Machine Learning for Face Recognition

GUAN Zhong
zhong.guan@connect.polyu.hk
BEng (Hons) Electronic & Information Eng (42470)
Supervisor: Prof. Kin-Man Lam

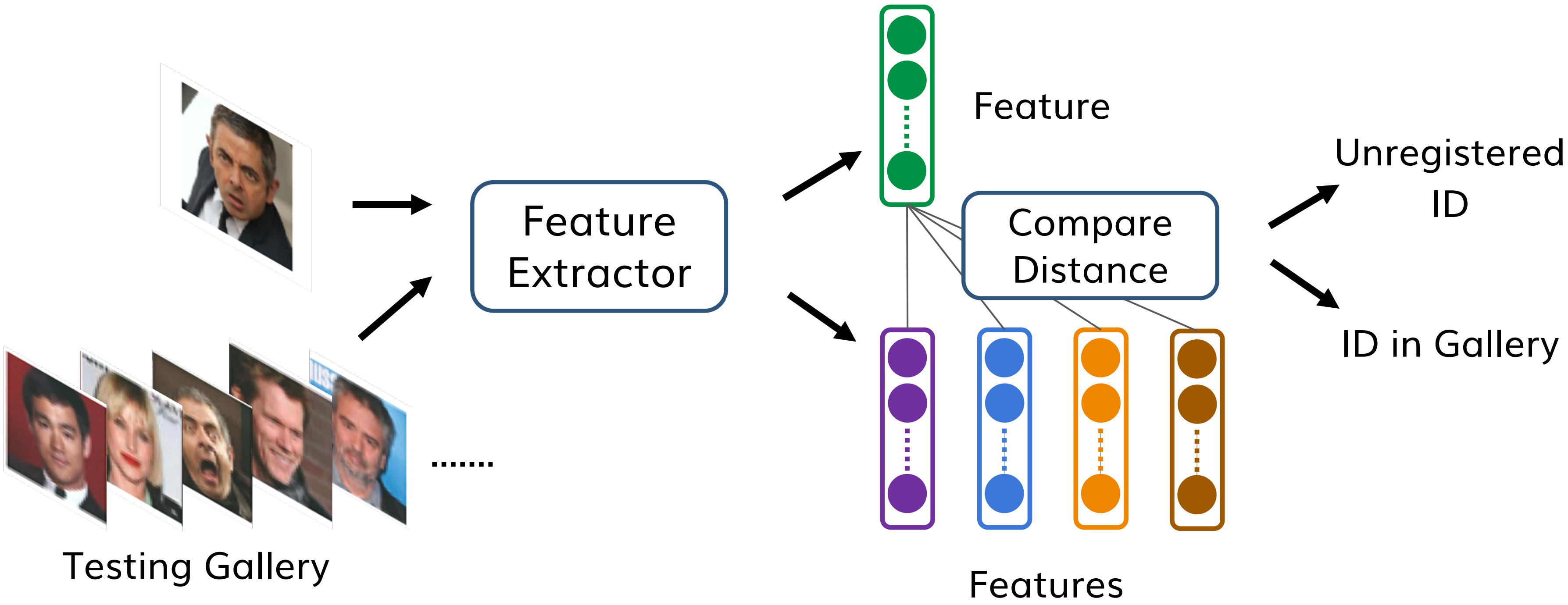
The Hong Kong Polytechnic University
香港理工大學

