

基于深度学习的人脸识别方法研究进展

黄怡涓¹,左勘¹,孙频捷²

(1. 四川大学计算机学院,成都 610065; 2. 上海政法学院,上海 200000)

摘要:

基于深度学习的人脸识别方法主要分为两个方向:多分类和度量学习。多分类的方法在标记的已知类别上训练,在未知类别上测试。测试集上的识别性能严重受限于训练集上模型的表达能力,近几年的研究工作主要是基于分类损失函数的改进,动机在于让模型在闭集上学习的特征具有更高的辨别性。度量学习的动机在于学习新的表征使得类间距离大于类内距离,训练阶段不需要知道目标的具体类别,只需要标记类别差异。近几年对于度量学习方向的研究工作主要集中在损失函数的改进,调整不同的策略减小类内距离方差,同时增大类间距离方差,学习到的度量可以直接作为特征比对阶段的相似度。对这两个方向的研究工作进行归纳和总结,并对其他可能的方向做一些展望,为基于深度学习的人脸识别方法的进一步研究提供一些参考。

关键词:

深度学习; 人脸识别; 多分类; 度量学习

基金项目:

国家重点研发计划项目(No.2017YFC0820103)

0 引言

在身份认证场景中,人脸已经成为显著的生物特征,近几年对于人脸识别的研究也越来越多。人脸识别广泛应用于金融、安防、通行和手机解锁等场景。在计算机视觉领域,人脸识别已经是一个研究了很长时间的话题。基于深度学习的方法近几年取得了突破性进展,在 LFW 等公开数据集上性能远远超过了传统的手工特征方法。深度学习方法的动机在于学习一个从输入到特征空间的映射函数,从而获得人脸图像在高维特征空间中的表达。特征在高维空间中的性质严重依赖于损失函数的设计,为了使得人脸图像的特征表达对于身份具有很强的可分性,基于多分类的方法主要对分类损失函数进行不同的改进,基于度量学习的方法主要探索如何使得类间距离大于类内距离。

1 相关工作

1.1 卷积神经网络

在神经网络中,卷积神经网络(Convolutional Neu-

ral Network, CNN)广泛应用于图像重建、图像分类、目标检测和人脸识别等。虽然卷积神经网络在 20 年以前已经被提出^[1],最近几年硬件的提升才让训练深度卷积神经网络成为可能。最初的 LeNet5^[2]仅有 5 层卷积、VGG19^[3]增加到 19 层,最近几年高速神经网络^[4]和残差神经网络^[5]才超过 100 层卷积层。卷积神经网络通常由卷积层、采样层、激活函数和批归一化^[6]等组成,常见的网络结构如图 1 所示。卷积层利用采用卷积核滑动的方式获取下一个特征图的神元值,每一次卷积的计算都是一个线性操作。为了让网络具有非线性表达能力,卷积层之后通常接一个非线性函数也叫激活函数,常用的激活函数有 Relu、Tanh 和 Sigmoid 等。当特征图的空间尺寸过大时,一次卷积的计算时间比较高且神经元的感受野较小。为了获得更大的感受野和减少特征计算时间,激活函数之后会增加一个采样层,常用的采样层有最大值采样和均值采样。对于一个分类任务,网络最后会接一个全连接层,全连接层的维度

和类别数一致。最后通过 Softmax 获得每个类别的概率分布。

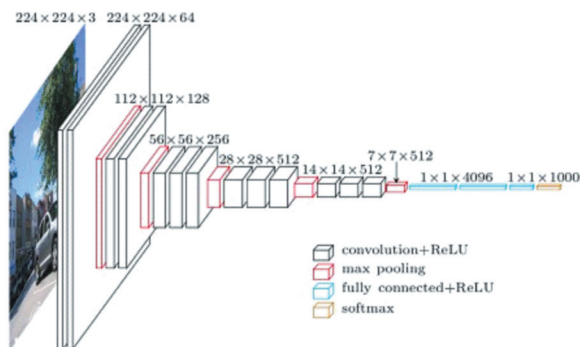


图1 卷积神经网络 Pipeline^[3]

