16-720 Computer Vision Term Project

An Exploration in Face Verification

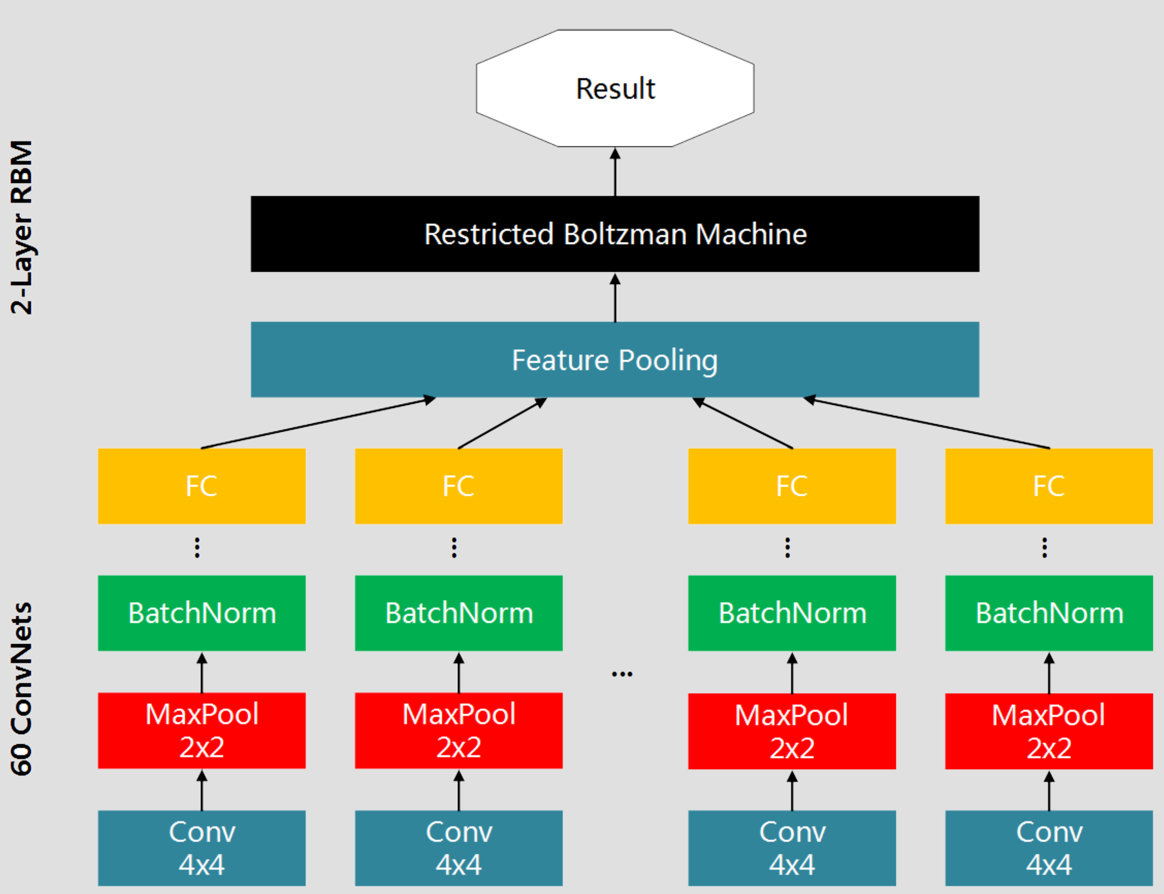
Wentao Han (whan1@andrew.cmu.edu) Hang Yuan (hy1@andrew.cmu.edu) Yunjin Wu (yunjinw@andrew.cmu.edu)

Introduction

General face recognition includes both verification and identification. In specific, face verification is the approach that tries to match a known face with an unknown one whereas face identification matches two known faces. In our project, we implemented two state-of-art face verification approaches using hybrid CNN model[1] and joint Bayesian techniques[2] respectively, and then compared their performances on the LFW dataset.

