

A Faster Lightweight Face Recognition Model

Shuai Peng¹, Hongbo Huang^{1,2}, Weijun Chen¹, Shuwen Pan^{1,2}

¹Computer School, Beijing Information Science and Technology University, Beijing

²Institute of Computing Intelligence, Beijing Information Science and Technology University, Beijing
Email: bistups@163.com, hhb@bistu.edu.cn

Received: Mar. 23rd, 2020; accepted: Apr. 7th, 2020; published: Apr. 14th, 2020

Abstract

With the development and application of deep learning-based approaches, face recognition algorithms have already been used on devices with sufficient computing resources and achieved high accuracy and fast speed. Face recognition models based on deep learning have better recognition accuracy, but requiring a large amount of computing resources. Aiming to this problem, this paper designs a model called Lite-Inception-ResNet, which is a lightweight network model based on deep learning algorithms and requires much fewer computing resources. The proposed model is based on the Inception-ResNet model and improved in network architecture and activation functions. Experiments on VGGFace2 and LFW show that the Lite-Inception-ResNet model can reduce the amount of parameters by 88.2% and the amount of calculation by 76.5% with only a 0.1% accuracy reduction, making the model more suitable for devices with less computing resources.

Keywords

Deep Learning, Face Recognition, Lightweight

一种更快捷的轻量级人脸识别模型

彭 帅¹, 黄宏博^{1,2}, 陈伟骏¹, 潘淑文^{1,2}

¹北京信息科技大学计算机学院, 北京

²北京信息科技大学计算智能研究所, 北京
Email: bistups@163.com, hhb@bistu.edu.cn

收稿日期: 2020年3月23日; 录用日期: 2020年4月7日; 发布日期: 2020年4月14日

摘 要

随着基于深度学习的人脸识别算法的发展和应用, 人脸识别算法已经可以运用在计算资源充足的设备上

并取得很高的精度和较快的速度,但是在计算资源受限设备上的应用有诸多困难。基于深度学习的人脸识别模型有着更好的识别精度,但是大多数基于深度学习的模型均需要大量的计算资源来支持运行。本文针对这一问题,设计出一个占用少量计算资源的基于深度学习算法的轻量级网络模型 Lite-Inception-ResNet,该模型基于具有良好性能的Inception-ResNet模型,在保持原模型良好性能的基础上对卷积核和网络架构进行了优化和重新设计,并选用了性能更好的激活函数。在VGGFace2和LFW上的实验表明,新模型可以在LFW数据集上仅降低0.1%正确率的情况下减少88.2%的参数量和76.5%的计算量,使该模型可以较好地应用于计算资源较少的设备上。

关键词

深度学习, 人脸识别, 轻量级

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着深度学习的发展,基于深度学习的人脸识别方法在现实中得到了广泛的应用。尤其在公共安全、智慧城市和金融交易等方面,对人脸识别技术的需求与日俱增。当前基于深度学习的人脸识别算法在精度上已经可以满足各类需求,但是其需要用到的计算量是非常大的,对运用的计算设备有着较高的需求,尤其高精确度的人脸识别模型通常需要用到大型的并行运算设备,这对于将模型迁移到计算资源少的小型设备上是一种挑战。因此,如何减少模型消耗的计算成本,设计出在计算资源少的小型设备上能高效运行的轻量级网络模型,已经成为进一步研究的热点之一。

轻量级网络模型吸引了大量研究人员的关注,目前已经取得不少的研究成果,如通过裁切、压缩和编码来降低网络模型的参数量和计算量的方法[1]。另有一些工作如 MobileNet [2]和 ShuffleNet [3]等则通过重新设计网络结构,引入逐通道卷积和逐点卷积等方式来降低参数量。这些模型在降低参数量和计算量的同时,其识别率等性能也下降明显,使得如何在轻量级网络模型的性能和计算复杂度之间取得恰当的均衡成为一个新问题。本文针对这一问题,提出一个以大型网络架构 Inception-ResNet [4]为基础,在识别率几乎没有下降的情况下优化出的一个轻量级网络模型 Lite-Inception-ResNet,在性能和计算量之间取得了更好的平衡。

2. 相关研究与方法基础

2.1. 人脸识别相关进展

近几年,基于卷积神经网络的方法已经成为人脸识别的主流方法,其主要分为两条研究路线,一条是对网络架构进行改进的方法[5] [6],通过改进网络结构来提升模型的性能和速度。另一条是对损失函数进行改进的方法[7] [8] [9],通过改进损失函数来增强网络在大规模人脸识别上的区分度。随着模型在精度上的不断提升,为了使网络结构能更好的运用于小型设备上,对于轻量级网络结构的研究[2] [3]也与日俱增。

