

SFace: 用于鲁棒性人脸识别的 S 形约束超球体损失

辛杨

摘要

SFace 这篇论文提出了类内和类间目标的适度优化来缓解传统严格优化思想所带来的过拟合问题, 并进一步提出了一种新的损失函数 Sigmoid-constrained 超球面损失 (SFACE)。Sigmoid 曲线精确地调整了类内梯度和类间梯度, 使训练样本在一定程度上得到优化。本人结合论文给出的 SFace-loss 的数理公式对该论文模型用 pytorch 进行了复现, 使用训练集: MS1MV2 对模型进行训练, 并在 LFW, YTF, CFP-FP, AgeDB-30, CALFW, CPLFW 五个数据集对论文实验进行了验证, 并得到了相似的结果。针对 SFace 削弱了 hard samples 的学习, 导致模型泛化较差的问题, 在 Sface 的基础上进行优化, 使分类器鼓励高质量 hard 样本学习, 不鼓励低质量 hard 样本学习, 提高模型泛化能力。

关键词: 人脸识别; 度量学习; 损失函数

1 引言

人脸识别算法相对来说已经很成熟了, 基于深度学习的 FR 可以分为两部分, 一是特征提取, 二是 Loss, 其中 Loss 是各 FR 算法改进的重点。目前本人研究主要集中在人脸识别的 Loss 优化, 到目前为止人脸识别方面较为有效的 loss 主要是 Softmax Loss^[1]、Triplet Loss^[2]、Center Loss^[3]、A-Softmax^[4]、CosFace^[5]、ArcFace^[6]。其中 Arcface 是目前相对较新的基于角度 loss 的研究成果, 也是目前本人研究关注的点, SFace^[7]这篇论文是基于 Arcface 的改进工作, 因此具有很大的参考价值。因此选择复现 SFace 这篇文章。

2 相关工作

面部识别领域, 是一个精细识别领域, Original Softmax Loss 只能进行普通的多分类任务, 分类间距小, 而且类内距离控制也不如 Center Loss 好, 也无法通过角度学习, Center Loss 能很好地控制类内距离, 但控制类间距离能力不足, 并且也无法通过角度进行更精细地学习, 故为了更好地实现在面部识别的高精度, 就必须要求算法可以实现两个功能, 即类内距离越小越好, 类间距离越大越好。近年来, 人脸识别研究领域的一个主要进展集中在 Softmax Loss 的改进上, 并且取得了很大的性能提升。主要的改进方法是做归一化以及增加类间 margin。Wen 等人在 ECCV 2016 中提出了一个新的 Loss: Center Loss, 用以辅助 Softmax Loss 进行人脸的训练, 为了让同一个类别压缩在一起, 最终获取更加 discriminative 的 features。L-Softmax 损失是一个灵活的学习目标, 具有可调的类间角 margin 约束。它提出了一种难度可调的学习任务, 随着 margin 变大, 难度逐渐增加。将向量相乘的方式变换为向量的模与角度的关系, L-Softmax 在 Softmax 的基础上增加了角度约束, 弥补了 Softmax 不能缩小类内距离的缺陷虽然 L-Softmax 的优势在于: 引入了对角度的约束, 使得类内更加紧凑, 从而产生更具判别性的人脸特征; 因为约束增强, 从而可以避免过拟合的问题。但是其仍然保留了模长项 (和) 导致模型还是会尝试增大模长来减小对角度的优化。A-Softmax 在 L-Softmax 的基础上将权重 W 做了归一化 $\|W\| = 1, b = 0$, A-Softmax 实验证明了类别样本相对越多, 模长越大, 所以对权重 W 做归一化可以减小样

本数量多少的影响。在 large margin Softmax 和 spherface 的 angular Softmax 中，margin 是以一种相乘 $\cos(m*\theta)$ 的方式出现的，这样带来的一个问题就是反向传播计算比较复杂，loss 优化不是很容易，再加上我们还要设置为了便于收敛的 lambda，有非常多的超参数需要处理，对于训练不是很友好。而 cosface 的出现则是弥补了这个缺陷，他是对 cosine loss 的结果设置加性 cosine margin，即 $\cos(\theta) - m$ ，这样优化比较简单，同时文章中还引入了 feature Normalization 以及相关的一些处理方式。上面是通过加大两个向量间的余弦距离来让分类间隔更大，而 Arcface[13] 的想法是用更加直接的角度距离来处理，给来做正则化。这样就会使得损失函数更直接的映射到空间位置上。公式如下，而几何结构来看，这两个 loss 的形状很像，但 Arcface 的是直接和相关的 loss，论文中也指出这个 Arcface 训练出来的模型准确率更高。