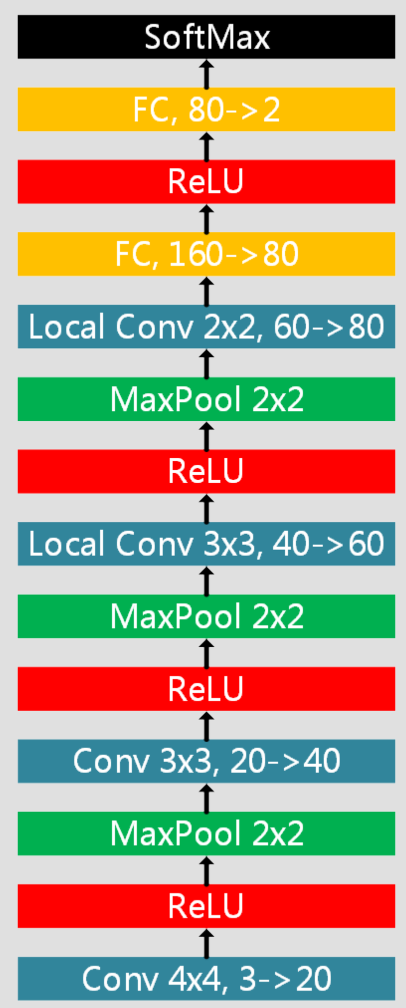
Deep Learning Model

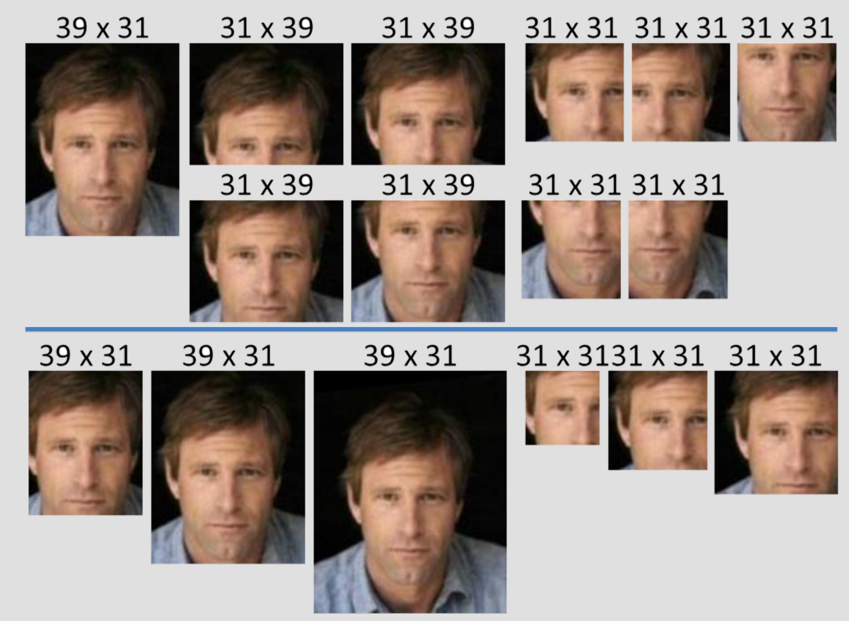
Model Architecture



- 60 ConvNets serve as complementary and efficient feature extractors
- 2 stages of average pooling to reduce feature dimensionality
- A 2-layer RBM for classification

ConvNet Architecture





- Proper kernel size and channel number at each stage of convolution to guarantee model capacity and efficiency
- The weights of the last two convolution layers are locally shared to capture high-level features of structured objects such as face
- Each of the 60 ConvNets is trained with images from different facial regions that vary in location, scale and color type
- Each type of the facial regions is centered at one of the five facial landmarks detected with the work of [3]

