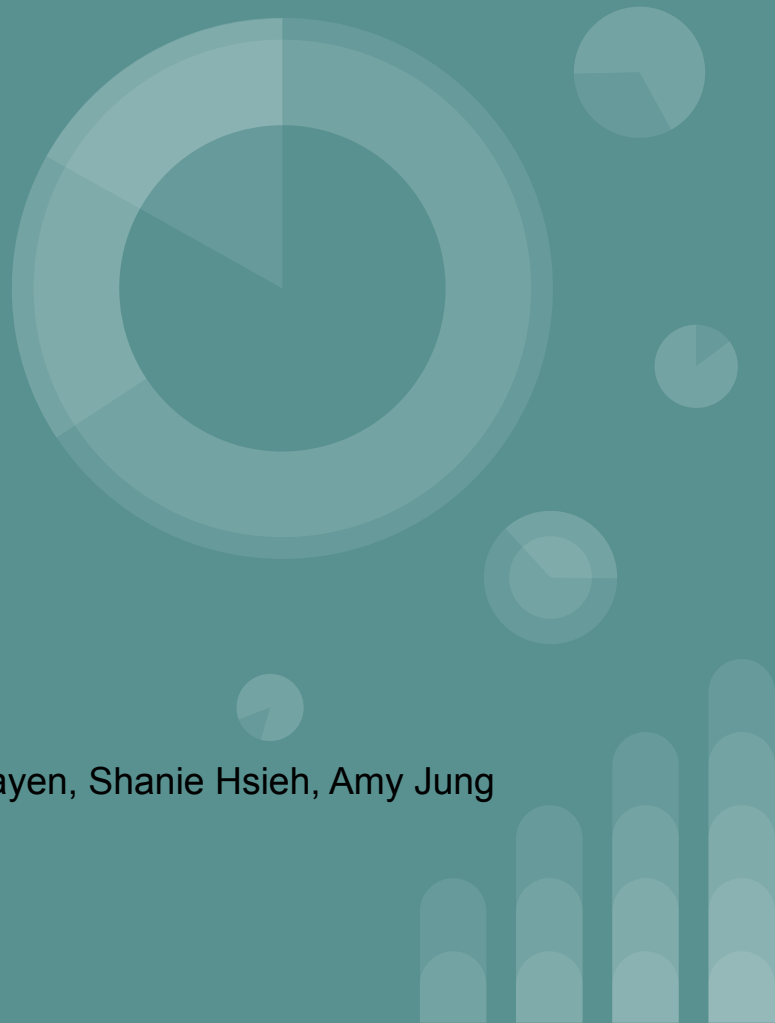


# Facial Recognition

W207 Final Project

Jeremy Yeung, Simran Sachdev, Gabriel Louis-Kayen, Shanie Hsieh, Amy Jung





**What is the question you will be  
working on?**

**Why is it interesting?**





# What is facial detection?

Our project focuses on **Facial Keypoints Detection**, detecting and predicting the location of keypoints on face images — the fundamental building block for various applications including:

- Tracking faces in images and videos
- Analyzing facial expressions
- Detecting dysmorphic facial signs for medical diagnosis
- Biometrics / face recognition.



**What is the data that you  
will be using?**



### **Data source:**

<https://www.kaggle.com/c/facial-keypoints-detection/overview>

The dataset is Facial Keypoints Detection data used to detect the location of keypoints on face images.

### **Size of dataset:**

The dataset has 7049 images that each have 30 columns variables.

### **Main features used:**

There are 30 columns associated with 15 features.

#### **The 15 features:**

- left\_eye\_center,
- right\_eye\_center,
- left\_eye\_inner\_corner,
- left\_eye\_outer\_corner,
- right\_eye\_inner\_corner,
- right\_eye\_outer\_corner,
- left\_eyebrow\_inner\_end,
- left\_eyebrow\_outer\_end,
- right\_eyebrow\_inner\_end,
- right\_eyebrow\_outer\_end
- nose\_tip
- mouth\_left\_corner
- mouth\_right\_corner
- mouth\_center\_top\_lip,
- mouth\_center\_bottom\_lip

Each of the 15 features has an x-axis column and a y-axis column corresponding to that feature's location on the image, leading to 30 columns overall.



