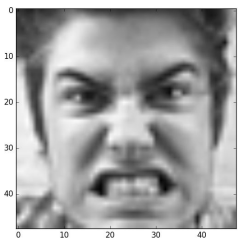
# Task - Image Sentiment Classification

本次作業為網路上收集到的人臉表情資料，

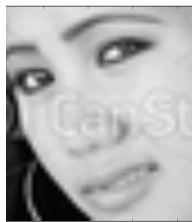
經過特殊處理，每張圖片，均是人臉部份佔大部分



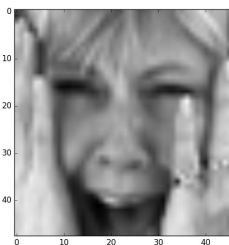
6(中立)



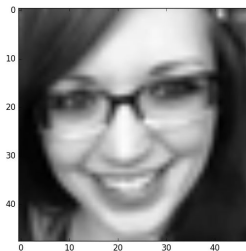
0(生氣)



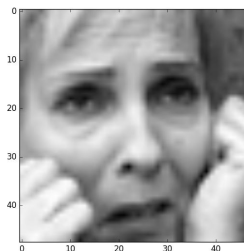
1(厭惡)



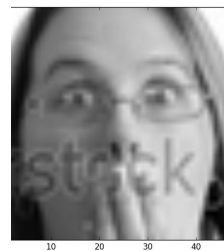
2(恐懼)



3(高興)



4(難過)



5(驚訝)

training data: about 29000 images

testing data: 7000 images (a half will be public test)

# Task and Dataset

- Task : **CNN**
  - Build your own model
  - Use pretrained model (ex : VGG, ResNet....)
- Dataset : Facial Expression Recognition Challenge (Fer2013)

Reference : [Challenges in Representation Learning](#)

# Kaggle Info & Deadline

- Link: [ML2019fall HW3 Image sentiment prediction](#)
- 個人進行、不須組隊
- Team Name:
  - 修課學生: 學號\_任意名稱 (ex: b09901666\_大助什麼都會)
  - 旁聽: 旁聽\_任意名稱
- Maximum Daily Submission: 5 times
- Kaggle Deadline: 11/7/2019 23:59:59 (GMT+8)
- Github Deadline: 11/11/2019 23:59:59 (GMT+8)
- This HW has no **simple baseline**
- test set的7000筆資料將被分為兩份, 3500筆public, 3500筆private
- Leaderboard上所顯示為public score, 在Kaggle Deadline前可以選擇2份submission作為private score的評分依據。

# Kaggle submission format

請預測test set中七千筆資料並將結果上傳Kaggle

1. 上傳格式為csv。
2. 第一行必須為id,label, 第二行開始為預測結果。
3. 每行分別為id以及預測的label, 請以逗號分隔。
4. Evaluation: Accuracy

| 1  | id, label |
|----|-----------|
| 2  | 0,0       |
| 3  | 1,0       |
| 4  | 2,0       |
| 5  | 3,0       |
| 6  | 4,0       |
| 7  | 5,0       |
| 8  | 6,0       |
| 9  | 7,0       |
| 10 | 8,0       |
| 11 | 9,0       |
| 12 | 10,0      |
| 13 | 11,0      |
| 14 | 12,0      |
| 15 | 13,0      |
| 16 | 14,0      |
| 17 | 15,0      |
| 18 | 16,0      |
| 19 | 17,0      |
| 20 | 18,0      |
| 21 | 19,0      |