2.2. Inception-ResNet 架构

Inception-resnet 是由 Google 提出的一个卷积神经网络架构,其有 v1 和 v2 两个版本,该网络架构由

几个简单的卷积层、下采样层和几十个 Inception-resnet 模块所组成。Inception-resnet 模块的设计思路源于 GoogLeNet [10]中的 Inception 模块和 ResNet [11]中的残差思想，每个 Inception-resnet 模块均由一个 Inception 模块和残差直连组成。Inception 模块可以从特征图里获取到不同尺度的特征信息，残差思想则可以保留下更多前几层网络所获取到的特征信息，并使网络得以训练的更深。结合了 Inception 模块和残差思想的 Inception-resnet 模块可以使整体网络拥有更为优秀的特征提取能力。

2.3. PReLU

PReLU [12]是由何恺明等人提出的一种可以被训练的激活函数，其设计思路源于 LeakyReLU [13]。常用的激活函数 ReLU 是将特征图中的所有负值置为零，这可能会使特征图损失部分提取出的特征信息。LeakyReLU 则是将特征图中的所有负值乘上同一个固定的常数 α ，这样既能起到非线性的作用，同时减少特征图的信息损失。PReLU 则是更进一步将这个 α 设置为可训练参数，这样使得网络拥有更好的可训练性。由于 Leaky ReLU 和 PReLU 均是只增加一个参数，所以如果将常用激活函数 ReLU 改进为 Leaky ReLU 或是 PReLU，则可以更好的发挥网络在特征提取上的性能，而增加的参数量和计算量是近乎忽略不计的。

3. 网络架构设计

本文以 Inception-ResNet v2 为基础网络架构，设计出一个参数量远比 Inception-ResNet v2 少的网络结构 Lite-Inception-ResNet。Inception-Resnet v2 网络结构由一个基础模块和四十个 Inception-Resnet 模块还有两个下采样 Inception 模块所组成，其中基础模块由几个卷积层和池化层组成，下采样 Inception 模块则是包含了下采样操作的 Inception 模块。本文提出的 Lite-Inception-ResNet 则是在此基础之上适当减少参数量并将使用的 Inception-Resnet 模块数减少到了十三个。Lite-Inception-ResNet 的网络架构如表 1 所示。在模型的参数总量上，Inception-ResNet v2 有 55 M 的参数量，而 Lite-Inception-ResNet 缩减到了 6.5 M 的参数量，在总参数量上有 8.4 倍的缩减。

Table 1. The network architecture of Lite-Inception-ResNet

表 1. Lite-Inception-ResNet 的网络架构

输出特征图的大小	网络层	大小	步长	补边
79 × 79 × 64	卷积	7 × 7, 64	2	否
39 × 39 × 64	最大池化	3 × 3	2	否
39 × 39 × 48	卷积	1 × 1, 48	1	是
39 × 39 × 128	卷积	3 × 3, 128	1	是
19 × 19 × 128	最大池化	3 × 3	2	否
19 × 19 × 96	卷积	1 × 1, 96	1	是
19 × 19 × 256	卷积	3 × 3, 256	1	是
19 × 19 × 256	简化的 Inception-Resnet a 模块 × 4			
9 × 9 × 512	简化的下采样 Inception a 模块			
9 × 9 × 512	简化的 Inception-Resnet b 模块 × 5			
4 × 4 × 1024	简化的下采样 Inception b 模块			
4 × 4 × 1024	简化的 Inception-Resnet c 模块 × 4			
1 × 1 × 1024	全局池化 + 特征图平铺			
512	全连接层	512		

为进一步优化模型的性能,在经过详细对比实验的基础上,本文提出在 Lite-Inception-ResNet 中使用 PReLU 激活函数替换 Inception-ResNet 模型的 ReLU 激活函数。这个优化只增加了上百个参数,对于总参数量而言是微不足道的。通过实验验证了 PReLU 激活函数对本文提出的模型在性能上有明显提升。

网络模型在最后一层输出的人脸特征向量为 512 维的特征向量,这样可以使模型提取出更具有泛化性的特征来表征人脸的特征信息。训练时模型时在最后加的损失函数用的是 Softmax Loss,通过 Softmax Loss 来指导整个网络模型的训练和参数权值更新。

4. 实验及结果

4.1. 数据集及其处理

本文使用的训练数据集是 VGGFace2 [14]数据集。VGGFace2 是当前大规模人脸数据集之一,包括九千多个人的三百多万张人脸图像,且数据集质量较好,总体数据包含的噪声较少。

