YI SUN, XIAOGANG WANG, XIAOOU TANG

2014 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION

DEEP LEARNING FACE REPRESENTATION FROM PREDICTING 10,000 CLASSES

Z.SAYGIN DOĞU

OUTLINE

- Introduction
- Related Work
- Deep Convolution Nets and Feature Extraction
- Face Verification
  - ➢ Joint Bayesian
  - ➢ Neural Network
- Experiments
- Results

INTRODUCTION

- **Yi Sun:** Department of Information Engineering, The Chinese University of Hong Kong
- **Xiaogang Wang:** Department of Electronic Engineering, The Chinese University of Hong Kong
- **Xiaoou Tang:** Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

INTRODUCTION

- *Face Verification*
- ConvNets for feature extraction
- *Patches*

RESEARCH BY AUTHORS ON FACE VERIFICATION

- Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
- **89.60%**
- Sun, Y., Wang, X. and Tang, X. (2014). Deep Learning Face Representation from Predicting 10,000 Classes. 2014 IEEE Conference on Computer Vision and Pattern Recognition.
- **97.45%**
- Sun, Yi, et al. "Deep learning face representation by joint identification-verification." *Advances in neural information processing systems*. 2014.
- **99.15%**
- Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." *arXiv preprint arXiv:1502.00873* (2015).
- **99.53%**