1.2 度量学习

度量学习是学习一个距离函数直接作为相似度的度量,学习的目标是使得在特征空间中相似的目标离得近、不相似的离得远。而深度度量学习主要是利用深度神经网络提取输入的特征表达,在特征所在的高维空间定义一个度量,例如 L2 距离、Cosine 距离等,根据定义的度量性质设计损失函数进行端到端学习。度量学习常用于人脸识别等任务中,如图 2 所示,学习的目标是使得网络收敛后相似的目标在定义的度量下(以 L2 距离为例)离得更近。

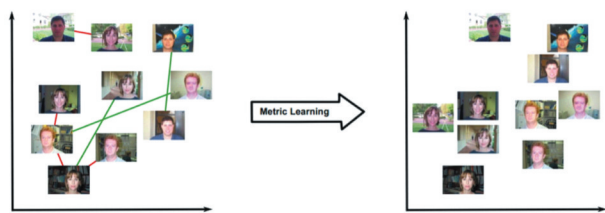


图2 度量学习在人脸识别中的应用

2 研究进展

2.1 基于多分类的人脸识别方法研究

多分类的方法是在闭集上训练一个多分类器,设计分类损失让闭集上的特征能具有强可分性。在测试阶段,底库和查询利用闭集上训练的模型来获取图像的特征表达。由于测试集是开集,基于多分类思路的识别性能具有如下缺陷:①更多的类别训练效果更好,但是更多的类别严重受限于硬件计算资源,网络最后

一层全连接层的参数 $W \in R^{d \times n}$ 随着 ID 数量 n 的增加而增加;②基于闭集学习到的特征可分性不一定在开集识别中足够的可分。

多分类方法广泛运用的损失函数是 Softmax loss,最近有很多工作提出如何让 Softmax loss 学习到的特征据有更强的可分性质[7,8,9,10,11,12,13,14,15]。Wen 等人提出 Center loss^[7],利用 L2 约束让 Batch 中每个样本“离类别中心尽可能的小”,也即是减少类内距离。这个方法存在一个明显的缺陷就是训练阶段更新类别中心时间成本太高。观察基于 Softmax loss 训练的模型,最后一个全连接层的参数近似等于每个类别的类中心特征。基于这个结论,文献[10,16]提出对分类的角度添加间隔来增大类间间隔并减少类内间隔,但是这些方法为了计算做了很多的近似,训练过程中并不稳定。CosFace^[11]直接对 Cosin 值添加间隔,在公开数据集上获得了比 SphereFace 更好的人脸识别性能。

目前在多个公开数据集上获得最佳性能的方法是 ArcFace^[17],网络框架如图 3 所示。对特征和最后一层全连接层的参数做了模长归一化之后,参数和特征的点积等价于 Cosine 相似度。利用 Arc-Cosin 函数计算出特征和参数(近似为类中心)的夹角,然后直接对角度增加间隔,从而获得更好的类间可分性。

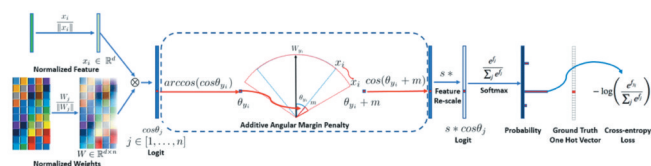


图3 ArcFace 网络框架^[17]

广泛使用的分类损失函数 Softmax loss,其计算公式(1):

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{w_j^T x_i + b_j}} \quad (1)$$

