**Targeting Ultimate Accuracy: Face Recognition via Deep Embedding**

Jingtuo Liu Yafeng Deng Tao Bai Zhengping Wei Chang Huang

Baidu Research – Institute of Deep Learning

***Abstract*—人脸识别已经研究了几十年。 与传统的手工制作功能（如LBP和HOG）相反，可以通过深度学习方法以数据驱动的方式自动学习更复杂的功能。 在本文中，我们提出了一个两阶段方法，它结合了多片段深度CNN和深度量学习，它提取低维但非常有辨别力的特征用于面部验证和识别。 实验表明，该方法在LFW数据集上优于其他最先进的方法，在其他两个更实用的协议下，实现了99.77％的成对验证精度和更高的精度。 本文还讨论了数据大小和补丁数量的重要性，为现实世界中实际的高性能人脸识别系统提供了明确的途径。**

# INTRODUCTION

最近，基于CNN的人脸识别问题[1,2,4,6,7,8,9,12]的方法优于传统的人工特征和分类器[10,11]。 LFW（野外标记面）[5]是一种广泛使用的数据集，用于评估无约束环境中的人脸识别算法，随着更加深入的基于CNN的方法的引入，它不断攀升。这些方法的共同流程包括两个步骤。首先，深度CNN用于提取具有相对较高维度的特征向量，并且网络可以通过多类损失和验证损失来监督[6,7,8,9]。然后，PCA [2]，联合贝叶斯[6,7,8,9]或度量学习方法[12]用于学习更有效的低维表示以区分不同身份的面部。有些人将这两个阶段纳入端到端的学习过程[12]。在第一步中使用了许多智能方法，例如联合学习[6,8,9]，多级特征和监督[6,7,9]，多补丁特征[2,6,7,8,9]和复杂的网络结构[12]。同时，大量标记的面部数据通常对性能很重要。训练数据的数量范围从100K到260M。

讨论了数据大小如何影响基于CNN的深层方法的结果，以及这些技巧对于不同的数据大小是否必不可少[2,12]。我们在实验中研究了这些问题。根据我们的实验，训练数据中的面部和身份数量对最终表现至关重要。此外，即使数据量增加，基于多补丁的特征和具有三重态丢失的度量学习仍然可以对深度CNN结果带来显着改善。

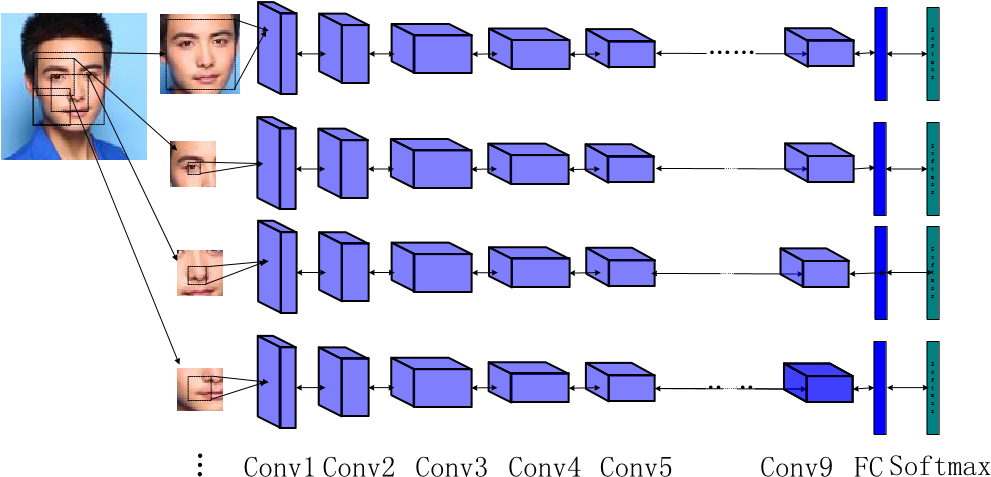
 在本文中，我们将介绍基于简单深度CNN的两阶段方法，用于多补丁特征提取和度量学习，以减少维数。我们在6000对评估协议以及其他两个协议下实现了LFW的最佳精度（99.77％）。实验将显示数据大小和多补丁如何影响性能。此外，我们将展示在现实世界中使用人脸识别技术的可能性，因为在其他两个更实用的协议下的结果也是非常有前途的。