# 配分 Grading Criteria-Kaggle(2%)

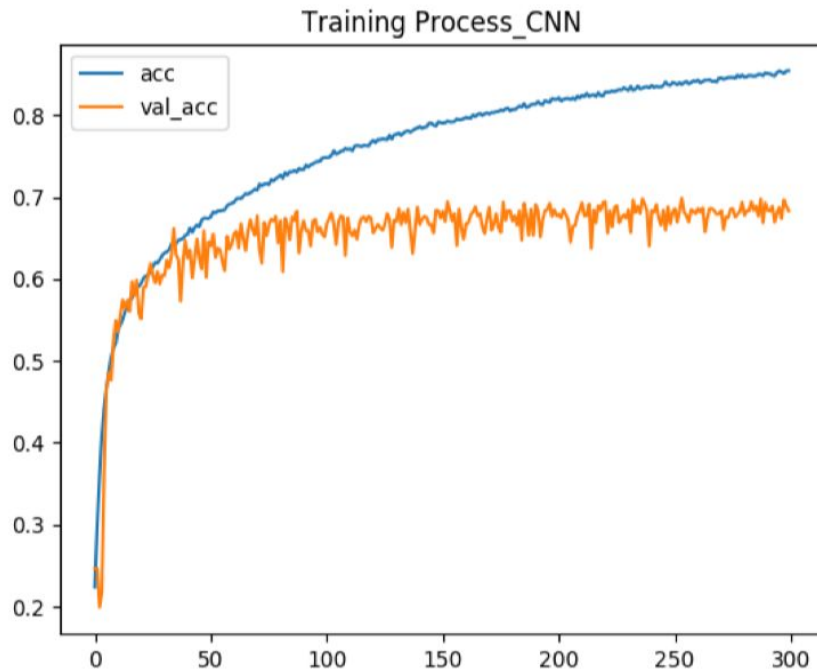
- Kaggle deadline : 11/7/2019 23:59:59 (GMT+8)
- Kaggle - 2%
  - ❑ 超過public leaderboard的strong baseline分數 : 1%
  - ❑ 超過private leaderboard的strong baseline分數 : 1%
- Bonus - 1%
  - (1.0%) private leaderboard 排名前五名且於助教時間上台分享的同學



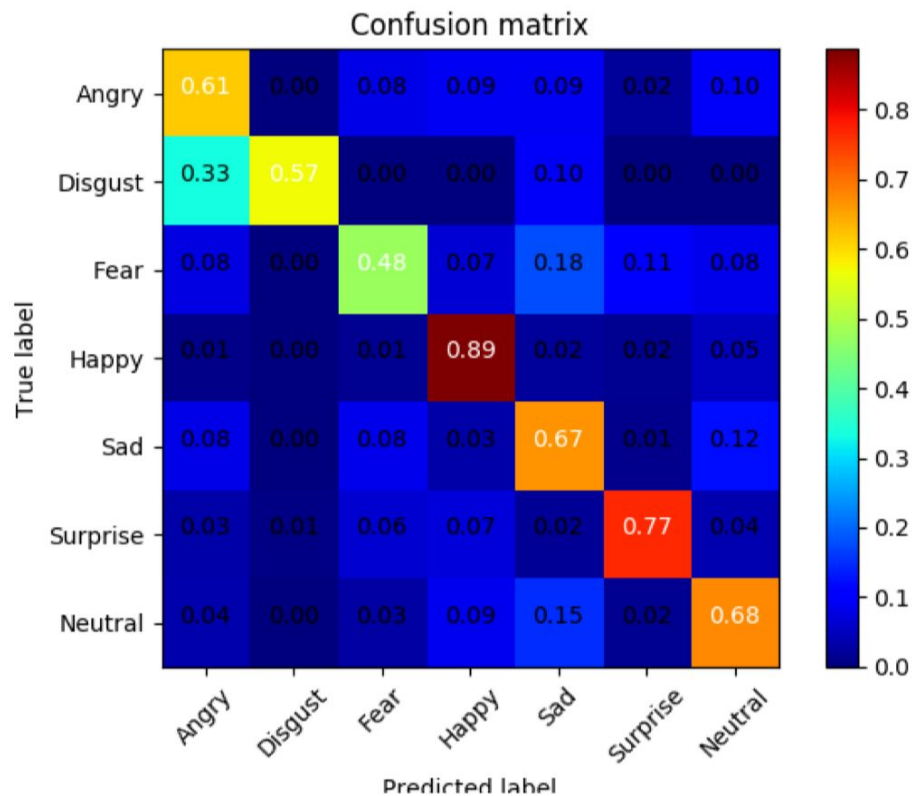
# 配分 Grading Criteria - report(8%)

- Programming Report - 5%
  - [https://docs.google.com/document/d/1ROAFfY7AuVoCPOIpbMCQizwGTj-YFKS2JE\\_39XJ2Wuc/edit](https://docs.google.com/document/d/1ROAFfY7AuVoCPOIpbMCQizwGTj-YFKS2JE_39XJ2Wuc/edit)
- Math Problem - 3%
  - [\[Hackmd\]](#)
  - Type in latex(preferable) or take pictures of your handwriting
- Write them in report.pdf