Department of Electronic and Information Engineering
電子及資訊工程學系

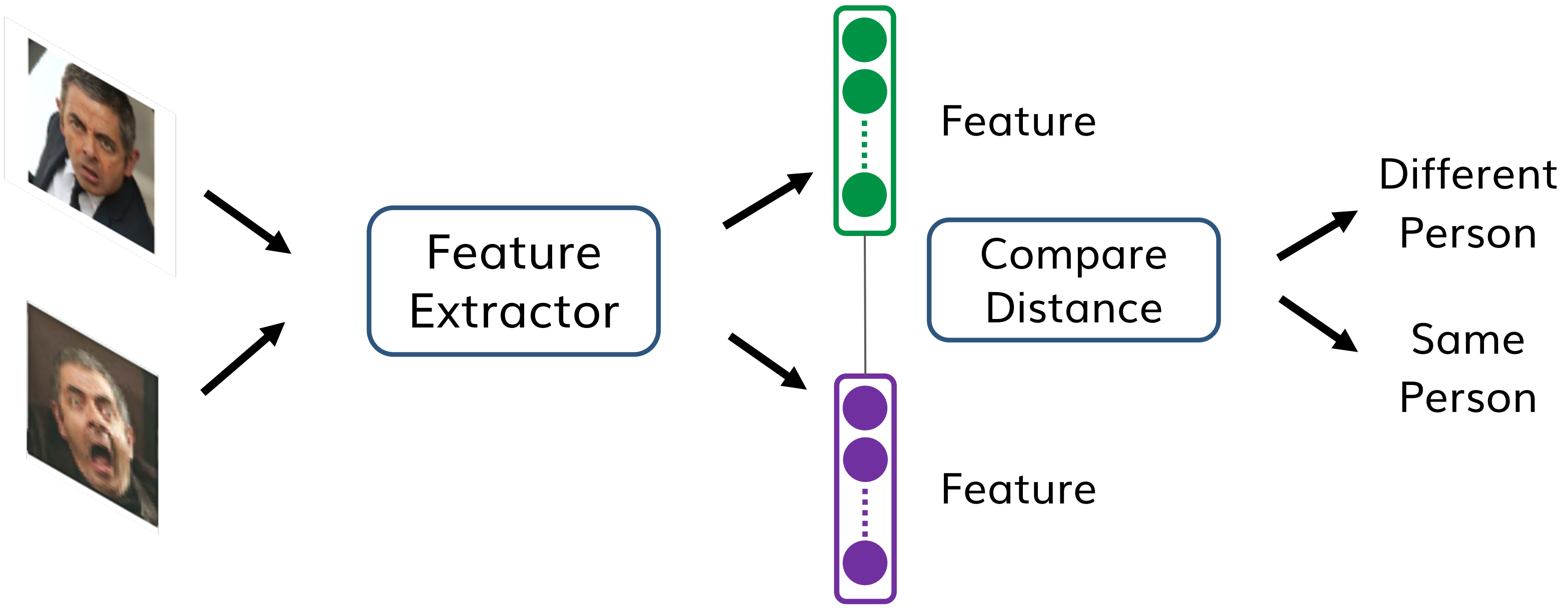
Agenda

1. Introduction
2. Methodologies
3. Experimental Setting
4. Evaluation Results
5. Conclusion

Introduction

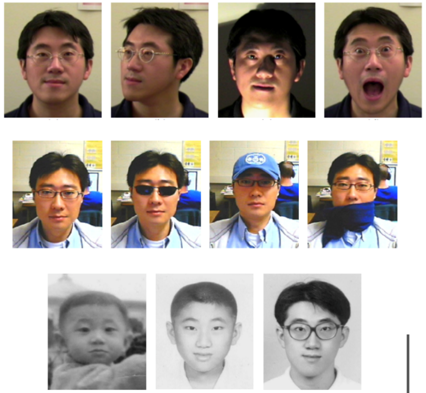


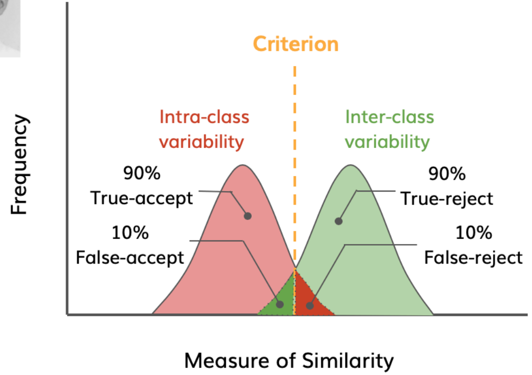
Introduction



Problems

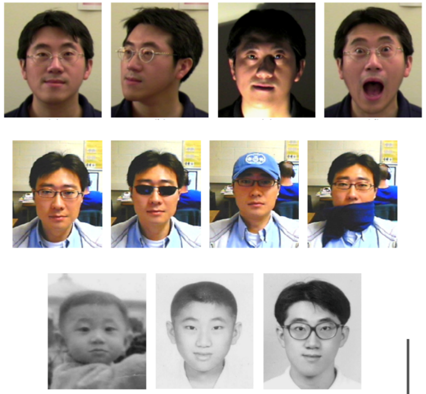
Intra-class variability > Inter-class variability

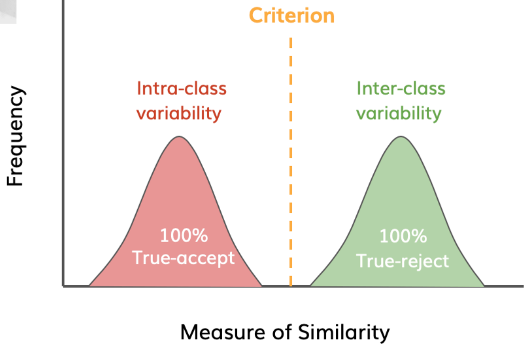




Problems

Intra-class variability < Inter-class variability





Objective

Design a model to learn discriminative feature representations such that intra-class distance < inter-class variance for <u>unconstrained conditions</u>

- Wild environment
- Poor lighting condition
- Large face angles
- Low resolution

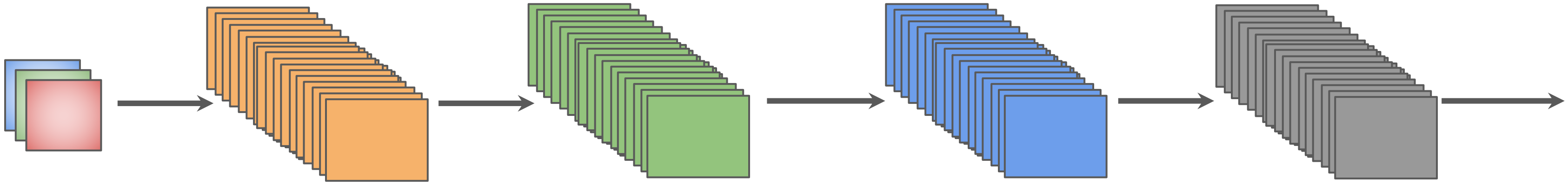
Objective

Design a model to learn discriminative feature representations such that intra-class distance < inter-class variance for <u>unconstrained conditions</u>

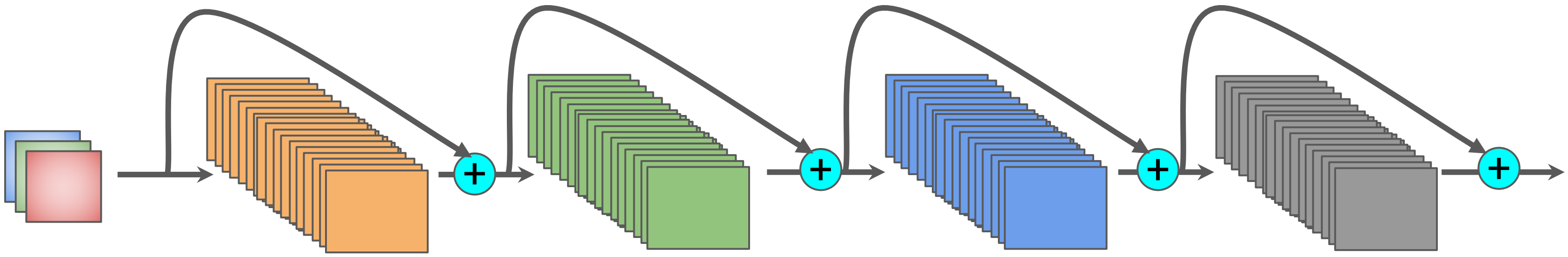
- Eigenface
- Local Binary Pattern (LBP)
- Deep face recognition

Methodologies

Standard Connectivity



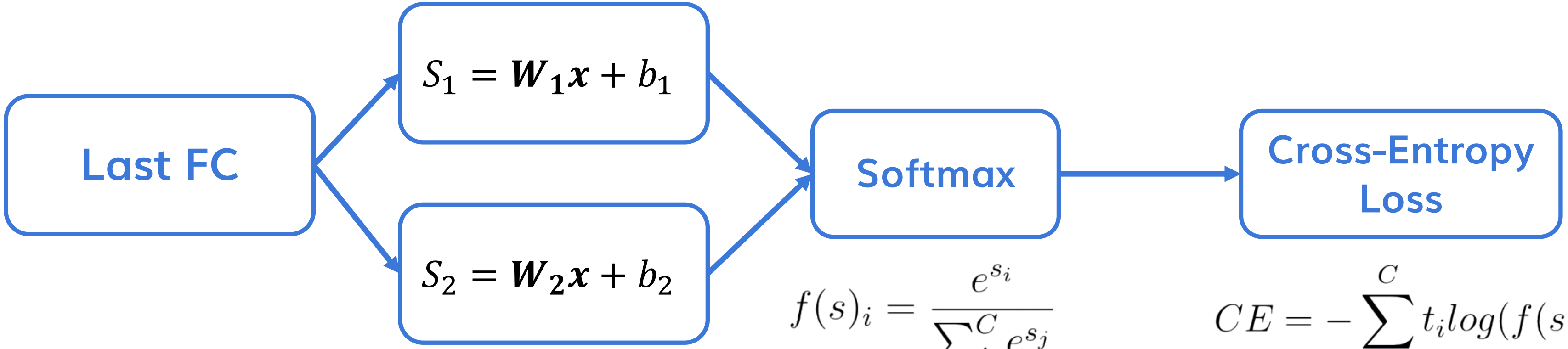
ResNet Connectivity



: Element-wise addition

K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” CoRR, vol. abs/1512.03385, 2015.