$x_i \in R^d$ 表示第 i 个样本的特征, y_i 表示分类结果。在文章里面 $d=512$, $w_j \in R^d$ 表示最后一层全连接层参数 $W \in R^{d \times n}$ 的第 j 列, $b_j \in R^n$ 是偏置项。 N 和 n 分别是 Batch size 和类别数。传统的 Softmax loss 广泛应用于深度人脸识别任务中^[18,19]。传统的 Softmax loss 函数并不能让学习到的特征具有更小的类内方差和更大的类

间方差,当类内方差变大的时候(姿态、光照和年龄等变化),识别性能会严重下降。

为了简洁,固定 $b_j=0$,然后对点积做恒等变换 $W_j^T x_i = \|W_j\| \|x_i\| \cos \theta_j$,其中 θ_j 是参数 W_j 和特征 x_i 之间的夹角。通过归一化,使得 $\|W_j\|=1$ 、 $\|x_i\|=1$,对特征做 Re-scale 使得模长是 s 。对特征和参数做归一化使得分类只依赖于特征和参数之间的夹角,学习到的特征被约束到一个半径是 s 的超球面上。特征和参数归一化之后的损失函数定义如公式(2)。

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos \theta_{y_i}}}{e^{s \cos \theta_{y_i}} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (2)$$

每个样本的特征都是在类别中心的长球面附近,对参数和特征的夹角添加一个惩罚因子 m 来增强类间差异和类内一致性,添加惩罚因子 m 后的损失函数定义如公式(3):

$$L_3 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (3)$$

作者选择了 8 个不同的 ID,每个 ID 包含 1500 张图像,训练了 Softmax loss 和 ArcFace,最后降维可视化后发现 ArcFace 可分性更好,对比见图 4。在多个公开数据集上,ArcFace 获得了最佳的识别性能。

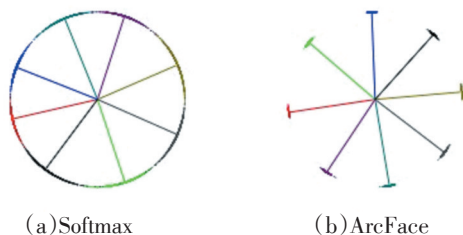


图4 Softmax 和 ArcFace 特征降维可视化对比图

2.2 基于度量学习的人脸识别方法研究

度量学习方法主要包含三个部分:特征提取网络、采样策略和损失函数。特征提取网络对于人脸识别任务主要是基于卷积神经网络,不同的计算平台可以根据需求选择或者设计合适 backbone。常用的损失函数有 Contrastive loss^[20]和 Triplet loss^[21]。Contrastive loss 的定义如公式(4):

$$l(i,j) = y_{ij} D_{ij}^2 + (1 - y_{ij}) [\partial - D_{ij}]_+^2 \quad y_{ij} \in \{0, 1\} \quad (4)$$

其中 i 和 j 分别表示样本 i 和 j , y_{ij} 表示 i 和 j 是否为相同 ID; 0 表示不是同一个人, 1 表示是同一个人;

D_{ij} 表示样本 i 和样本 j 之间 L2 距离。对比损失优化的目标使得同一个 ID 的距离更小,不同 ID 之间的距离区域常数 ∂ 。对比损失没有充分考虑到正负样本距离之间的大小关系,F.Schroff 等人^[21]提出将 Triplet loss 作为度量学习的损失函数。Triplet loss 定义了 anchor、正样本点和负样本点,学习的目标是使得 anchor 和正样本点之间的距离小于 anchor 和负样本点之间的距离,直观的理解可以参考图 5。



图5 Triplet 优化目标^[21]

Triplet loss 定义如公式(5):

$$L = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \partial]_+ \quad (5)$$

其中 N 是 mini-batch 的大小; i 表示第 i 各样本。Triplet loss 的优化目标是使得类间距离大于类内距离加上要给间隔 ∂ 。相比 Contrastive loss, Triplet loss 增加了类内距离和类间距离的大小关系,最后在公开的人脸数据集上的 1:1 和 1:n 的识别性能均超过了 Contrastive loss。其中 1:1 表示人脸验证, 1:n 表示人脸识别。虽然三元组有很多优点,但是三元组的采样和 Semi-hard 挖掘依然是影响模型性能的重要超参。