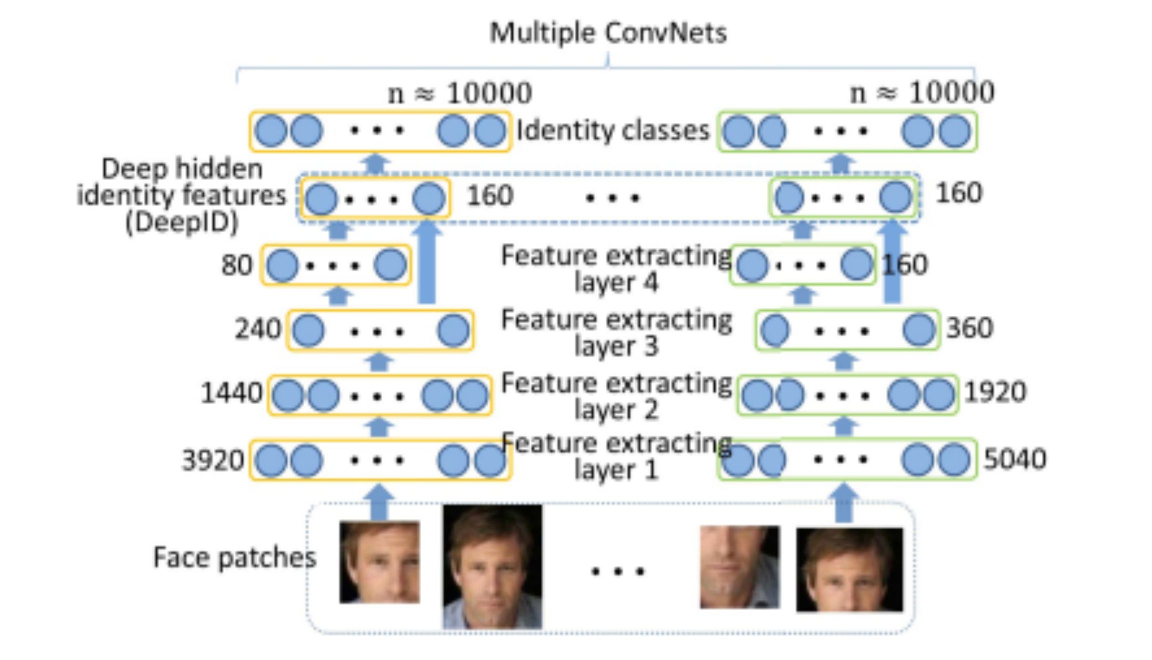
RELATED WORK

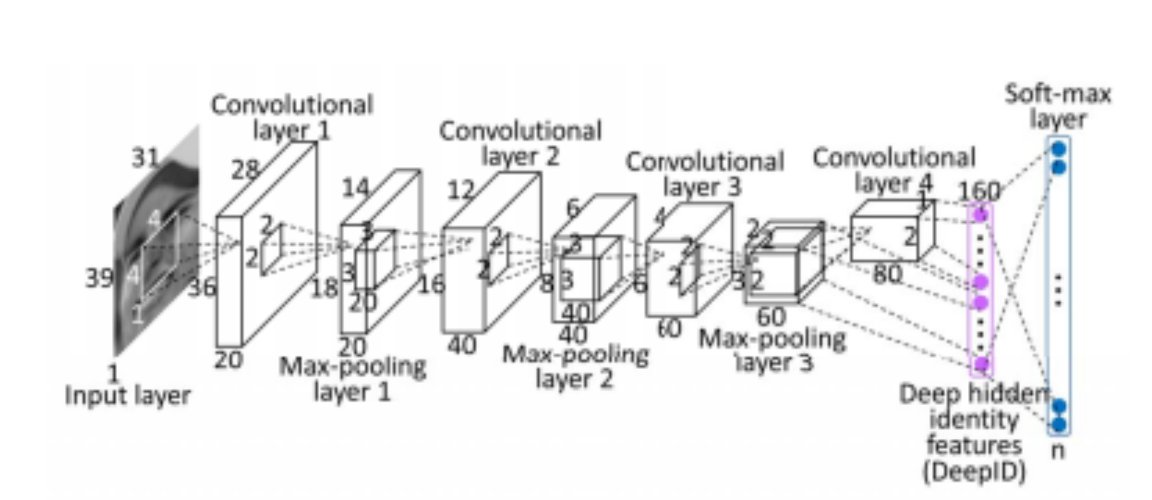
- Generally shallow models
- High dimensional & Low Level Features
  - Learning-based Learning Entropy(LE) descriptors
    - *L*2 distance between the LE descriptors [4]
  - Local Binary Patterns (LBP) descriptors
    - and Joint Bayesian [5] after *PCA.*
  - *Scale Invariant Feature Transform (*SIFT) features into Fisher vectors[6]
- *Deep Models*
  - Convolutional Deep Belief Networks (CDBN) => Information-Theoretic Metric Learning(ITML) and linear SVM
  - Siamese Networks[8]

THIS PAPER

- Feature extraction and recognition is not jointly learned, instead they do it in 2-steps:
  - ConvNet for feature extraction
  - Joint Bayesian or Neural Network for recognition
- Last hidden layer is used as features *not* the output layer.
- Classifying all the identities simultaneously instead of training binary classifiers

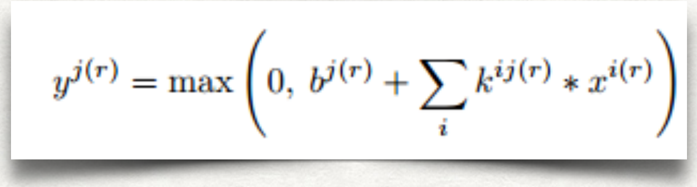
LEARNING DEEP ID FOR FACE VERIFICATION


Multiple ConvNets
n ≈ 10000
n ≈ 10000
Identity classes
Deep hidden identity features (DeepID)
160
160
80
160
Feature extracting layer 4
240
360
Feature extracting layer 3
1440
1920
Feature extracting layer 2
3920
5040
Feature extracting layer 1
Face patches



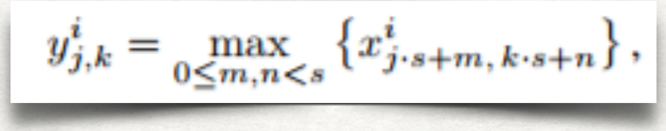
DEEP ID - CONV NET

- Convolution operation:



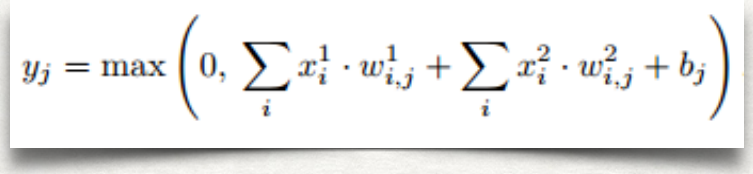
$$y^{j(r)} = \max\left(0,\ b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)}\right)$$