# Considerations

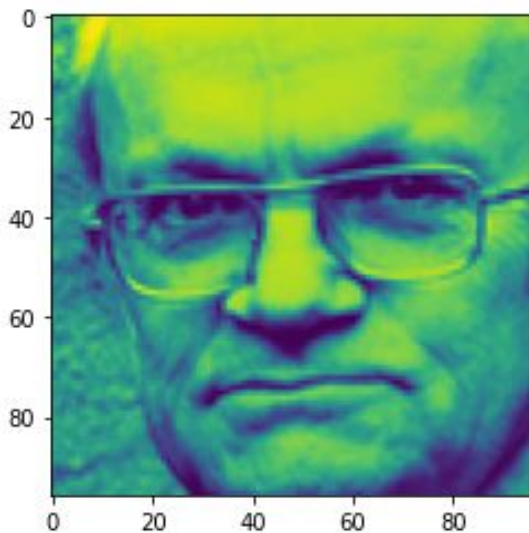
Using the 2000+ images with 15 features

- Higher train accuracy
- Prone to overfitting

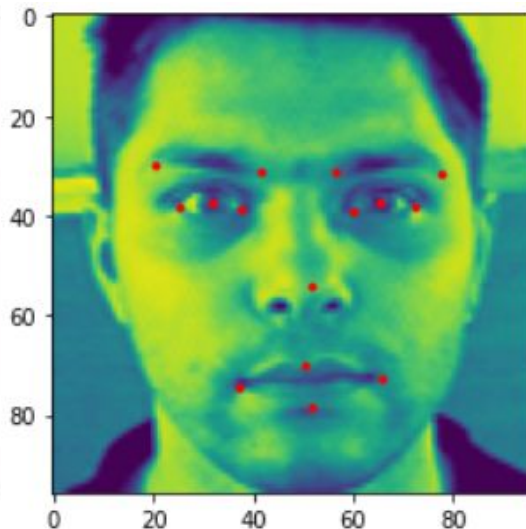
Using the 7000+ images with 4 features

- Easier to generalize
- More training data
- Less accuracy in identifying facial keypoints

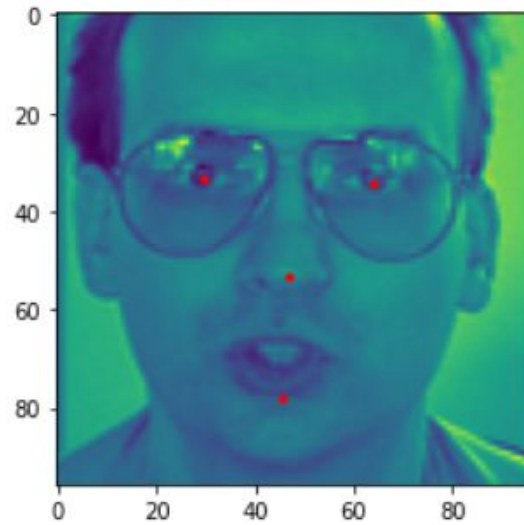
# Example Images and Facial Key Points



Example Image



Example Facial Key Points (15 features)



Example Facial Key Points (4 features)



# Summary statistics

7000 images have data for 4  
facial key points

|                           | count  | mean               | std                |
|---------------------------|--------|--------------------|--------------------|
| left_eye_center_x         | 7000.0 | 66.34940047635854  | 3.377149279603647  |
| left_eye_center_y         | 7000.0 | 37.61810350559417  | 3.0365916551708043 |
| right_eye_center_x        | 7000.0 | 30.303406587113066 | 2.9489464986052036 |
| right_eye_center_y        | 7000.0 | 37.94265611325309  | 2.884111354058055  |
| nose_tip_x                | 7000.0 | 48.372452384140466 | 4.1715876082560435 |
| nose_tip_y                | 7000.0 | 62.68202743453441  | 5.621674878670501  |
| mouth_center_bottom_lip_x | 7000.0 | 48.57167648140966  | 4.237941397514037  |
| mouth_center_bottom_lip_y | 7000.0 | 78.97570952261637  | 5.407682797359353  |