# METHOD

我们的方法在训练中需要两个步骤。 它们将在如下的单独部分中说明。

## Deep CNNs on Multi-patch

我们简单地使用具有9个卷积层和最后的softmax层的网络结构用于监督多类学习。 网络的输入是2D对齐的RGB面部图像。 池化和标准化层位于某些卷积层之间。 在面部区域上以不同地标为中心的重叠图像块上使用相同的结构。 每个网络都在GPU上单独进行培训。 选择每个网络的最后一个卷积层的输出作为面部表示，并且我们简单地将它们连接在一起以形成高维特征。



…

**Figure 1.** Overview of deep CNN structure on multi-patch.

## Metric Learning

高维特征本身具有代表性，但它对于人脸识别来说效率不高而且非常冗余。 由三元组损失监督的度量学习方法用于将特征减少到低维度，例如128/256浮点，同时使其在验证和检索问题中更具辨别力。 具有三重态丢失的度量学习旨在缩短属于相同身份的样本的L2距离并且在来自不同身份的样本之间扩大它。 因此，与多类损失函数相比，三元组损失更适合于验证和检索问题。

…

…

Multi

-

patch

S

o

f

t

m

a

x

…

…

…

…

…

…

…

…

…

…

conv9 Concatenate 128 float Triplet loss

**Figure 2.** Metric learning with triplet loss

# EXPERIMENTS

## Training Datasets

我们在网站上收集名人的图像，检测图像中的面部，并用手标记每个人的面部以消除噪音。 在通过名字移除LFW中的人之后，我们得到了大约18000人，具有大约120万个面部图像。 每个面都由地标定位和对齐。 我们使用数据集来训练我们的模型。

## Evaluation Protocols

LFW是实际情况下最受欢迎的人脸识别评估基准。 评估LFW的性能有三种评估方案。 第一个协议是测试6000个面对的准确性，这是Huang等人提出的。人。 在[5]中更新并在[13]中更新，我们遵循“无限制，标记的外部数据”任务来评估我们的方法。 第二种协议在[3]中提出，该协议包括闭集识别任务和开集识别任务。 在[14]中提出了第三个协议，其中包括验证任务和开放集识别任务。 因此，有5个任务用于评估和比较我们的模型与其他方法。 有关所有协议的详细信息，请参阅[5,3,13,14]。

## Data Driven

我们训练了三种具有150K，450K和1.2M面部图像的嵌入模型。 它们中的每一个都输出128-d矢量作为面部的表示，并且在这样的128-d空间中的欧几里德距离适当地测量任何两个面的相似性。 基于训练数据的分布，估计某个阈值以判断两个面是否是同一个人。 我们使用第一个协议对LFW上的三个模型进行了测试，并观察到了更多训练数据的显着改进，如表1所示：使用来自18K人的1.2M面部图像，6000对验证错误率降低了三分之二以上 来自1.5K人的150K面部图像。

**TABLE 1.** PAIR-WISE ERROR RATE WITH DIFFERENT AMOUNT OF TRAINING DATA

|  |  |  |
| --- | --- | --- |
| ***Identities*** | ***Faces*** | ***Error rate*** |
| 1.5K | 150K | 3.1% |
| 9K | 450K | 1.35% |
| 18K | 1.2M | 0.87% |

## Effect of Multi-patch

为了研究使用多个贴片的有效性，9个CNN模型分别用1.2M面部图像进行训练，每个面部图像采用以一定比例的不同界标为中心的裁剪贴片。 基于一个，四个，七个或九个CNN模型的输出的串联来学习四种不同的嵌入模型。 我们发现这种方法非常有效，因为局部补丁通常对诸如姿势或表情之类的变化更具鲁棒性。 如表2所示，6000对错误率随着补丁数量的增加而减少，但在七个补丁后以某种方式饱和。

**TABLE 2.** PAIR-WISE ERROR RATE WITH DIFFERENT NUMBER OF

PATCHES

|  |  |
| --- | --- |
| ***Number of patch*** | ***Error rate*** |
| 1 | 0.87% |
| 4 | 0.55% |
| 7 | 0.32% |
| 9 | 0.35% |