测试数据集使用的是 LFW [15] [16]数据集。LFW 数据集是专门用来测试算法性能的人脸图片数据集,它包含一万多无约束的人脸照片。其所有人脸照片都标注有人物名字,且进行过简单的处理,保证每张图片中只有一张完整人脸。在该数据集上可以较好的测试小型人脸识别网络模型的性能。

所有的训练数据集及测试数据集均通过 MTCNN [17]在同一标准下裁剪出标准的人脸图像,以使模型对人脸特征的提取更为一致。

4.2. 模型训练细节

模型的训练是在 NVidia GeForce RTX 2080Ti GPU 上进行的,使用 TensorFlow [18]框架来构建和更新网络模型。网络模型的输入图像大小设定为 163×163 ,值得一提的是,为更进一步提升模型的泛化性能,我们在训练的时候对图像进行了随机水平翻转。每个模型均训练 100 轮,每一轮训练 1000 次迭代并在该轮结束时在 LFW 数据集上验证一次,每次迭代的 batch size 大小设定为 100,即每次迭代用 100 张图片来训练网络。使用的优化方法是 Batch Gradient Descent,初始学习率为 0.05,每 20 轮下降一次学习率,每次下降学习率乘上 0.2。

4.3. 实验结果及分析

实验针对三个模型进行观察比对及分析,分别是 Inception-ResNet v2 模型、Lite-Inception-ResNet 模型和 Lite-Inception-ResNet (+PReLU)模型。在训练过程中,使用 TensorBoard [18]对其进行可视化来观察训练效果,如图 1 所示,其中 Value 是模型在 LFW 数据集上的正确率。从图中可以看出,Lite-Inception-ResNet 模型并没有因为网络的缩减而明显影响网络模型的训练,缩减后的模型在性能上并没有显著的劣化。

实验进行了两组比对,第一组是 Inception-ResNet v2 模型和 Lite-Inception-ResNet 模型的比对,其参数量、计算量及在 LFW 数据集上的测试结果比对如表 2 所示。第二组是 Lite-Inception-ResNet 模型和 Lite-Inception-ResNet (+PReLU)模型进行对比,其在 LFW 数据集上的测试结果比对如表 3 所示。

由表 2 可以看出,我们的模型在参数量和计算量上分别减少了 7.5 倍和 3.3 倍,其在 LFW 上的精确度有 0.43% 的损失。对网络结构的简化虽然导致了在精确度上有所降低,但是其影响并不显著,仍可以达到较高的识别精度。其 6.5 M 的参数量和 628 M 的计算量显著的减小了对计算资源的依赖,可以布置在更多计算资源少的设备上,使模型有更大的应用空间。

在对于模型激活函数的改进上,由表 3 可以看出,将激活函数从 ReLU 换成 PReLU,可以明显的提升 Lite-Inception-ResNet 模型的性能,且在参数量和计算量上几乎没有增加。该模型最终仅比 Inception-ResNet v2 模型相差 0.1% 的精确度,在参数量和计算量的消耗上则有数量级上的明显降低。实

验结果表明 Lite-Inception-ResNet 的简化是成功的, PReLU 的运用对 Lite-Inception-ResNet 起到了非常积极的作用, Lite-Inception-ResNet (+PReLU)在保持识别率几乎没有下降的情况下,显著减少了网络参数量和计算量,加大了模型的泛用性。

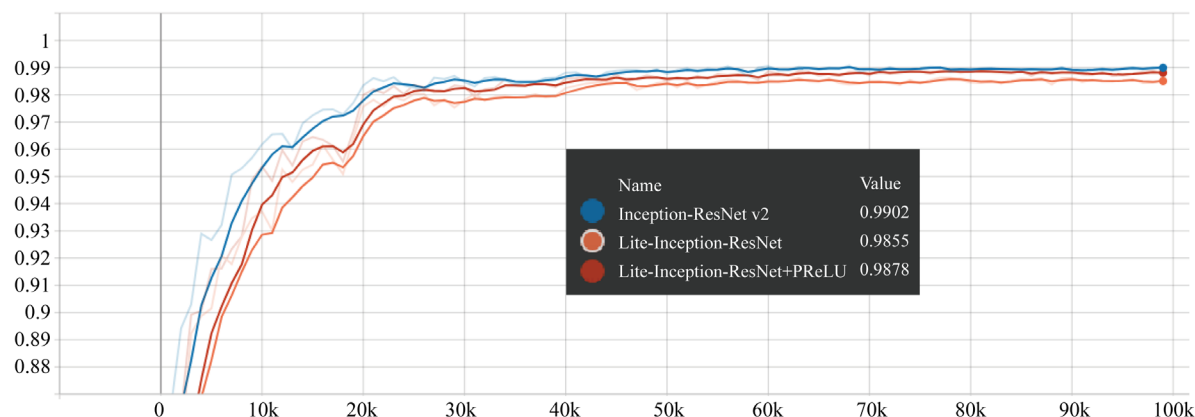


Figure 1. Validation results of the model on the LFW dataset during training

图 1. 训练过程中模型在 LFW 数据集上的验证结果

Table 2. The first set of comparative models and their test results on LFW

表 2. 第一组比较模型及其在 LFW 上的测试结果

模型	参数量	计算量	正确率
Inception-ResNet v2	55 M	2666 M	99.08%
Lite-Inception-ResNet	6.5 M	628 M	98.65%