# Example - training process



# Example - confusion matrix

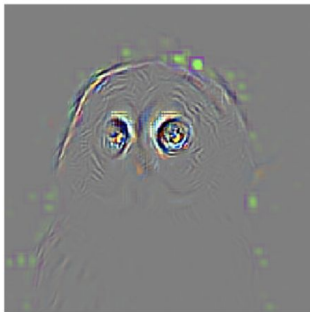


# Example - saliency map

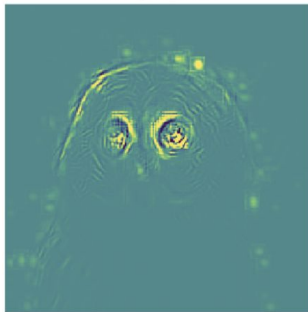
Input image



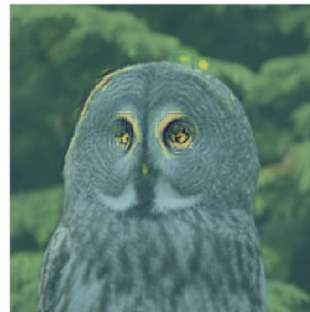
Gradients across RGB channels



Max gradients



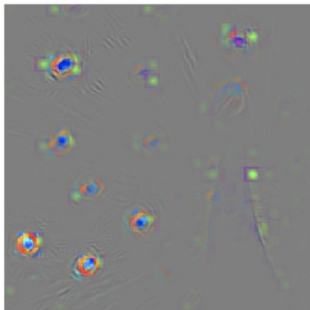
Overlay



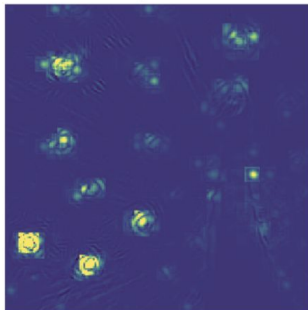
Input image



Gradients across RGB channels



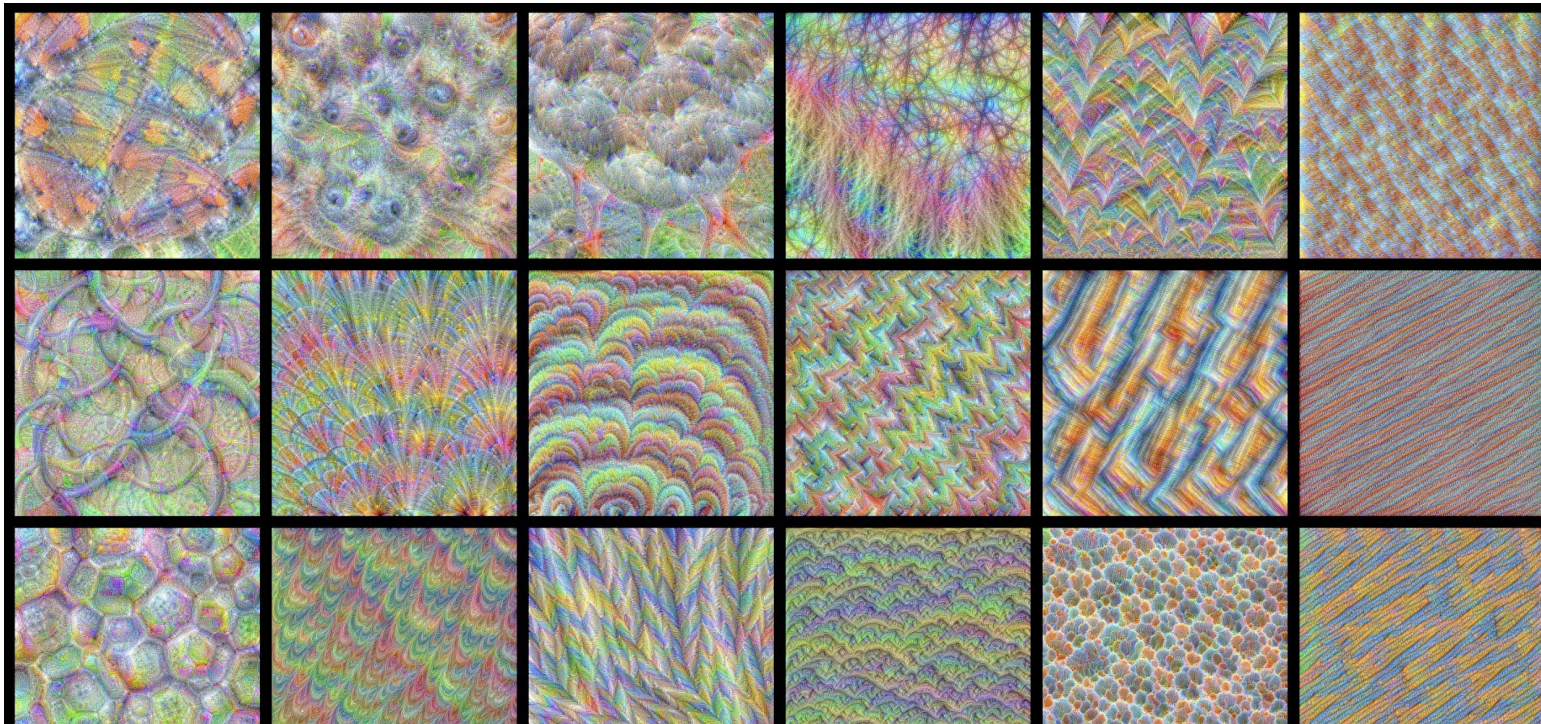
Max gradients



Overlay



# Example- filters visualization



# Assignment Regulation

- Only Python 3.6 available !!!!
- 開放使用套件
  - numpy  $\geq 1.14$
  - pandas  $\geq 0.24.1$
  - python standard library
  - pytorch  $\geq 1.2.0$  (torchvision  $= 0.4.0$ )
  - tensorflow  $= 1.14.0$
  - keras  $= 2.2.4$
  - cv2
  - pillow  $\geq 6.1.0$
- 若需使用其他套件，請儘早寄信至助教信箱詢問，並請闡明原因。



# Requirements

1. 請使用CNN
2. 不能使用額外data
3. 請附上訓練好的model (及其參數)至dropbox, 並於hw3\_test.sh中寫下載的command
4. hw3\_test.sh要在10分鐘內跑完

# Github Submissions

你的github上ML2019FALL/hw3/裡面請至少包含：

1. **report.pdf** : Please refer to report template
2. **hw3\_train.sh** : how to train your model
3. **hw3\_test.sh** : meet the highest score you choose in kaggle
4. your python files
5. model參數 (Make sure it can be downloaded by your script.)

請不要上傳dataset, 請不要上傳dataset, 請不要上傳dataset



# Shell script

助教在批改程式部分時，會執行以下指令：

- `bash hw3_train.sh <training data folder> <train.csv>`