Softmax Loss



$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$$

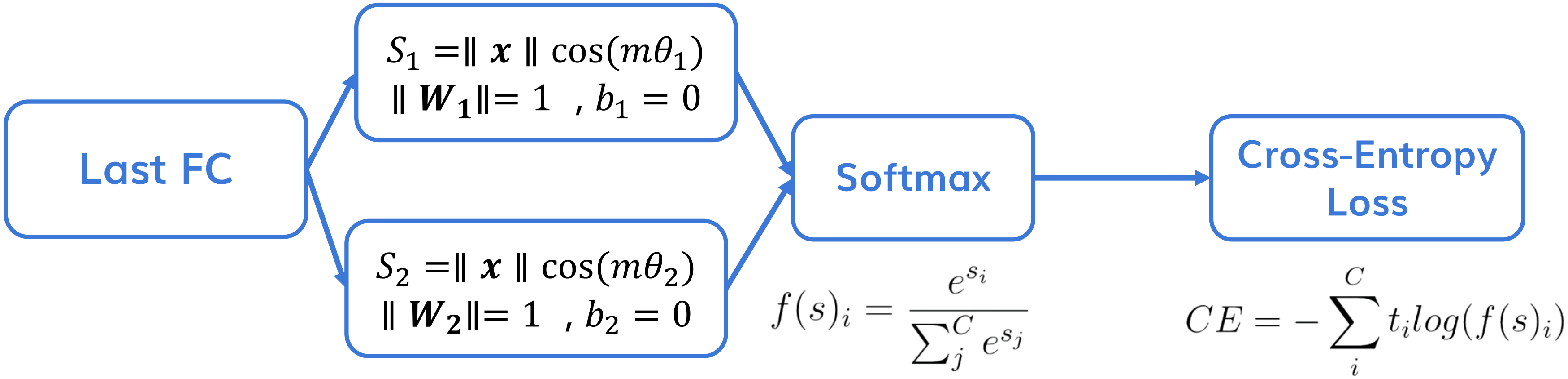
$$CE = -\sum_i^C t_i log(f(s)_i)$$

Decision Boundary:

$$(\mathbf{W_1} - \mathbf{W_2})\boldsymbol{x} + b_1 - b_2 = 0$$

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human- level performance in face verification," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708, June 2014.

A-Softmax Loss



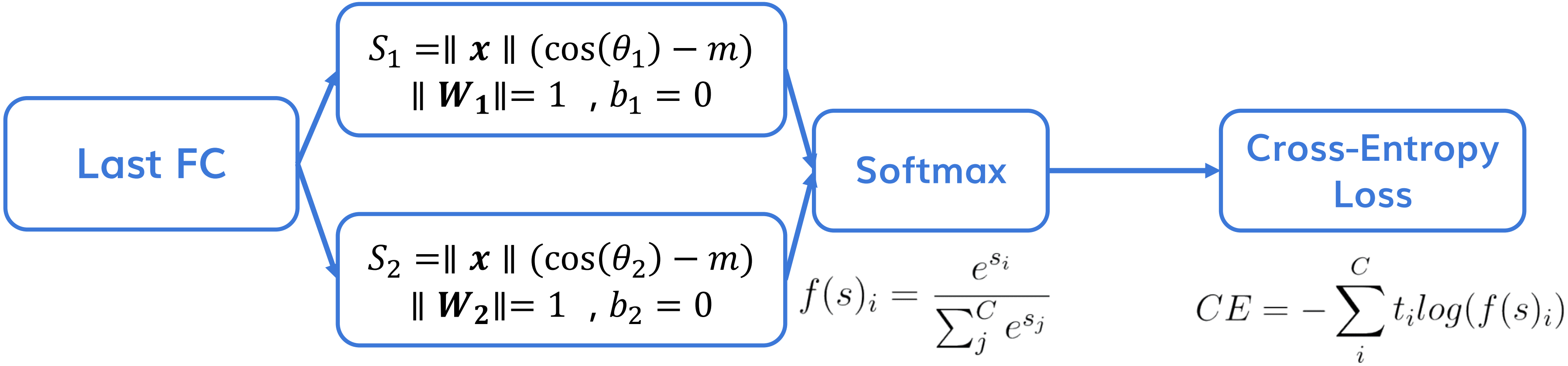
Decision Boundary:

$\| \boldsymbol{x} \| (\cos(m\theta_1) - \cos(\theta_2))$ = 0, for class 1

$\| \boldsymbol{x} \| (\cos(\theta_1) - \cos(m\theta_2))$ = 0, for class 2

W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “Sphereface: Deep hypersphere embedding for face recognition,” CoRR, vol. abs/1704.08063, 2017.

AM-Softmax Loss



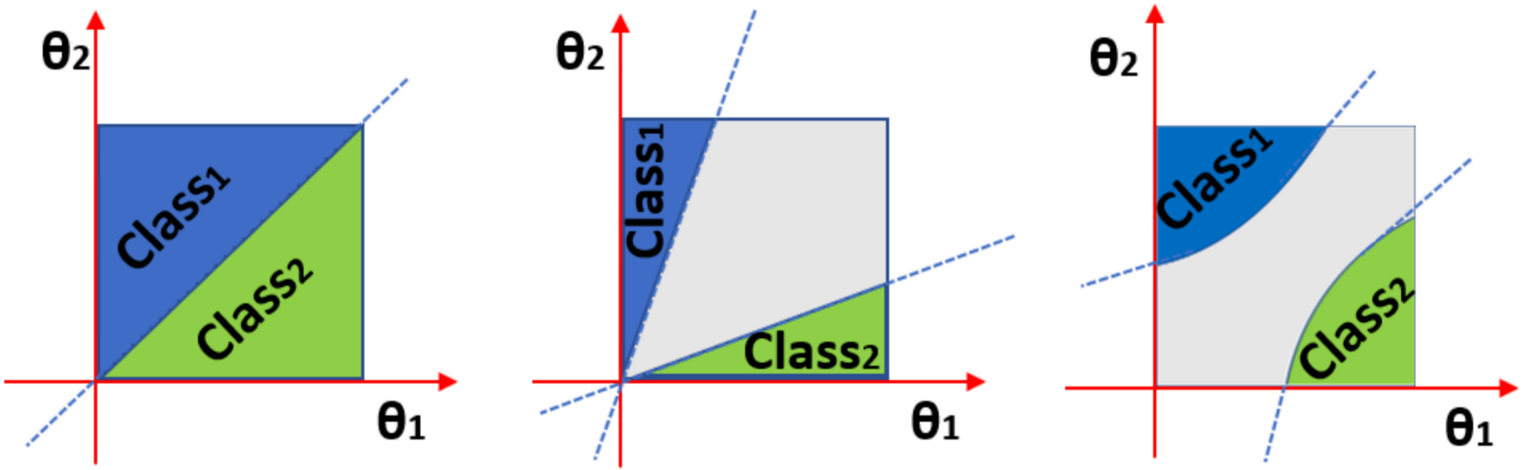
Decision Boundary:

$\| \boldsymbol{x} \| (\cos(\theta_1) - m - \cos(\theta_2)) = 0$, for class 1

$\| \boldsymbol{x} \| (\cos(\theta_1) - \cos(\theta_2) + m) = 0$, for class 2

F. Wang, W. Liu, H. Liu, and J. Cheng, “Additive margin softmax for face verification,” CoRR, vol. abs/1801.05599, 2018.

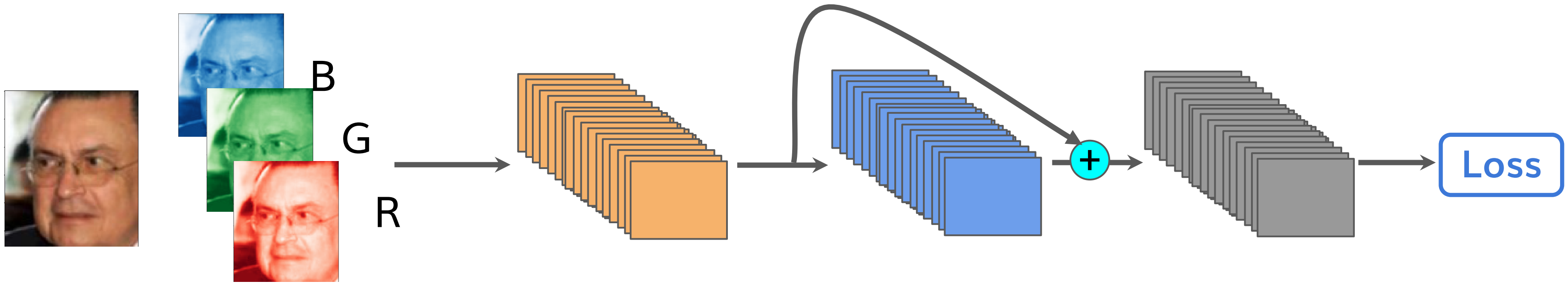
Decision Boundary Comparison



Softmax A-Softmax AM-Softmax

J. Deng, J. Guo, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," CoRR, vol. abs/1801.07698, 2018

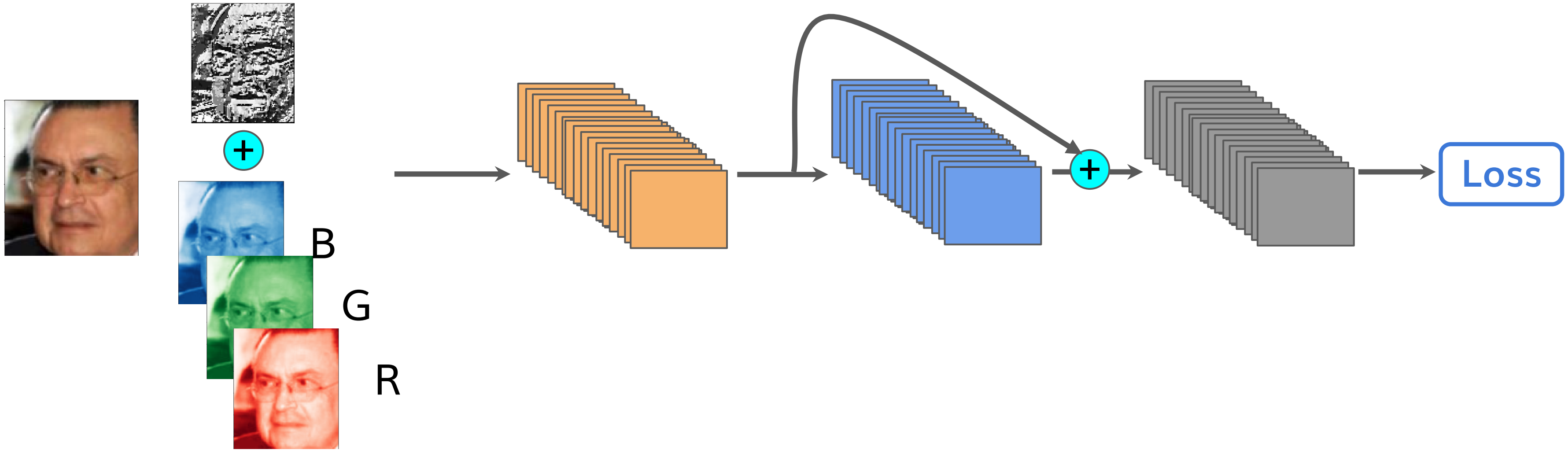
Model Design



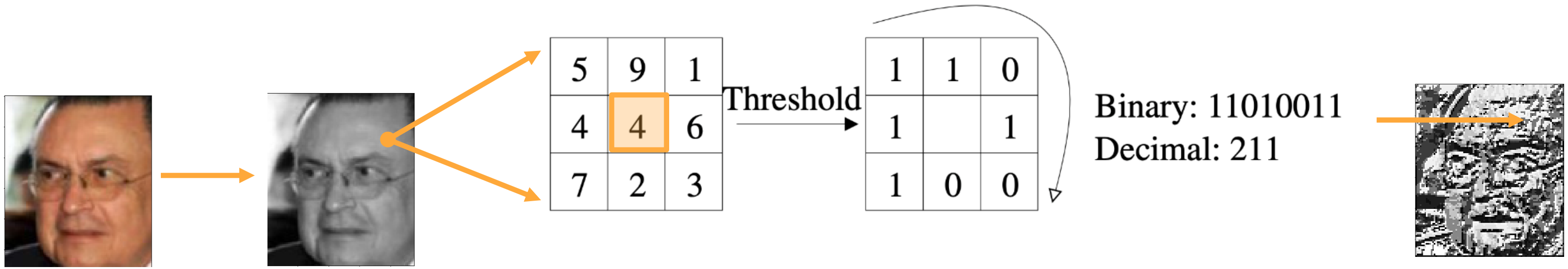
Methodologies

Model Design

LBP Feature Channel

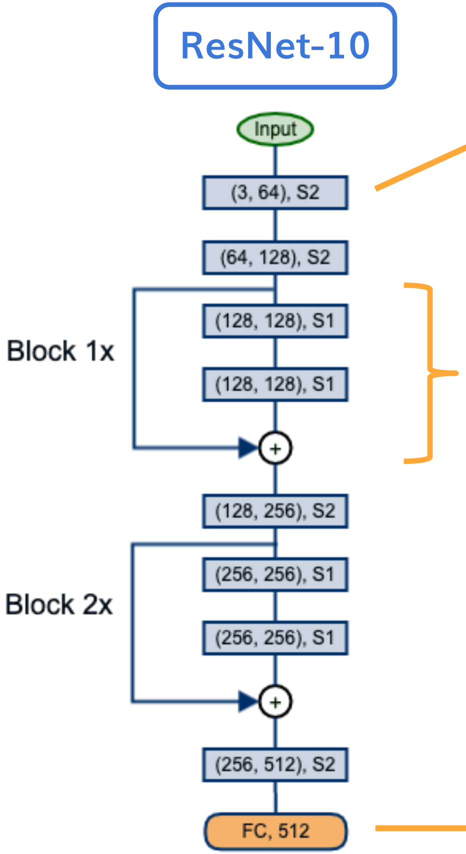