3 本文方法

3.1 本文方法概述

深度人脸识别由于大规模的训练数据库和快速发展的损失函数而取得了巨大的成功。现有的算法致力于实现一个理想的思想：最小化类内距离和最大化类间距离。然而，他们大都忽略了数据集中也有很多低质量的训练图像，不应该以这种严格的方式优化。SFace 这篇论文提出了类内和类间目标的适度优化来缓解这种严格优化思想所带来的过拟合问题，并进一步提出了一种新的损失函数 Sigmoid-constrained 超球面损失 (SFACE)。如图 1所示：SFace 对一个超球面流形施加了类内和类间约束，这两个约束分别由两个 Sigmoid 梯度缩放函数控制。Sigmoid 曲线精确地调整了类内梯度和类间梯度，使训练样本在一定程度上得到优化。因此，SFACE 可以在减少干净样本的类内距离和防止标签噪声过拟合之间取得较好的平衡，从而贡献出更鲁棒的深度人脸识别模型。

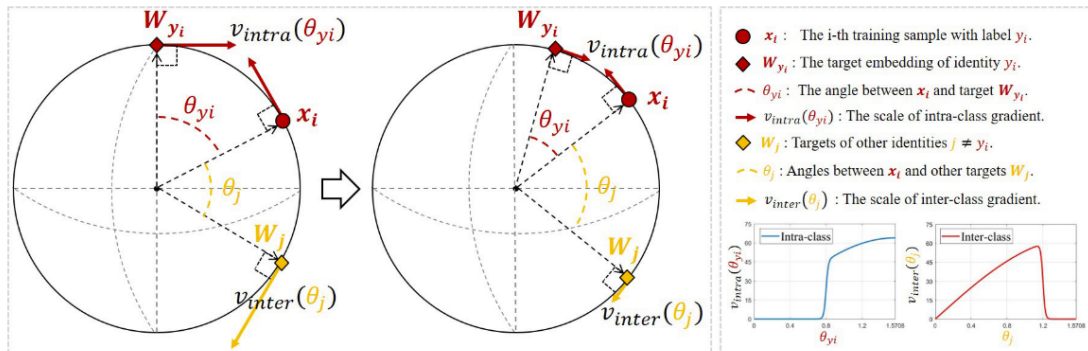
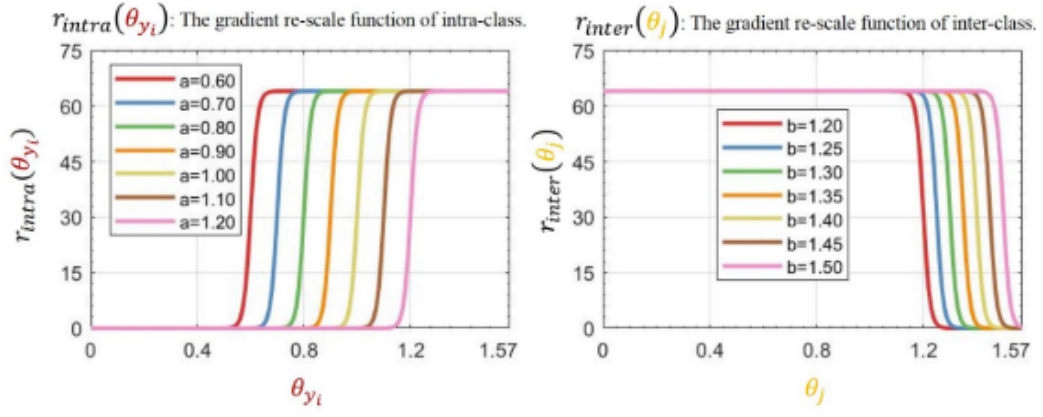


图 1: sigmoid 约束超球面损失的示意图

3.2 Loss 函数优化的核心理想

样本和目标嵌入的优化方向总是沿着超球面的切线方向，运动速度分别由两条 sigmoid 曲线控制。图 2展示了 SFace 所使用的类内、类间梯度尺度缩放函数 sigmoid 曲线。具体地说，深度特征及其目标中心的移动速度随着它们相互接近而逐渐减小，而其他目标中心的移动速度随着它们开始相互接近而迅速增加。SFACE 对具有精确控制的梯度和类内梯度的超球面流形施加类内和类间约束，从而在一定程度上优化了类内距离和类间距离。为了便于理解，论文中给出了一个简单易懂的例子。在标签噪声设置下，通过将带噪样本严格拖动到错误的标签标识，模型将过度适应标签噪声。相比之下，SFace 可以在一定程度上缓解这一问题，因为它以温和的方式优化了噪声样本。通过精确控制，干净的训练样本可以更早、更轻松地进行优化，而标签噪声可以留在后面。



(a)