  - training data folder : 包含所有的training image
  - train.csv : train image label 的路徑
- `bash hw3_test.sh <testing data folder> <prediction.csv>`
  - testing data folder : 包含所有的testing image
  - prediction.csv : 預測結果的csv檔路徑

# Shell script

- 請務必在訓練過程中，隨時存取參數
- test data會shuffle過，請勿直接輸出事先存取的答案
- hw3\_test.sh需要在10分鐘內執行完畢，否則該部分將以0分計算
- 切勿於程式內寫死輸入檔案是output file的路徑，否則該部分將以0分計算
- [Conda file](#) (同學可自行下載改prefix測試)

# Report 格式

- 限制
  - 檔名必須為 report.pdf !!!
  - 檔名必須為 report.pdf !!!
  - 檔名必須為 report.pdf !!!
  - 請用中文撰寫report(非中文母語者可用英文)
  - 請標明系級、學號、姓名，並按照report模板回答問題，切勿隨意更動題號順序
  - 若有和其他修課同學討論，請務必於題號前標明 collaborator(含姓名、學號)
- Report模板連結
  - 連結：[Link](#)
- 截止日期同 Github Deadline: 11/11/2019 23:59:59 (GMT+8)

## 其他規定 Other Policy

- Lateness
  - Github每遲交一天(不足一天以一天計算) hw3所得總分將x0.7
  - 不接受程式 or 報告單獨遲交
  - 不得遲交超過一天, 若有特殊原因請儘速聯絡助教
  - Github遲交表單: 遲交請先上傳遲交檔案至自己的github後再填寫遲交表單, 助教群會以表單填寫時間作為繳交時間手動 clone檔案。
  - 表單連結: [Link](#) (遲交才必需填寫)
- Script Error
  - 當script格式錯誤, 造成助教無法順利執行, 請在公告時間內寄信向助教說明, 修好之後重新執行所得kaggle部分分數將x0.5。
  - 可以更改的部分僅限 syntax及io的部分, 不得改程式邏輯或是演算法, 至於其他部分由助教認定為主。
  - 只能在助教面前更改你的 script。

# 其他規定 Other Policy



- Cheating

- 抄code、抄report (含之前修課同學)
- 開設kaggle多重分身帳號註冊competition
- 於訓練過程以任何不限定形式接觸到testing data的正確答案
- 填寫前人的github repo url
- 不得上傳之前的kaggle競賽
- 教授與助教群保留請同學到辦公室解釋coding作業的權利, 請同學務必自愛

# TA Hour

- 11/1助教課 手把手教學
- 10/29, 11/5 (Tue) @BL530
- AM09:20~PM12:00