Model Design



T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, pp. 2037–2041, Dec. 2006

Network Design



Each conv layer is followed by PReLU (He et al., 2015).

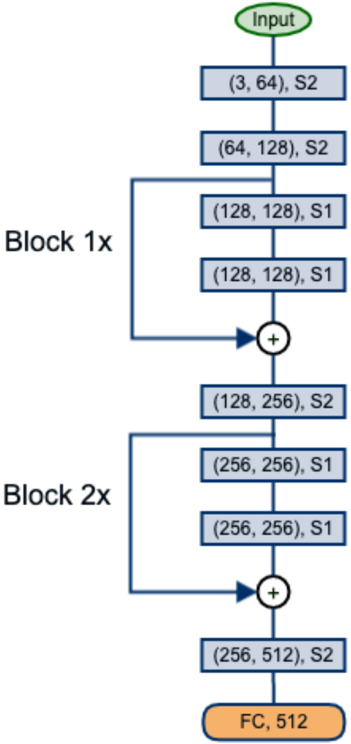
Each residual block has 2 conv Layers

Output features dimension is 512

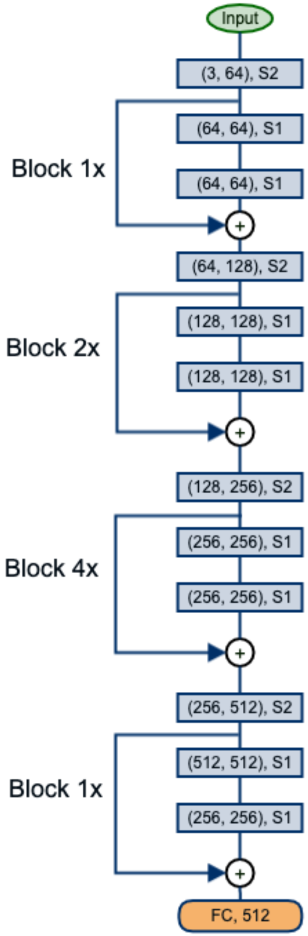
Methodologies

Network Design

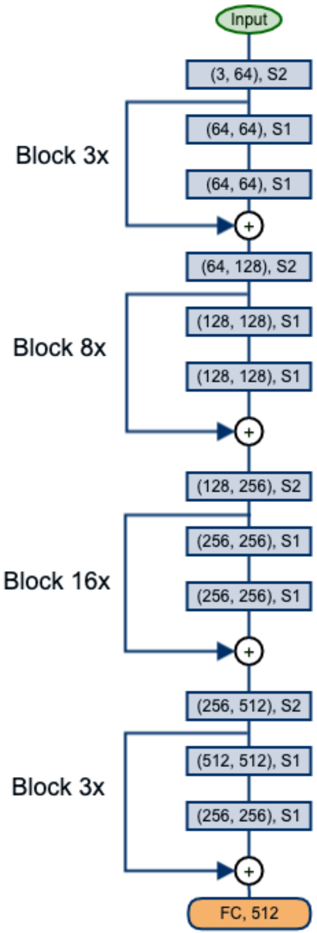
ResNet-10



ResNet-20



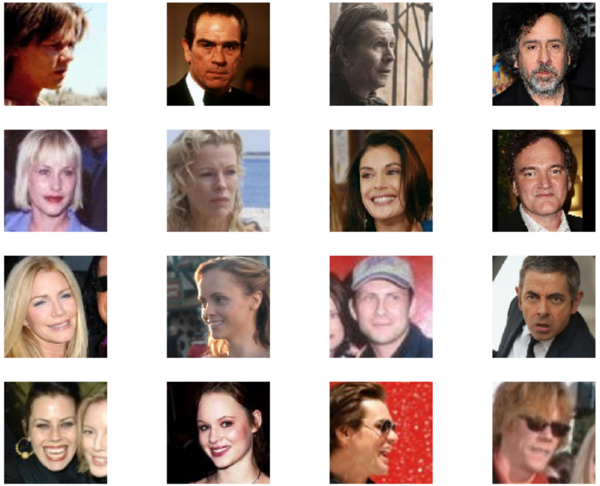
ResNet-64



Training Data

WebFace

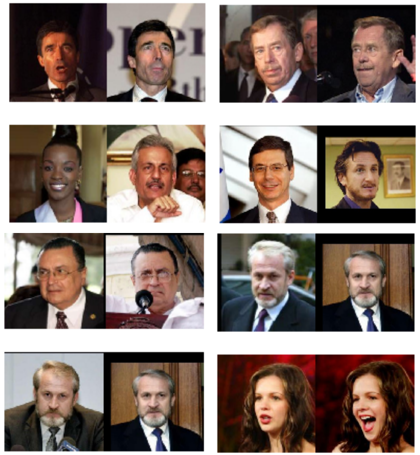
- Faces collected from web across time, places, and ages
- 494,414 face images belonging to 10,575 identities