- Max-pooling can be formulated as:



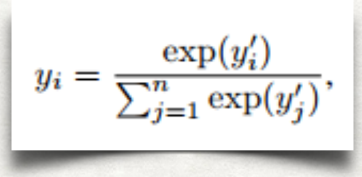
$$y^i_{j,k} = \max_{0\le m,n<s}\left\{x^i_{j\cdot s+m,\,k\cdot s+n}\right\},$$

- ReLU nonlinearity is used for hidden neurons (y = max(0,x)).
- Last hidden layer takes:



$$y_j = \max\left(0,\ \sum_i x^1_i \cdot w^1_{i,j} + \sum_i x^2_i \cdot w^2_{i,j} + b_j\right)$$

- Softmax:



$$y_i = \frac{\exp(y'_i)}{\sum_{j=1}^{n} \exp(y'_j)},$$

LEARNING FEATURES

- From Each image:
- 60 Face Patches
    - 3 Scales
    - 2 Channels (RGB or Gray)
- 60 ConvNets trained for each patch (+ Horizontally Flipped Counterpart)
    - 160 Dimensional Features
- (160x2x60) Dimensional Features

WHAT TO DO WITH THE DEEPID FEATURES

- Train Classifiers using DeepID Features
  - Joint Bayesian
  - Neural Networks

MULTI SCALE CONV NETS