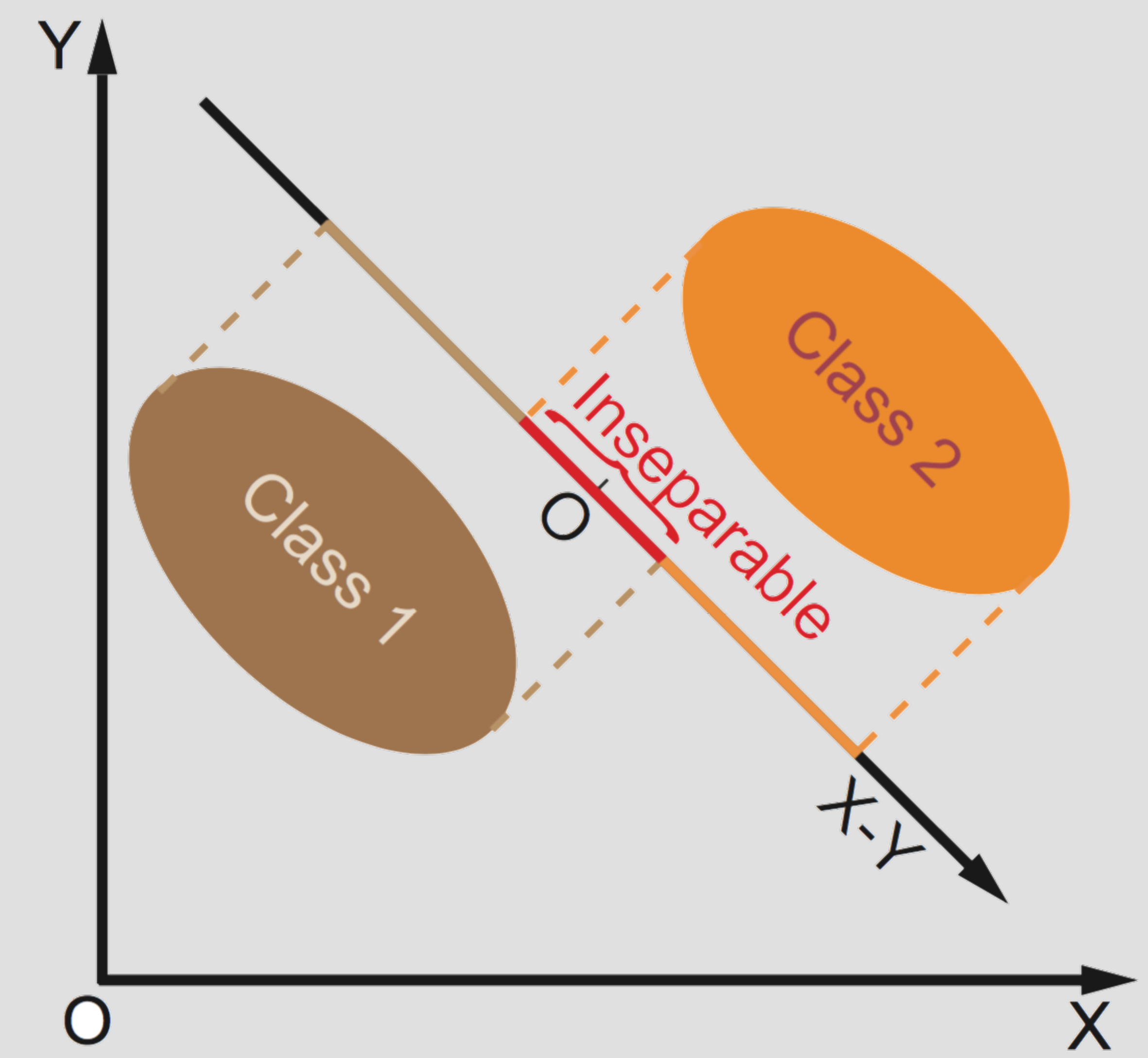
Classification RBM

- Given input, the probability of its output can be explicitly expressed as

$$p(y_c \mid x) = \frac{e^{d_c} \prod_j \left(1 + e^{c_j + U_{jc} + \sum_k W_{jk} x_k}\right)}{\sum_i e^{d_i} \prod_j \left(1 + e^{c_j + U_{ji} + \sum_k W_{jk} x_k}\right)}$$

- The 2-layer RBM is trained with features extracted by the ConvNets
- Top-down fine-tuning of the entire hybrid model is not performed due to limited computation resources

Joint Bayesian Model



Problem description

Traditional approaches project 2D data to 1-D such that the separable joint representation becomes inseparable.