Of those 7000 images, only 2140  
images of those have data for all 15  
facial key points

|                           | count  | mean               | std                |
|---------------------------|--------|--------------------|--------------------|
| left_eye_center_x         | 2140.0 | 66.22154868409592  | 2.087683355101556  |
| left_eye_center_y         | 2140.0 | 36.842274165726266 | 2.294027490805707  |
| right_eye_center_x        | 2140.0 | 29.64026856456148  | 2.051575209871264  |
| right_eye_center_y        | 2140.0 | 37.06381489055456  | 2.2343335854467448 |
| left_eye_inner_corner_x   | 2140.0 | 59.27212810062244  | 2.005630683413952  |
| left_eye_inner_corner_y   | 2140.0 | 37.85601445389234  | 2.03450012751805   |
| left_eye_outer_corner_x   | 2140.0 | 73.41247343419627  | 2.701639370765223  |
| left_eye_outer_corner_y   | 2140.0 | 37.6401096830805   | 2.68416217097158   |
| right_eye_inner_corner_x  | 2140.0 | 36.6031065182916   | 1.8227836818129908 |
| right_eye_inner_corner_y  | 2140.0 | 37.92085164154391  | 2.0095047120860614 |
| right_eye_outer_corner_x  | 2140.0 | 22.36161709895906  | 2.7688040797668125 |
| right_eye_outer_corner_y  | 2140.0 | 38.03457131359977  | 2.654902542892582  |
| left_eyebrow_inner_end_x  | 2140.0 | 56.14799092743679  | 2.819913666924865  |
| left_eyebrow_inner_end_y  | 2140.0 | 29.22230444909996  | 2.8671313510347325 |
| left_eyebrow_outer_end_x  | 2140.0 | 79.61752316513792  | 3.3126467711070138 |
| left_eyebrow_outer_end_y  | 2140.0 | 29.65657017639958  | 3.627186873003011  |
| right_eyebrow_inner_end_x | 2140.0 | 39.27208385866163  | 2.6096476570044818 |
| right_eyebrow_inner_end_y | 2140.0 | 29.41374657993314  | 2.8422186447220557 |
| right_eyebrow_outer_end_x | 2140.0 | 15.76170725407129  | 3.3379012928231457 |
| right_eyebrow_outer_end_y | 2140.0 | 30.452946698618238 | 3.6443422006653514 |
| nose_tip_x                | 2140.0 | 47.95214068998041  | 3.276053208468195  |
| nose_tip_y                | 2140.0 | 57.25392567086902  | 4.528635210886218  |
| mouth_left_corner_x       | 2140.0 | 63.419076094887814 | 3.650131009318928  |
| mouth_left_corner_y       | 2140.0 | 75.88765965132447  | 4.438565027075064  |
| mouth_right_corner_x      | 2140.0 | 32.96736460044271  | 3.5951027258262207 |
| mouth_right_corner_y      | 2140.0 | 76.13406536660167  | 4.259513821121693  |
| mouth_center_top_lip_x    | 2140.0 | 48.081324634435525 | 2.7232735346715224 |
| mouth_center_top_lip_y    | 2140.0 | 72.6811245530104   | 5.108675344728991  |
| mouth_center_bottom_lip_x | 2140.0 | 48.1496539871852   | 3.032388960435935  |
| mouth_center_bottom_lip_y | 2140.0 | 82.63041245065179  | 4.813557334126184  |