## Final Performance on LFW

如3.4节所述，七补丁嵌入模型实现了99.68％的成对分类精度，这已经是该协议下发布的最佳结果。此外，我们还训练了另外9个具有不同参数的嵌入模型，并将这10个模型的欧几里德距离测量结合起来进行进一步改进。我们简单地遵循十倍交叉验证规则：线性集合模型在每次迭代中训练九次，并用剩余的一次进行测试。通过使用启发式网格搜索算法直接最小化训练数据的分类误差来获得模型的权重和分类阈值。十次迭代的最终平均测试精度为99.77％。至于第二和第三协议中的其他四个任务，我们简单地平均所有十个嵌入模型的距离测量。如表3所示，我们的单个模型和集合模型都优于以前在五个任务上发布的所有结果。

成对精度是LFW上最流行的协议。通过组合10个模型，所提出的方法达到99.77％1，这将[12]中报告的先前技术水平的误差减少了约38％。在所有6000对中，只有14对被这个整体模型错误分类，其中五个实际上根据最新的LFW勘误标记错误标记。图3中列出了14个错误分类的对，以及它们由集合模型给出的分数。

虽然几种算法在6000对验证任务中已经达到了近乎完美的准确度，但面部验证应用的一个更实用的标准是在极低的错误接受率（例如，@ 0.1％）下的错误拒绝率，这是在远端。 ROC曲线而不是其中心部分。 此外，低错误接受率下的开放式识别甚至更具挑战性，但适用于许多情况。 我们还将我们的单一模型和集合模型与先前发布的方法在表3中的这些扩展协议上进行了比较。特别是在第二协议的开集识别任务[3]中，最佳公布的识别率为81.4％[9]，而 我们的集成模型可以达到95.8％的识别率，相对降低错误率约77％。

**TABLE 3.** COMPARISONS WITH OTHER METHODS ON SEVERAL EVALUATION TASKS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Method*** | ***Performance on tasks*** | | | | |
| ***Pair-wise Accuracy(%)*** | ***Rank-1(%)*** | ***DIR(%) @ FAR =1%*** | ***Verification(%***  ***)@ FAR=0.1%*** | ***Open-set***  ***Identification(%***  ***)@ Rank =***  ***1,FAR = 0.1%*** |
| ***IDL Ensemble Model*** | ***99.77*** | ***98.03*** | ***95.8*** | ***99.41*** | ***92.09*** |
| ***IDL Single Model*** | ***99.68*** | ***97.60*** | ***94.12*** | ***99.11*** | ***89.08*** |
| ***FaceNet[12]*** | ***99.63*** | ***NA*** | ***NA*** | ***NA*** | ***NA*** |
| ***DeepID3[9]*** | ***99.53*** | ***96.00*** | ***81.40*** | ***NA*** | ***NA*** |
| ***Face++[2]*** | ***99.50*** | ***NA*** | ***NA*** | ***NA*** | ***NA*** |
| ***Facebook[15]*** | ***98.37*** | ***82.5*** | ***61.9*** | ***NA*** | ***NA*** |
| ***Learning from Scratch[4]*** | ***97.73*** | ***NA*** | ***NA*** | ***80.26*** | ***28.90*** |
| ***HighDimLBP[10]*** | ***95.17*** | ***NA*** | ***NA*** | ***41.66(reported in [4])*** | ***18.07(reported in [4])*** |



S

core =

-

0.06

0

(

pair

#113

)

S

core =

-

0.0

22

(

pair

#

202

)

S

core =

-

0.0

34

(

pair

#

656

)



Score =

-

0.031

(

pair

#

1230

)

Score =

-

0.073

(

pai

r

#

1862

)

Score =

-

0.09

1(

pair

#

2499

)



Score = -0.024 (pair #2551) Score = -0.036 (pair #2552) Score = -0.089 (pair #2610)



Score = -0.071 (pair #3652) Score = -0.021 (pair #3766) Score = -0.001 (pair #5636)

1. **False Reject**



Score = 0.032 (pair #3503) Score = 0.069 (pair #4585)

1. **False Accept**

**Figure 3.** Failed cases in the LFW pair-wise verification task(including cases with wrong label): (a) False Reject Pairs. (b) False Accept Pairs. The score under the image is the similarity score of the above pair, and the score is in the range of [-1.0, 1.0] with threshold 0.0. The 5 pairs with red rectangle are wrong labeled ones.