Face formation:

$$x = \mu + \varepsilon$$

$$\mu \sim N(0, S_\mu) \quad \text{and} \quad \varepsilon \sim N(0, S_\varepsilon)$$

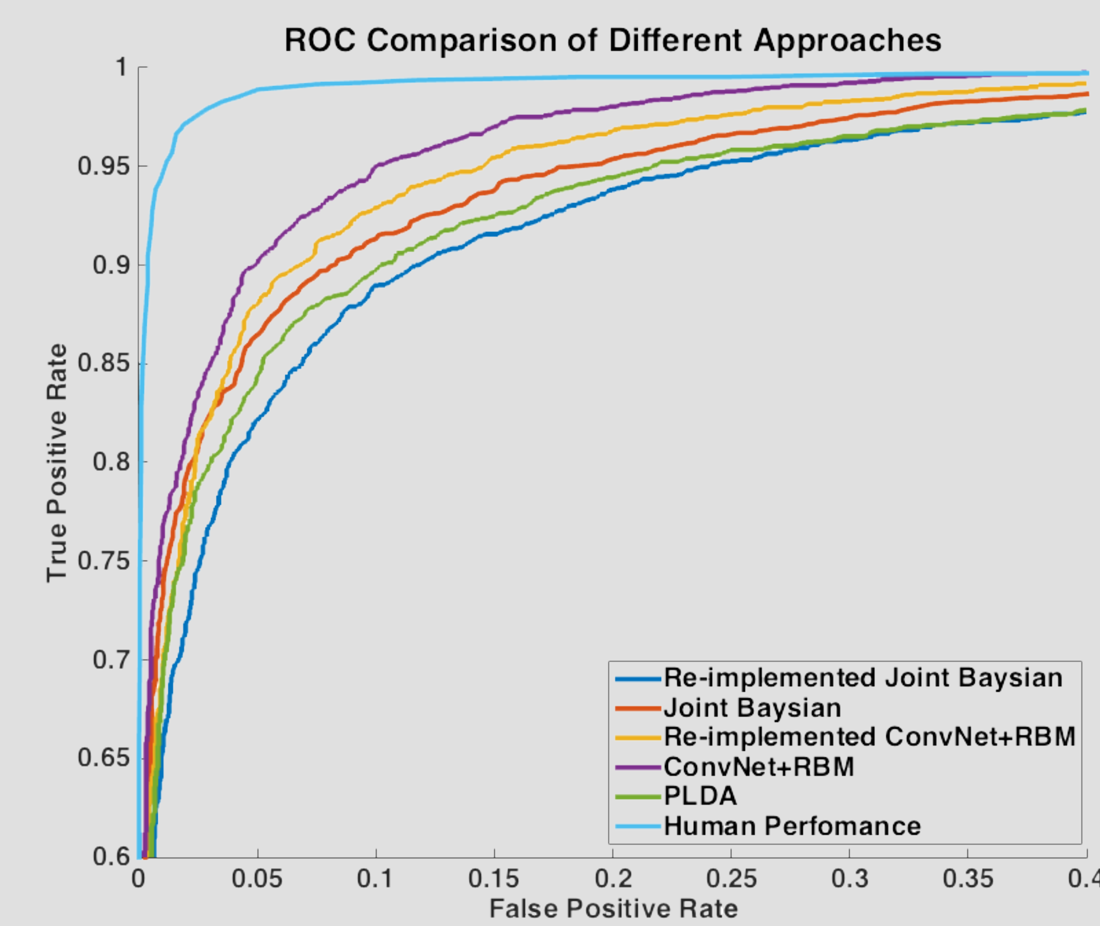
- Each face has two latent variables, identity μ and inter-personal variations ε.
- These two variables follow two Gaussian distributions with mean of 0.

Output generation:

$$r(x_1, x_2) = \log\left(\frac{P(x_1, x_2 \mid H_I)}{P(x_1, x_2 \mid H_E)}\right)$$

- HI hypothesizes these two images belong to the same person in the intra-personal representation. HE hypothesizes these two images belong to different people in the extra-personal representation. The two covariances matrices parameters are computed using an EM-like algorithm.

Results



Method	Accuracy(%)
Re-implemented Joint Baysian	89.82%
Joint Baysian	90.90%
Re-implemented ConvNet-RBM	92.08%
ConvNet-RBM	93.83%
PLDA	87.33%
Human Performance	97.53%

- Evaluation follows the unrestricted protocol of the LFW dataset
- Competitive results are reached compared to the original works and other state-of-art methods

References

1. Yi Sun, Xiaogang Wang, and Xiaoou Tang. Hybrid deep learning for face verification. In Proceedings of the IEEE International Conference on Computer Vision, pages 1489–1496, 2013.
2. Chen, Dong, et al. "Bayesian face revisited: A joint formulation." European Conference on Computer Vision. Springer Berlin Heidelberg, 2012.
3. Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.