Benchmark Data

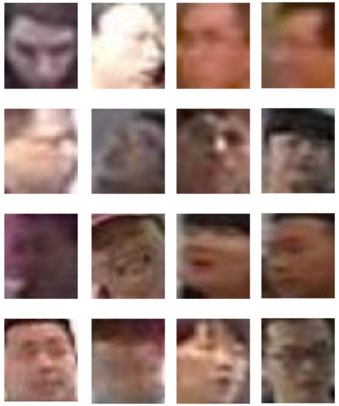
Labeled Faces in the Wild (LFW)

- Faces collected from web across time, places, and ages



Surveillance Face (SurvFace)

- Low resolution faces from surveillance camera



Multi-PIE

- 20 different illuminations and 8 pose angles within ±90°

Benchmark Data

Labeled Faces in the Wild (LFW)

- Faces collected from web across time, places, and ages

Surveillance Face (SurvFace)

- Low resolution faces from surveillance camera

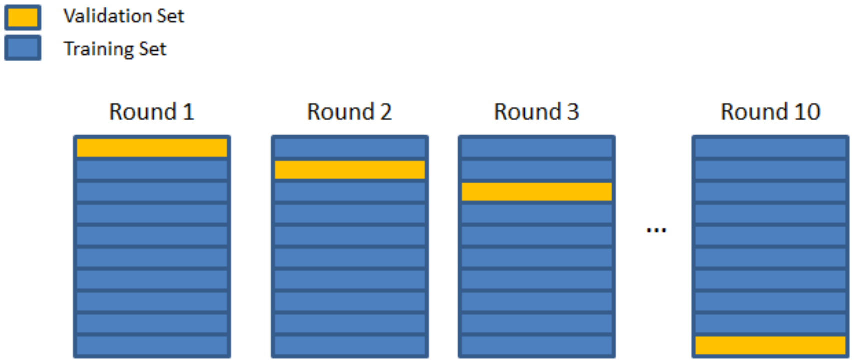
Multi-PIE

- 20 different illuminations and 8 pose angles within ±90°

Dataset	# photos	# subjects	Metrics
LFW [20]	13,323	5,749	1:1: Acc, TAR vs. FAR (ROC)
Survface [3]	463,507	15,573	1:1: Acc, TAR vs. FAR (ROC)
MultiPIE [21]	126,819	249	1:N: Rank-N (CMC)

Evaluation Result on LFW

10-Fold Cross Validation



Validation Accuracy: Acc1 Acc2 Acc3 … Acc10

Final accuracy = Average(Acc1, Acc2, …)

$$\hat{\mu} = \frac{\sum_{i=1}^{10} p_i}{10}$$

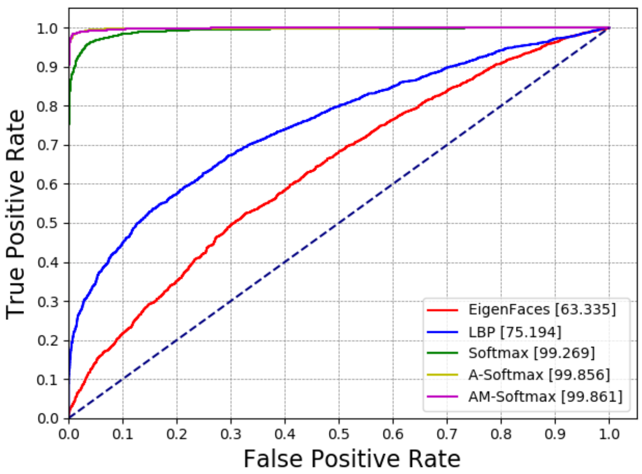
The reported standard deviation is given by:

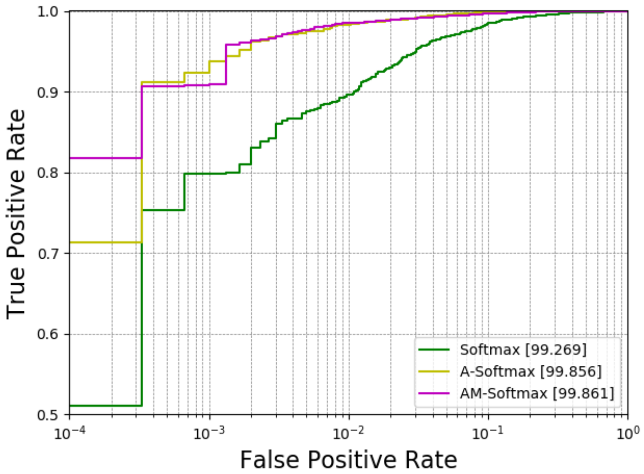
$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{10} (p_i - \hat{\mu})^2}{9}}$$

Evaluation Result on LFW

Method	Accuracy (%)
Eigenface	60.85 ± 2.36
Local Binary Pattern	68.47 ± 1.11
ResNet-10 + Softmax	96.00 ± 0.65
ResNet-10 + A-Softmax	98.68 ± 0.70
ResNet-10 + AM-Softmax	$\mathbf{98.78 \pm 0.51}$