Table 3. The second set of comparative models and their test results on LFW

表 3. 第二组比较模型及其在 LFW 上的测试结果

模型	正确率
Lite-Inception-ResNet	98.65%
Lite-Inception-ResNet (+PReLU)	98.98%

5. 结论

本文针对目前基于深度学习的人脸识别模型普遍需要大量计算资源的问题,设计了一种轻量级的人脸识别模型。该网络模型在保持有较好识别精度的情况下,显著减少了计算量和参数量,使得模型可以应用到计算资源受限的设备上,对于人脸识别算法的更进一步推广和普及具有积极意义。

基金项目

北京市教委科技计划一般项目(KM201811232024);北京信息科技大学促进高校内涵发展“信息+”项目-多源光谱生物特征活体识别平台建设;北京信息科技大学高教研究重点项目(2019GJZD01)。

参考文献

- [1] Han, S., Mao, H. and Dally, W.J. (2015) Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. arXiv Preprint arXiv:1510.00149.

-
- [2] Howard, A., Sandler, M., Chu, G., *et al.* (2019) Searching for Mobile Net v3. *Proceedings of the IEEE International Conference on Computer Vision*, 1314-1324. <https://doi.org/10.1109/ICCV.2019.00140>
- [3] Ma, N., Zhang, X., Zheng, H.T., *et al.* (2018) Shufflenet v2: Practical Guidelines for Efficient CNN Architecture Design. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 116-131. https://doi.org/10.1007/978-3-030-01264-9_8
- [4] Szegedy, C., Ioffe, S., Vanhoucke, V., *et al.* (2017) Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, 4-9 February 2017.
- [5] Taigman, Y., Yang, M., Ranzato, M.A., *et al.* (2014) Deepface: Closing the Gap to Human-Level Performance in Face Verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708. <https://doi.org/10.1109/CVPR.2014.220>
- [6] Parkhi, O.M., Vedaldi, A. and Zisserman, A. (2015) Deep Face Recognition. <https://doi.org/10.5244/C.29.41>
- [7] Wang, F., Cheng, J., Liu, W., *et al.* (2018) Additive Margin Softmax for Face Verification. *IEEE Signal Processing Letters*, **25**, 926-930. <https://doi.org/10.1109/LSP.2018.2822810>
- [8] Wang, H., Wang, Y., Zhou, Z., *et al.* (2018) Cosface: Large Margin Cosine Loss for Deep Face Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5265-5274. <https://doi.org/10.1109/CVPR.2018.00552>
- [9] Deng, J., Guo, J., Xue, N., *et al.* (2019) Arcface: Additive Angular Margin Loss for Deep Face Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4690-4699. <https://doi.org/10.1109/CVPR.2019.00482>
- [10] Christian, S., Wei, L., Yangqing, J., *et al.* (2015) Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-9.
- [11] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 26 June-1 July 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [12] He, K., Zhang, X., Ren, S., *et al.* (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 1026-1034. <https://doi.org/10.1109/ICCV.2015.123>
- [13] Maas, A.L., Hannun, A.Y. and Ng, A.Y. (2013) Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of ICML*, **30**, 3.
- [14] Cao, Q., Shen, L., Xie, W., *et al.* (2018) Vggface2: A Dataset for Recognising Faces across Pose and Age. 2018 13th *IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, 15-19 May 2018, 67-74. <https://doi.org/10.1109/FG.2018.00020>
- [15] Huang, G.B., Mattar, M., Berg, T., *et al.* (2008) Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.
- [16] Huang, G.B. and Learned-Miller, E. (2014) Labeled Faces in the Wild: Updates and New Reporting Procedures. Technical Report, Department of Computer Science, University of Massachusetts, Amherst, MA.
- [17] Zhang, K., Zhang, Z., Li, Z., *et al.* (2016) Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, **23**, 1499-1503. <https://doi.org/10.1109/LSP.2016.2603342>
- [18] Abadi, M., Agarwal, A., Barham, P., *et al.* (2016) Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv Preprint arXiv:1603.04467.