**What prediction algorithm  
do you plan to use?**



# Linear Regression

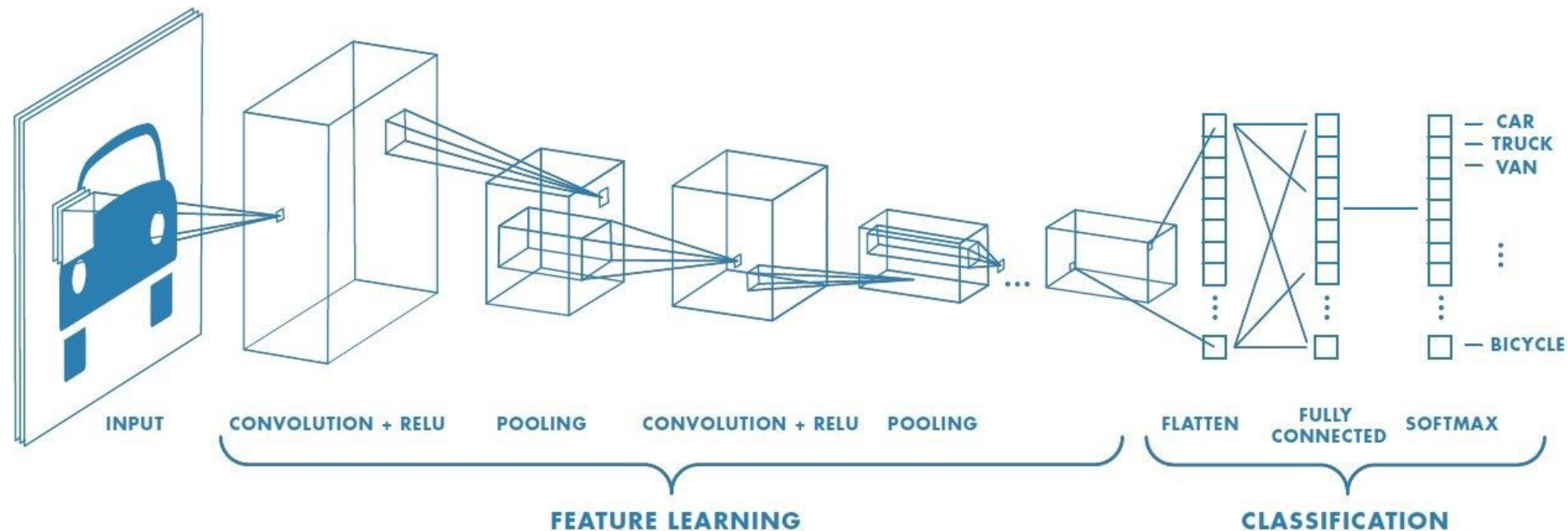
- N different multiple linear regressions, where N is the number of facial keypoints (8 or 30)
- Each pixel is a feature
- Continuous Output

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\theta}$$

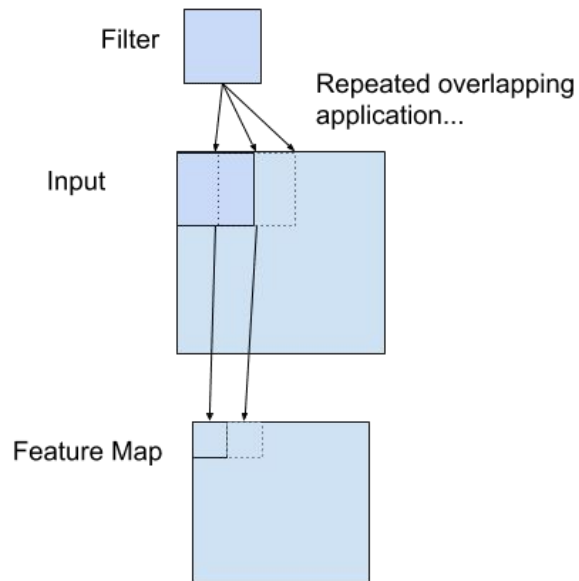
$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_p \end{bmatrix}$$

# Convolutional Neural Network

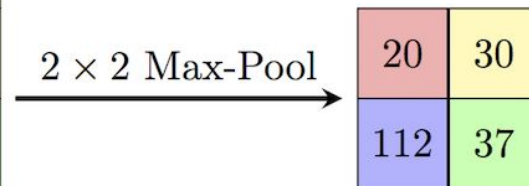
Outputs 8 or 30 x- or y-coordinate values indicating the location of a facial keypoint



# Convolution and Pooling



|     |     |    |    |
|-----|-----|----|----|
| 12  | 20  | 30 | 0  |
| 8   | 12  | 2  | 0  |
| 34  | 70  | 37 | 4  |
| 112 | 100 | 25 | 12 |





**How will you evaluate  
your results?**





# Metrics: Mean Squared Error

MSE measures of goodness of fit

- Works well for both of our models
- We will take the average of the square of the difference between the original values and the predicted values
- Larger mistakes are more pronounced than smaller mistakes

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

# Questions?

