Face Recognition

Raj, Tom, Donna, Olga, Sahil, Aditya

October 10, 2017

Introduction - MegaFace

- Face recognition and verification problem
- Current Face recognition

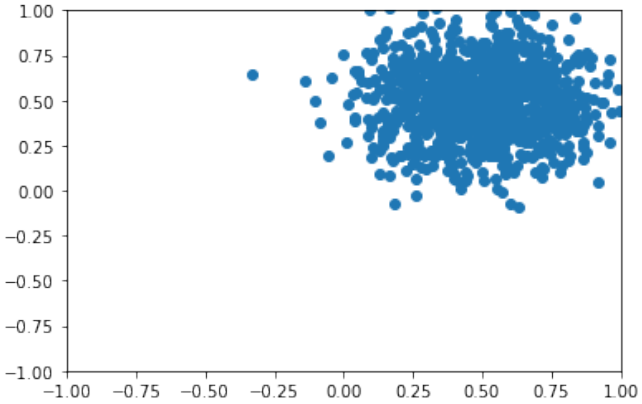
VGG Very Deep 16 Architecture (Input)

Images: $\mathbf{I} = \{i_1, \cdots, i_N\} \in \mathbb{R}^{3\times224\times224}$
Input to Network: $\mathbf{X} = \{x_1, \cdots, x_N\}, x_i \in \mathbb{R}^{3\times224\times224}$

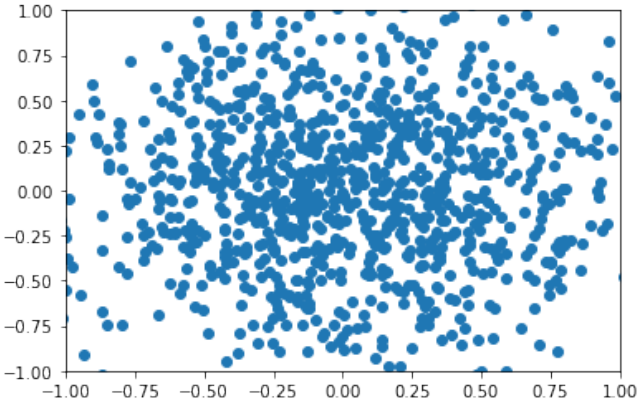
$$x_i = i_i - \frac{\sum_{j=1}^{N} i_j}{N} \tag{1}$$

Makes faces more separable after normalization.

VGG Very Deep 16 Architecture (Face Normalization)



(a) biased faces



(b) normalized faces

Figure 1: Subtracting the average face from all of the input images makes it easier to separate each image in space after normalization

VGG Very Deep 16 Architecture (Function Review)

Convolution (Padding: 1, Stride: 1):

$$f(x) = K^j \circledast x + diag(b^j)\mathbf{1}_{c_j \times n_j \times n_j} \in \mathbb{R}^{c_j \times n_j \times n_j} \tag{2}$$

with

$$K^j \in \mathbb{R}^{c_j \times \hat{c}_j \times (2k+1) \times (2k+1)}, \quad b^j \in \mathbb{R}^{c_j} \tag{3}$$

and

$$\mathbf{1}_{c_j \times n_j \times n_j} = 1_{c_j} \otimes 1_{n_j} \otimes 1_{n_j}, \quad k = 0, 1, 2, 3, ... \tag{4}$$

Relu:

$$g(x) = \begin{cases} 0 & x \leq 0, \\ x & x > 0. \end{cases} \tag{5}$$

VGG Very Deep 16 Architecture (Function Review)

Softmax (Padding: 0, Stride: 2):

$$p = \sigma(y), y \in \mathbb{R}^c \tag{6}$$

$$p_j = \frac{e^{y_j}}{\sum_{i=1}^{c} e^{y_i}}, j = 1 : c \tag{7}$$

$$p_j \geq 0, \sum_{j=1}^{c} p_j = 1 \tag{8}$$

Max Pooling:

$$r : \mathbb{R}^{a \times b \times b} \rightarrow \mathbb{R}^{a \times \frac{b}{2} \times \frac{b}{2}} \tag{9}$$

$$r(X)_{i,j} = \max_{-1 \leq k,l \leq 0} X_{2i+k,2j+1}, \quad i,j = 1 : \frac{b}{2} \tag{10}$$

VGG Very Deep 16 Architecture (Encoding)

Layer	Operation	Filter	Dimension	Filters	Size
0	$K_0 \circledast x$	3x3	3 (r.g.b.)	64	$64 \times 224 \times 224$
1	$K_1 \circledast g \circ f^0$	3x3	64	64	$64 \times 224 \times 224$
2	$K_2 \circledast r \circ g \circ f^1$	3x3	64	128	$128 \times 112 \times 112$
3	$K_3 \circledast g \circ f^2$	3x3	128	128	$128 \times 112 \times 112$
4	$K_4 \circledast r \circ g \circ f^3$	3x3	128	256	$256 \times 56 \times 56$
5	$K_5 \circledast g \circ f^4$	3x3	256	256	$256 \times 56 \times 56$
6	$K_6 \circledast g \circ f^5$	3x3	256	256	$256 \times 56 \times 56$
7	$K_7 \circledast r \circ g \circ f^6$	3x3	256	512	$512 \times 28 \times 28$
8	$K_8 \circledast g \circ f^7$	3x3	512	512	$512 \times 28 \times 28$
9	$K_9 \circledast g \circ f^8$	3x3	512	512	$512 \times 28 \times 28$
10	$K_{10} \circledast r \circ g \circ f^9$	3x3	512	512	$512 \times 14 \times 14$
11	$K_{11} \circledast g \circ f^{10}$	3x3	512	512	$512 \times 14 \times 14$
12	$K_{12} \circledast g \circ f^{11}$	3x3	512	512	$512 \times 14 \times 14$
13	$K_{13} \circledast r \circ g \circ f^{12}$	7x7	512	4096	$4096 \times 7 \times 7$

VGG Very Deep 16 Architecture (Decoding)

Layer	Operation	Input Size	Output Size
14	$\theta_{14} \circ g \circ f^{13}$	$4096 \times 7 \times 7$	4096
15	$\theta_{15} \circ g \circ f^{14}$	4096	2622
16	$\sigma \circ f^{15}$	2622	2622

Transfer Learning in the Context of Facial Recognition

- K-Class classification
  - $p \in \mathbb{R}^{2262}, p_i \in [0, 1]$
- Internal Face Representation (f^{13} from above layers)
  - 7×7 representations instead of 224×224 image pixels
  - Reduce noise
  - Learn to extract important facial features
- Tune Model for Verification
  - Triplet Loss
  - Siamese Loss
  - L2 Loss

Triplet Loss Function

- Show Math and motivation for Triplet Loss function.
- You should have the equation and explain in Dr. Xu's notation
- You should explain why picking good matches helps

Implementation and Results

- Put Results Here