- They used the output of the 4th conv layer as DeepID

- Adding the output of the 3rd conv layer improves validation

  accuracy by 4.72%

  - Actually improves final accuracy from 95.35% to 96.05%

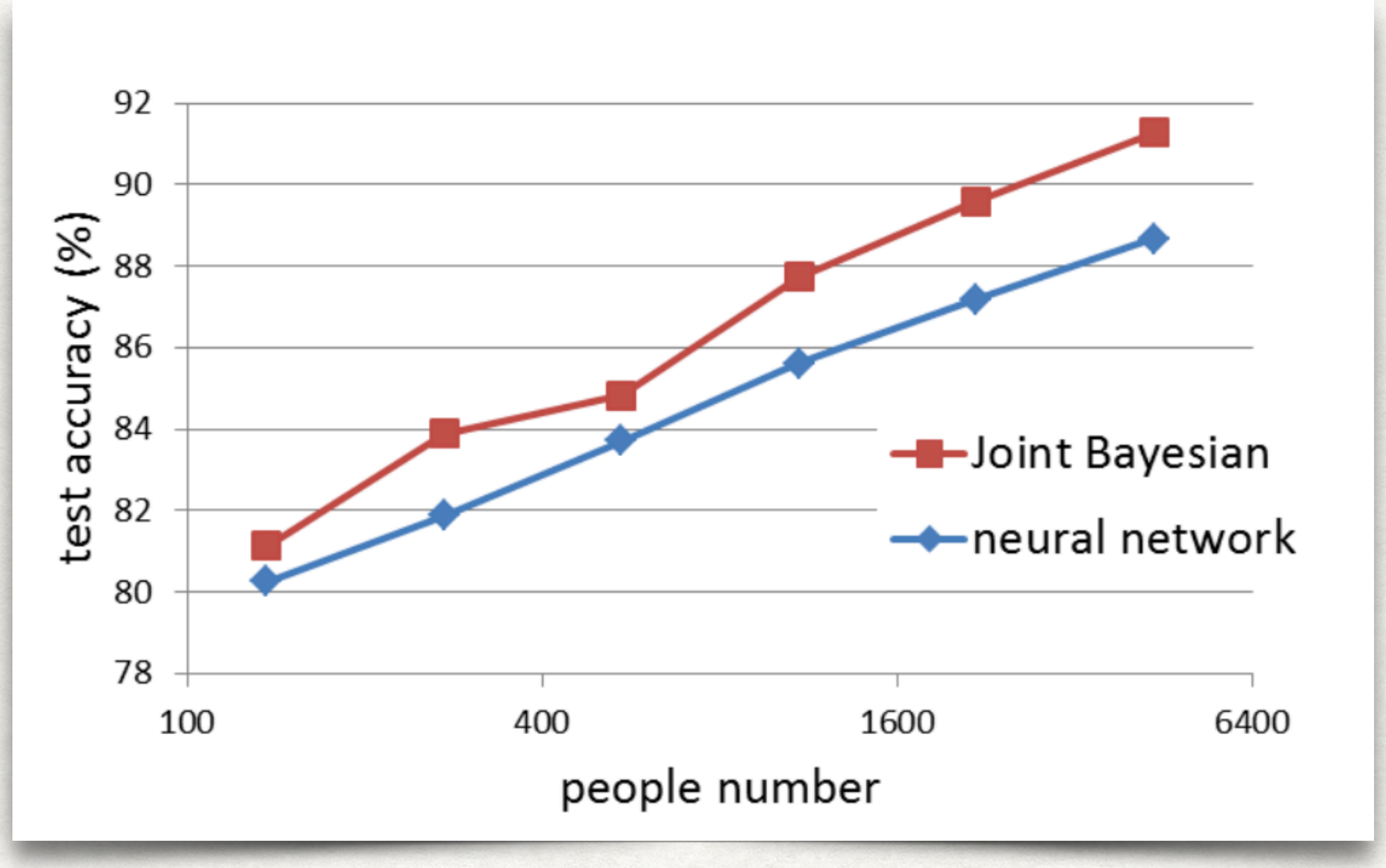
LEARNING EFFECTIVE FEATURES

- Learning from large number of classes is the key to learn compact and expressive DeepID Features

- They doubled the number of identity classes from 136 to 4349

- Significant average accuracy gain
  - 2.03% for Joint Bayesian
  - 1.68% for Neural Network

LEARNING EFFECTIVE FEATURES



JOINT BAYESIAN METHOD

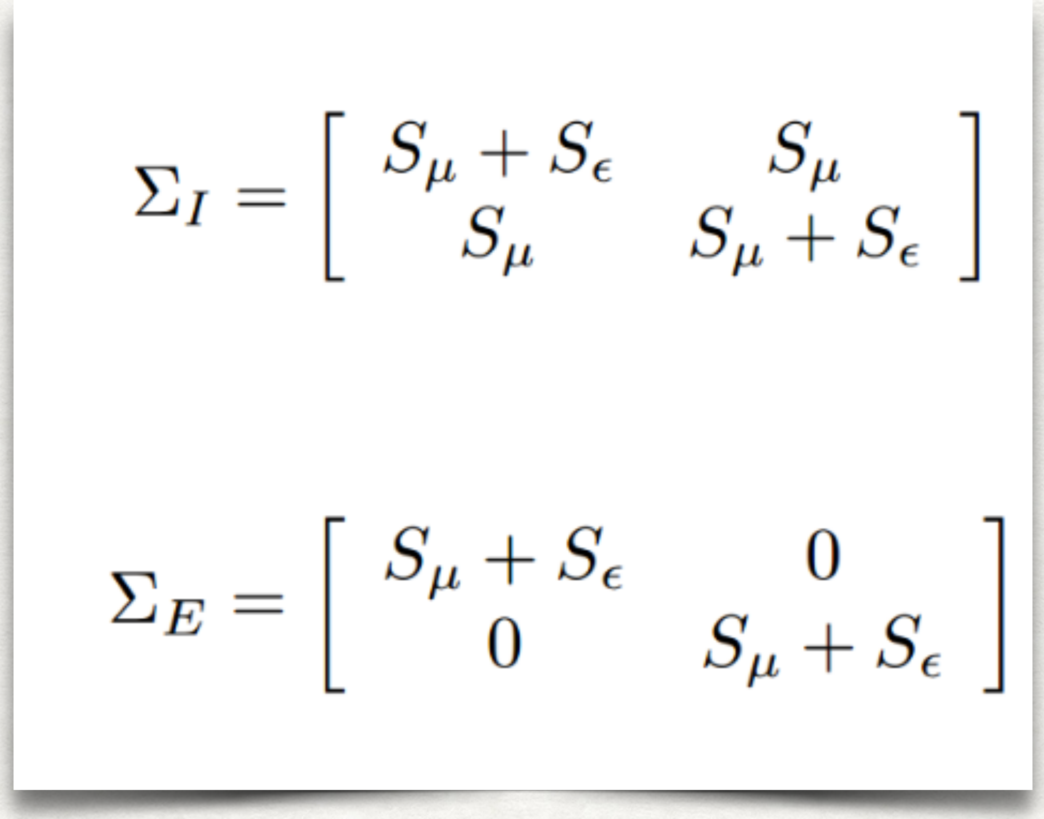
- Normal Bayesian Method uses the difference of the features from two face images
- Train on both images instead of the “difference”
- Chen et al. [2] achieved 92.4% test accuracy on LFW dataset.