All approaches



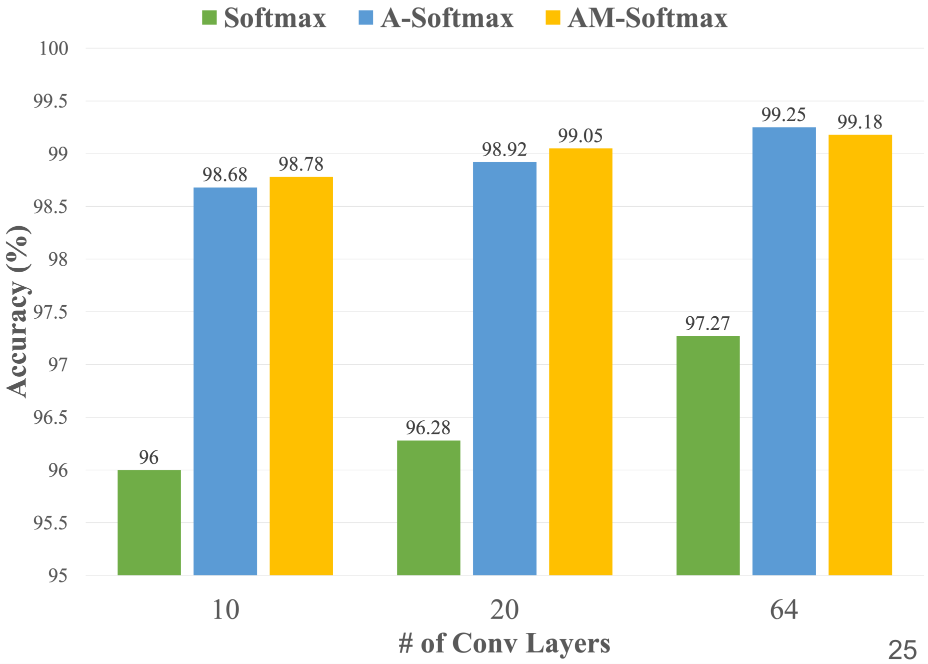


Deep Face Recognition

Evaluation Result on LFW

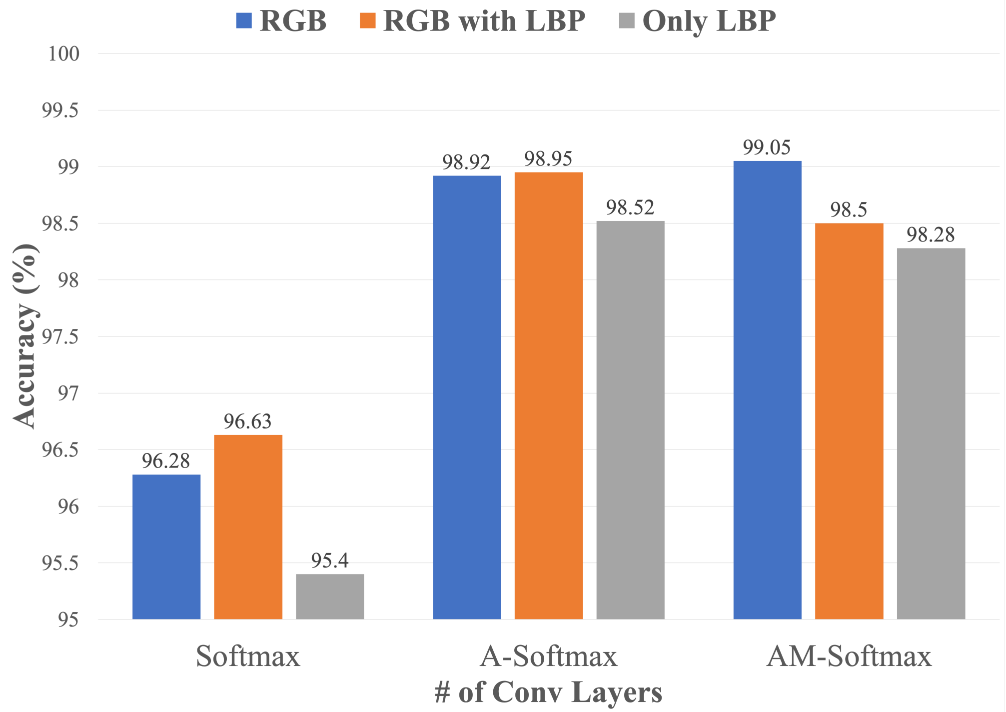
ResNet with
Different # of layers

Backbone	Loss	Accuracy (%)
ResNet-10	Softmax Loss	96.00 ± 0.65
	A-Softmax Loss	98.68 ± 0.70
	AM-Softmax Loss	**98.78 ± 0.51**
ResNet-20	Softmax Loss	96.28 ± 1.17
	A-Softmax Loss	98.92 ± 0.52
	AM-Softmax Loss	**99.05 ± 0.43**
ResNet-64	Softmax Loss	97.27 ± 0.82
	A-Softmax Loss	**99.25 ± 0.36**
	AM-Softmax Loss	99.18 ± 0.38



Evaluation Result on LFW

ResNet-20 with Different Input



Findings

- Deep CNNs can also learn from LBP features
- Adding LBP channel cannot have significant improvement on LFW
- Adding LBP channel may be helpful to the Softmax-based CNNs

Evaluation Result on Multi-PIE

Accuracy on Different Poses

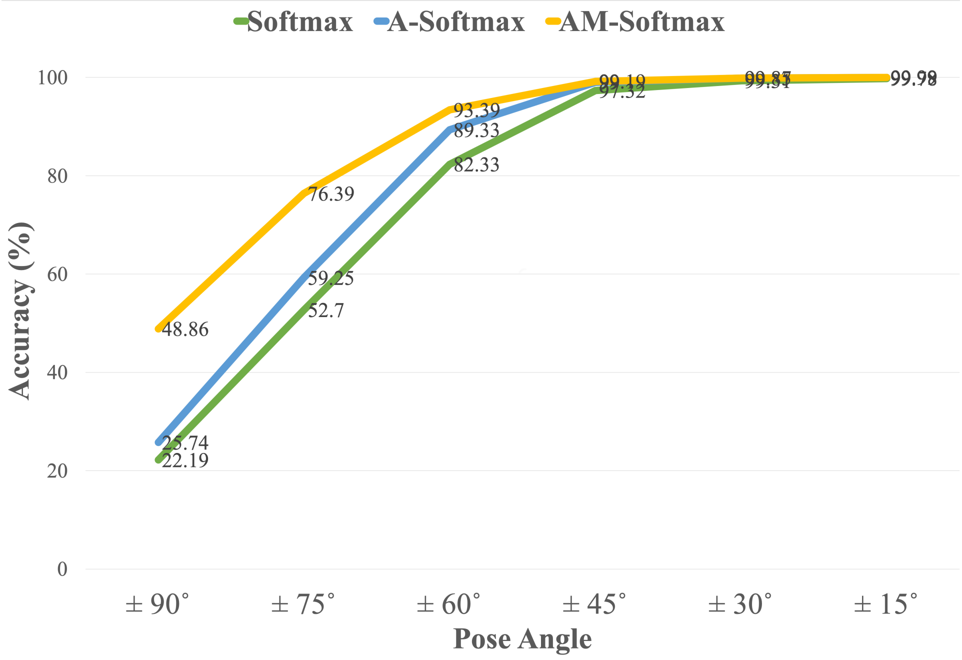
Method	±90°	±75°	±60°	±45°	±30°	±15°
ResNet-20 + Softmax	22.19	52.70	82.33	97.32	99.31	99.78
ResNet-20 + A-Softmax	25.74	59.25	89.33	99.00	99.85	99.98
ResNet-20 + AM-Softmax	**48.86**	**76.39**	**93.39**	**99.19**	**99.87**	**99.99**

Findings

- AM-Softmax-based ResNet-20 has the best performance across all pose angles

Evaluation Result on Multi-PIE

Accuracy on Different Poses



Findings

- AM-Softmax-based ResNet-20 has the best performance across all pose angles
- The performance gradually decreases as the pose angle becomes larger