3 结语

基于多分类和度量学习的人脸识别方法从优化的目标上可以分为两种不同的方法。多分类的思路是利用大量的闭集数据训练 Softmax loss 等损失函数,让网络学习到的特征具有很强的可分性从而能泛化到开集识别数据集上。多分类方法学习到的人脸特征,其可分性质严重受限闭集上类别数量,由于网络的参数量和 ID 数量成正比,其类别数量严重的被计算能力限制住。基于度量学习的方法不管是训练还是测试都是面向开集,当训练集中样本较小的情况下依然可以训练。多分类和度量学习各有优势,针对人脸识别任务,不管是多分类还是度量学习的方法都能在公开 benchmark 上获得不错的识别性能。虽然在公开的数据集上多分类方法获得了最佳的识别性能,但是当需要解决

的问题规模很大的时候,例如工业界安防、通行和手机解锁等场景,可能度量学习更具有优势,这些场景天生就是开集识别问题。未来可以考虑将多分类和度量学习做多任务学习一起优化,利用多任务学习的方法优化模型。开集上的识别问题,往往需要模型推荐一个阈值,当两张比对的人脸分数超过阈值认为是同一个

人、低于阈值则不是同一个人。当样本分布变化很大的时候,例如多光照、多姿态、不同年龄和遮挡变化等,这些场景会导致类内和类间的距离方差过大。当类内和类间距离方差过大阈值便不能对齐,表现为正常光和暗光场景的 thr@far 差异过大,目前还没有阈值对齐方面的研究,未来可以考虑增加这方面的工作。

参考文献:

- [1]LeCun Y.,Boser B.,Denker J.S.,Henderson D.,Howard R.E.,Hubbard W.,Jackel L.D. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 1989, 1(4):541-551.
- [2]LeCun Y.,Bottou L.,Bengio Y.,Haffner P. Gradientbased Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [3]Russakovsky O.,Deng J.,Su H.,Krause J.,Satheesh S.,Ma S.,Huang Z.,Karpthy A.,Khosla A.,Bernstein M.,et al. Imagenet Large Scale Visual Recognition Challenge[J]. IJCV:1-7.
- [4]Srivastava R.K.,Greff K.,Schmidhuber J. Training Very Deep Networks[J]. NIPS, 2015:1-2+5.
- [5]HE K.,ZHANG X.,REN S.,SUN J. Deep Residual Learning for Image Recognition[J]. CVPR, 2016:1-6.
- [6]Ioffe S.,Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. ICML, 2015(3).
- [7]WEN Y.,ZHANG K.,LI Z.,QIAO Y. A Discriminative Feature Learning Approach for Deep Face Recognition[J]. ECCV, 2016:2-6+7.
- [8]DENG J.,ZHOU Y.,Zafeiriou S. Marginal Loss for Deep Face Recognition[J]. CVPR Workshop, 2017:2-6.
- [9]ZHANG X.,FANG Z.,WEN Y.,Li Z.,QIAO Y. Range Loss for Deep Face Recognition with Long-Tail[J]. ICCV, 2017(2).
- [10]LIU W.,WEN Y.,YU Z.,LI M.,Raj B.,SONG L. Sphereface:Deep Hypersphere Embedding for Face Recognition[J]. CVPR, 2017:1-7.
- [11]WANG H.,WANG Y.,ZHOU Z.,JI X.,LI Z.,GONG D.,ZHOU J.,LIU W. Cosface:Large Margin Cosine Loss for Deep Face Recognition[J]. CVPR, 2018:1-7.
- [12]WANG F.,LIU W.,LIU H.,CHENG J. Additive margin Softmax for Face Verification[J]. IEEE Signal Processing Letters, 2018:2-7.
- [13]CHEN B.,DENG W.,DU J. Noisy Softmax:Improving the Generalization Ability of DCNN Via Postponing The Early Softmax Saturation. CVPR, 2017(2).
- [14]WAN W.,ZHONG Y.,LI T.,CHEN J. Rethinking Feature Distribution for Loss Functions in Image Classification. arXiv:1803.02988, 2018(2).
- [15]QI X.,ZHANG L. Face Recognition Via Centralized Coordinate Learning. arXiv:1801.05678, 2018(2).
- [16]LIU W.,WEN Y.,YU Z.,YANG M. Large-Margin Softmax Loss for Convolutional Neural Networks. CML, 2016:2-3.
- [17]DENG Jian-kang, GUO Jia, XUE Nian-nan, Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. <https://arxiv.org/abs/1801.07698>.
- [18]Parkhi O.M., Vedaldi A., Zisserman A. Deep Face Recognition[J]. BMVC, 2015:1-2+6.
- [19]CAO Q., SHEN L., XIE W., Parkhi O.M., Zisserman A. Vggface2: A Dataset For Recognising Faces Across Pose and Age[J]. FG, 2018: 1-8.
- [20]SUN Y., CHEN Y., WANG X., TANG X. Deep Learning Face Representation by Joint Identification-Verification[J]. NIPS, 2014:1-6+7.
- [21]Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering[J]. CVPR, 2015:1-7.