JOINT BAYESIAN METHOD

$$x = \mu + \epsilon$$

- Subtract the mean from the features
- Represent them as addition of 2 Gaussian distributions
- μ ~ N (0, Sμ) face identity
- **ε** ~ N (0, S **ε**) intra personal variations

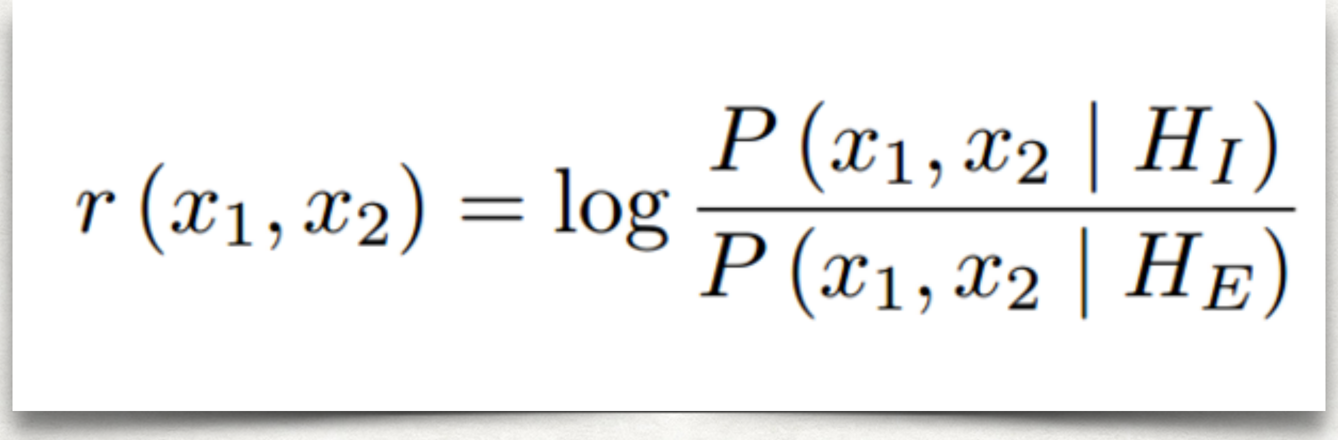
JOINT BAYESIAN METHOD



$$\Sigma_I = \left[\begin{array}{cc} S_\mu + S_\epsilon & S_\mu \\ S_\mu & S_\mu + S_\epsilon \end{array} \right]$$