Evaluation Result on Multi-PIE

Accuracy on Different Poses

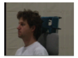
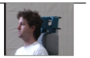
Loss Function	Input Data	±90°	±75°	±60°	±45°	±30°	±15°
Softmax	RGB image	22.19	52.70	82.33	**97.32**	99.31	99.78
	RGB plus LBP	**23.35**	**53.81**	**84.56**	97.28	**99.55**	**99.85**
	Only LBP	12.04	37.56	72.62	93.45	98.45	99.67
A-Softmax	RGB image	**25.74**	**59.25**	**89.33**	99.00	99.85	**99.98**
	RGB plus LBP	23.91	56.21	88.29	**99.21**	**99.88**	**99.98**
	Only LBP	5.61	25.26	68.80	95.58	99.22	99.63
AM-Softmax	RGB image	**48.86**	**76.39**	**93.39**	**99.19**	**99.87**	99.99
	RGB plus LBP	42.72	72.07	91.86	98.78	99.83	**100.00**
	Only LBP	30.34	65.71	89.56	98.43	99.86	99.94

Findings

- Only using LBP features can result in worse performance
- Adding LBP features help Softmax-based CNNs but not the other loss-based CNNs

Evaluation Result on Multi-PIE

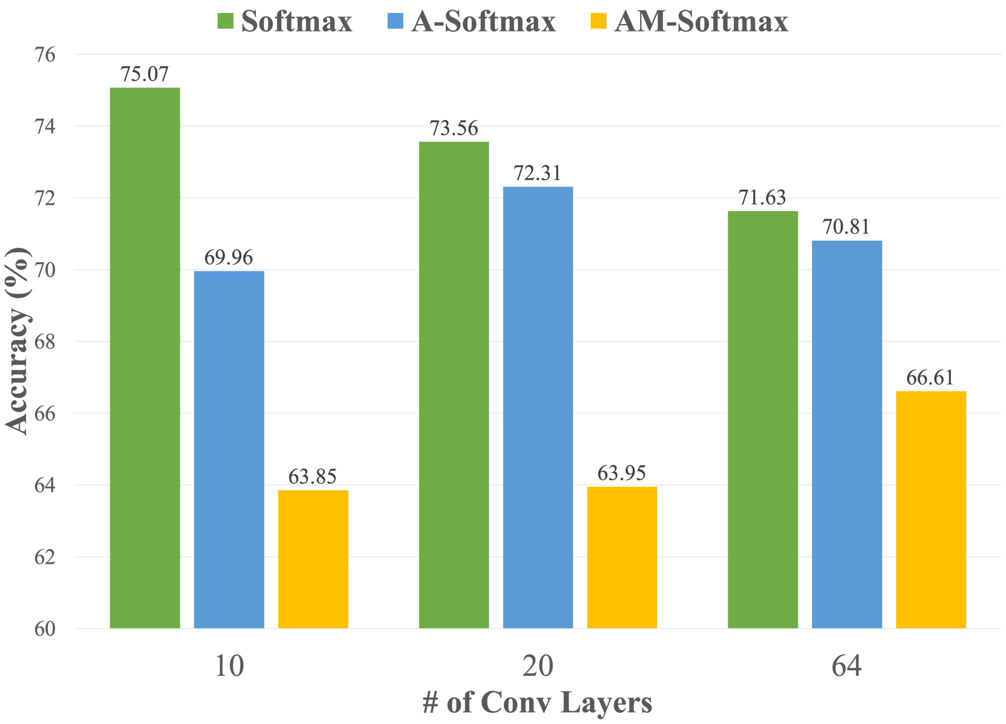
Accuracy on Different Illuminations

Loss Function	Input Data						
							
	Input Data	00	01	02	03	04	05
Softmax	RGB image	74.07	78.63	73.39	67.21	64.49	**86.89**
	RGB plus LBP	**79.84**	**81.45**	**75.00**	**67.61**	**69.8**	84.62
							
	Input Data	06	07	08	09	10	11
	RGB image	95.97	96.76	95.97	93.55	90.65	86.23
	RGB plus LBP	**96.37**	**98.38**	**96.77**	**94.35**	**94.31**	**89.47**
							
	Input Data	12	13	14	15	16	17
	RGB image	81.85	**73.98**	70.97	85.94	92.37	91.16
	RGB plus LBP	**85.08**	72.36	**75.40**	**89.96**	**96.79**	**95.18**
							
	Input Data	18	19	Avg			
	RGB image	91.16	77.79	83.45			
	RGB plus LBP	**95.58**	**82.30**	**86.03**			

Findings

- Adding the LBP features can be helpful when the lighting condition is poor

Evaluation Result on SurvFace



Findings

- Huge domain difference between WebFace and SurvFace
- Much better on WebFace, then much worse on SurvFace
- Domain Problems can significantly affect accuracy

Conclusion and Future Work

Conclusion

- Deep face recognition is far better than traditional method
- Loss functions with large decision margin and more layers can improve CNNs
- All the CNNs cannot achieve good performance when the pose angle is large, illumination condition is poor, and resolution is low
- Adding LBP features can improve Softmax-based CNN when lighting condition is poor
- Domain problems can significantly affect performance

Conclusion and Future Work

Future Work

- Domain adaptation methods can be applied into the CNNs to solve the unmatched domain problems between training data and testing data
- More experiments can be done to verify the performance improvement when adding the LBP features channel
- Use more local image descriptors or change the color space may improve the performance of deep CNNs

References

D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014.

Z. Cheng, X. Zhu, and S. Gong, "Low-resolution face recognition," *CoRR*, vol. abs/1811.08965, 2018.

Z. Cheng, X. Zhu, and S. Gong, "Surveillance face recognition challenge," *CoRR*, vol. abs/1804.09691, 2018.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, June 2014.

W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," *CoRR*, vol. abs/1704.08063, 2017.

F. Wang, W. Liu, H. Liu, and J. Cheng, "Additive margin softmax for face verification," *CoRR*, vol. abs/1801.05599, 2018.

Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016.

T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 2037–2041, Dec. 2006.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, pp. 84–90, May 2017.

G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," in *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–8, Sep. 2008.

H. Wang, Y. Wang, Z. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," in *CVPR*, 2018.

M. Wang and W. Deng, "Deep face recognition: A survey," *CoRR*, vol. abs/1804.06655, 2018.

K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015.

K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multi-task cascaded convolutional networks," *CoRR*, vol. abs/1604.02878, 2016.

D. E. Rumelhart, G. E. Hinton, R. J. Williams, *et al.*, "Learning representations by back-propagating errors," *Cognitive modeling*, vol. 5, no. 3, p. 1, 1988.

D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

Thank You!

Machine Learning for Face Recognition

GUAN Zhong supervised by **Prof, Kin-Man Lam**

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

DEPARTMENT OF Electronic and Information Engineering
電子及資訊工程學系