作者简介:

黄怡涓(1994-),女,四川成都人,硕士,研究方向为多媒体计算、计算机视觉

通信作者:左劼(1977-),男,博士,副教授,研究方向为数据挖掘、数据库技术、分布式系统

孙频捷,社会学博士,研究方向为社交媒体、数据挖掘、网络恐怖主义

收稿日期:2019-11-19 修稿日期:2019-11-20

Research Progress of Face Recognition Based on Deep Learning

HUANG Yi-juan¹, ZUO Jie¹, SUN Pin-jie²

(1. College of Computer Science, Sichuan University, Chengdu 610065;

2. Shanghai University of Political Science and Law, Shanghai 200000)

Abstract:

Face recognition methods based on deep learning are mainly divided into two directions: multi classification and metric learning. The multi classification method is trained on the known classes of tags and tested on the unknown classes. In recent years, the research work is mainly based on the improvement of the classification loss function. The motivation is to make the characteristics of the model learning on the closed set more discriminative. The motivation of metric learning is to learn new representation so that the distance between classes is greater than the distance within the class. In the training stage, we do not need to know the specific category of the target, only need to mark the category difference. In recent years, the research on metric learning mainly focuses on the improvement of loss function, adjusting different strategies to reduce the same class distance variance, and increasing the different class distance variance. The learned metric can be directly used as the similarity of feature comparison stage. Summarizes the research work of these two directions, and makes some prospects for other possible directions, which provides some reference for the further research of face recognition method based on deep learning.

Keywords:

Deep Learning; Face Recognition; Classification; Metric Learning

(上接第 55 页)

作者简介:

吴东(1981-),男,广西合浦人,硕士,副教授,研究方向为智能信息处理

收稿日期:2019-10-28 修稿日期:2019-11-29

Learning Resource Recommendation Method Based on Social Relationship

WU Dong, CHEN Chang-jiang

(School of Information Engineering, Lingnan Normal College, Zhanjiang 524048)

Abstract:

Aiming at the Information Overload problem caused by the explosive growth of primary and secondary school teaching platform resources, designs a learning resource recommendation algorithm based on social relationship. The algorithm uses the social relationship in the teaching platform to consider the user's familiarity with friends and the similarity of interest, and realizes the recommendation of learning resources. The experimental results show that with the increase of user friends, the MAE and RMSE of the algorithm are reduced, and the recommended performance is enhanced. The research results are expected to provide new ways for learning resources in primary and secondary schools.

Keywords:

Recommended Technology; Learning Resources; Social Relationship