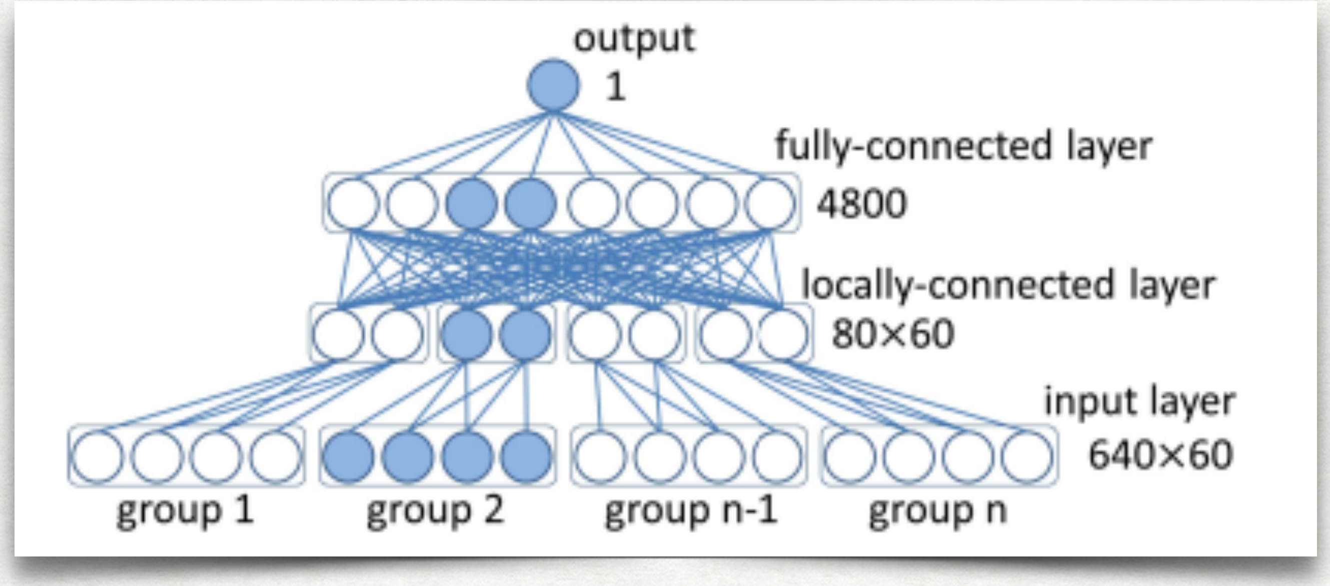
$$\Sigma_E = \left[\begin{array}{cc} S_\mu + S_\epsilon & 0 \\ 0 & S_\mu + S_\epsilon \end{array} \right]$$

JOINT BAYESIAN METHOD



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

NEURAL NETWORK METHOD



EXPERIMENTS

- They Used LFW(Labelled Faces in the Wild) dataset in order to test the methods
- 5749 people
- 85 have > 15 images
- 4069 people have one image
- Inadequate to train on this
- They used CelebFaces for training
- 87, 628 face images
- 5436 celebrities
- approximately 16 images per person

EXPERIMENTS

- 80% of Celebfaces to learn DeepID
- 20% to train neural network and joint bayesian
- Used PCA to reduce number of features to 150 in Joint bayesian

EXPERIMENTS

- Use 4349 dimensional softmax output instead of DeepID
  - 66% accuracy on Joint Bayesian
  - Neural Network fails

EXPERIMENTS

- Try Different k values for patches
  - k = 1, 5, 15, 30, 60
- Comparison of k = 1 and k = 60
  - 4.53% Neural Network
  - 5.27% Joint Bayesian
- 96.05% Joint Bayesian
- 94.32% Neural Network

EXPERIMENTS

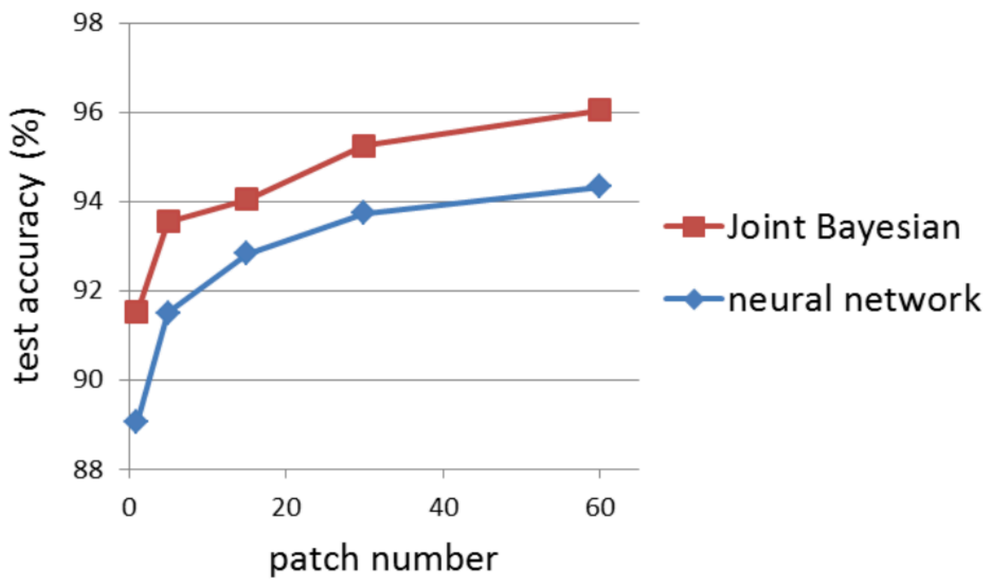


Figure 9. Test accuracy of Joint Bayesian (red line) and neural networks (blue line) using features extracted from 1, 5, 15, 30, and 60 patches. Performance consistently improves with more features. Joint Bayesian is approximately 1.8% better on average than neural networks.