# DISCUSSION

众所周知，面部验证和开放式识别是人脸识别的最常见应用。对于验证任务，当误报率为0.001时，召回我们的方法达到99.41％，即使误报率为0.0001，召回率为97.38％。它表明，面部验证性能足以满足实际应用的需求。但对于开集识别，当误报率为0.0001时，召回率约为80％。虽然这是这项任务的最佳结果，但非常有希望，考虑到它在识别场景中更容易出现误报，我们认为开放式识别的性能仍不足以满足实际应用的需要。

我们发现训练数据对于人脸识别的表现非常重要。我们通过手机摄像头收集了评估数据集，其中包括大约3300名中国人，并且在不同时间收集了一个人的所有面部。我们使用由名人训练的模型来测试评估数据集上的验证任务，当误报率为0.0001时，我们达到了85％。在我们添加网站收集的中国名人培训相同参数的模型后，当误报率为0.0001时，我们达到了92.5％。我们相信，如果我们在与评估数据集相同的情况下添加更多面部，我们可以获得更好的结果。我们认为数据和算法一样重要，我们建议在实际情况下收集大量数据之前，最好不要得出人脸识别方法不够好的结论。

LFW一直是最受欢迎的人脸识别评价基准，在促进人脸识别社会改进算法方面发挥了非常重要的作用。但是，在剩下只有9个错误的对（除了错误的标记对）之后，这可能达到LFW数据集的最终性能，预计新的基准将更有效地比较不同的方法。

# CONCLUSION

我们提出了一种结合深度CNN和度量学习的两阶段人脸识别方法。 受益于多补丁的功能，我们的方法可以很好地处理具有变体姿势，遮挡和表达的情况。 随着训练数据中每个身份的身份和面部数量的增加，性能也相应提高。

所提出的方法在主协议下优于LFW的最新方法，并且当FAR相当低时获得相当高的验证率。 随着算法的不断改进，我们希望人脸识别技术最终能够在现实世界中更具挑战性的条件下得到广泛应用。

## ACKNOWLEDGEMENT

我们要感谢Erik G. Learned-Miller教授在本文中提供的响应性反馈和有用的建议。

## REFERENCES

1. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In Proc. CVPR, 2014.
2. Erjin Zhou, Zhimin Cao, Qi yin. Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not? Technical report, arXiv:1501.04690.
3. L. Best-Rowden, H. Han, C. Otto, B. Klare, and A. K. Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. TR MSU-CSE-14-1, 2014.
4. D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. In arXiv:1411.7923, 2014.
5. 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. Technical Report07-49, University of Massachusetts, Amherst, October 2007.
6. Y. Sun, Y. Chen, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. In Advances in Neural Information Processing Systems, pages 1988–1996, 2014.
7. Y. Sun, X.Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1891–1898. IEEE, 2014.
8. Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse,selective, and robust. arXiv preprint arXiv:1412.1265, 2014.
9. Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. DeepID3: Face Recognition with Very Deep Neural Networks. *arXiv:1502.00873*, 2014.

1. D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: Highdimensional feature and its efficient compression for face verification. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 3025–3032. IEEE, 2013.
2. Xudong Cao, David Wipf, Fang Wen, and Genquan Duan.A Practical Transfer Learning Algorithm for Face Verification.International Conference on Computer Vision (ICCV), 2013.
3. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. CVPR, 2015.
4. Gary B. Huang and Erik Learned-Miller. Labeled Faces in the Wild: Updates and New Reporting Procedures. Technique report,

University of Massachusetts , 2015.

1. S. Liao, Z. Lei, D. Yi, and S. Z. Li. “A benchmark study of largescale unconstrained face recognition”. In IAPR/IEEE International Joint Conference on Biometrics, Clearwater, Florida, USA, 2014.
2. Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. “Web-Scale Training for Face Identification”. CVPR, 2015.