METHOD COMPARISON

- Used CelebFaces+ (larger than CelebFaces)
- Used 100 Patches
- Used 5 Different scales instead of 3
- 97.20% test accuracy on LFW. (Joint Bayesian)

METHOD COMPARISON

- Training with CelebFaces+ doesn’t generalize well to LFW.
- Cao et al. [3] proposes a way to adapt joint bayesian from source domain to target domain.
- After adaptation Joint Bayesian gets 97.45% accuracy.
- 97.53% is the human accuracy.

METHOD COMPARISON

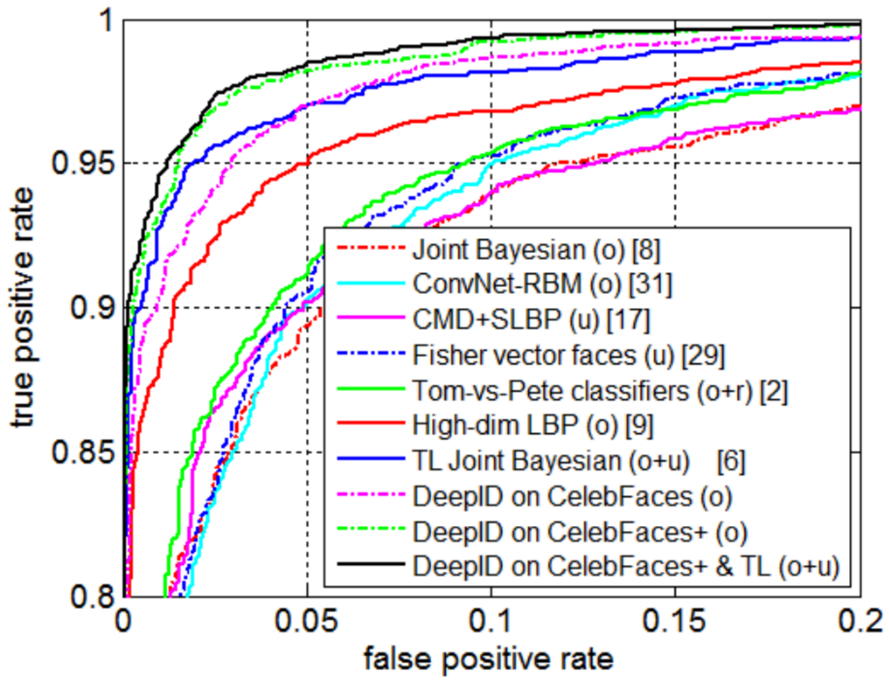


Figure 10. ROC comparison with the state-of-the-art face verification methods on LFW. TL in our method means transfer learning Joint Bayesian.

DISADVANTAGES

- 2 Step Learning process
- Shallow Model for Verification (Using DeepID)
- Need to extract patches

REFERENCES

- [1] Sun, Y., Wang, X. and Tang, X. (2014). Deep Learning Face Representation from Predicting 10,000 Classes. *2014 IEEE Conference on Computer Vision and Pattern Recognition*.

- [2] D. Chen, X. Cao, L. Wang, F. Wen and J. Sun, "Bayesian Face Revisited: A Joint Formulation", *Computer Vision - ECCV 2012*, pp. 566-579, 2012.

- [3] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun. A practical transfer learning algorithm for face verification. In Proc. ICCV, 2013. 1, 4,7, 8

- [4] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning based descriptor. In *Proc. CVPR*, 2010

- [5] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In *Proc. ECCV*, 2012

- [6] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *Proc. BMVC*, 2013.

- [7] T. Berg and P. Belhumeur. POOF: Part-based one-vs-one features for fine-grained categorization, face verification, and attribute estimation. In *Proc. CVPR*, 2013

- [8] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proc. CVPR*, 2005