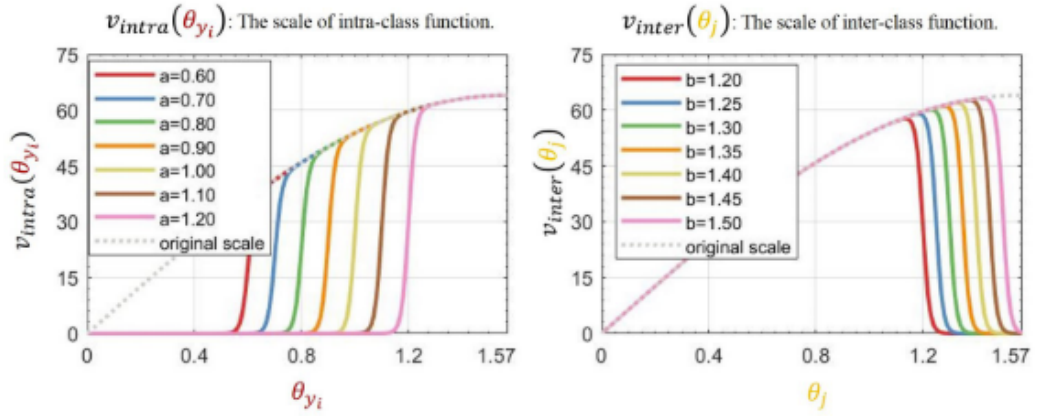


图 2: 类内类间的尺度缩放 sigmoid 曲线

4 复现细节

4.1 与已有开源代码对比

参考 BohdanNikoletti/SFaceCompare (c++) 以及 zhongyy/SFace (tensorflow) 的代码, 自己使用 pytorch 框架实现了 SFace, 并得到了与论文相似的实验结果。

4.2 实验环境搭建及实验结果

实验设置: 训练集: MS1MV2 验证集: LFW, YTF, CFP-FP, AgeDB-30, CALFW, CPLFW 模型: Resnet 50 + SFace 将论文中的实验在 MS1MV2 上训练模型, 并在 LFW, YTF, CFP-FP, AgeDB-30, CALFW, CPLFW 进行验证。具体实验界面如图 3 所示:

```

IR_50 Backbone Generated
=====
SFaceLoss()
=====
SGD (
Parameter Group 0
  dampening: 0
  lr: 0.1
  momentum: 0.9
  nesterov: False
  weight_decay: 0.0005

Parameter Group 1
  dampening: 0
  lr: 0.1
  momentum: 0.9
  nesterov: False
  weight_decay: 0
)
Optimizer Generated
=====

```

图 3: 操作界面示意

具体实验结果如图 4所示:

```

Learning rate 0.100000
Perform Evaluation on ['lfw', 'calfw', 'cplfw', 'cfp_fp', 'agedb_30'] , and Save Checkpoints...
(12000, 512)
[lfw][60000]XNorm: 40.44536
[lfw][60000]Accuracy-Flip: 0.98867+-0.00386
[lfw][60000]Best-Threshold: 1.35600
(12000, 512)
[calfw][60000]XNorm: 39.67394
[calfw][60000]Accuracy-Flip: 0.92367+-0.01327
[calfw][60000]Best-Threshold: 1.37000
(12000, 512)
[cplfw][60000]XNorm: 35.67893
[cplfw][60000]Accuracy-Flip: 0.85033+-0.02159
[cplfw][60000]Best-Threshold: 1.43900
(14000, 512)
[cfp_fp][60000]XNorm: 35.37814
[cfp_fp][60000]Accuracy-Flip: 0.91257+-0.01028
[cfp_fp][60000]Best-Threshold: 1.56000
(12000, 512)
[agedb_30][60000]XNorm: 39.47153
[agedb_30][60000]Accuracy-Flip: 0.92467+-0.01388
[agedb_30][60000]Best-Threshold: 1.49000
highest_acc: [0.9913333333333332, 0.9261666666666667, 0.8596666666666666, 0.9184285714285714, 0.9246666666666667]

```

图 4: 实验结果示意

4.3 界面分析与使用说明

Databases: The training databases, CASIA-WebFace, VGGFace2 and MS1MV2, evaluation datasets can be downloaded from Data Zoo of InsightFace. Train: Run the code to train a model.

(1) Train ResNet50, CASIA-WebFace, ArcFace.

```
python3 -u train-softmax.py --workersid0, 1 --batch_size256 --lr0.1 --stages50, 70, 80 --data_modedcasia --netIR50 --headArcFace --outdir./results/IR50-arc-casia2 > 1|tee./logs/IR50-arc-casia.log
```

(2) ResNet50, CASIA-WebFace, SFace.

```
python3 -u train - SFacetorch.py --workersid0, 1 --batch_size256 --lr0.1 --stages50, 70, 80 --data_modedcasia --netIR50 --outdir./results/IR50-sface-casia --parama0.87 --paramb1.22 > 1|tee./logs/IR50-sface-casia.log
```

5 总结与展望

与以往的类内距离最小化和类间距离最大化不同，SFace 引入了一种新的思想，即在一定程度上优化和类内类距离，以缓解对不完善训练数据库的过拟合问题。同时，SFace 与基于 Softmax 损失函数相比的优势在于类内和类间优化的精确控制能力。提出的 SFace 在过拟合和欠拟合之间取得了较好的平衡，进一步提高了深层人脸模型的泛化能力。本人结合论文给出的 SFace-loss 的数理公式对该论文模型用 pytorch 进行了复现，使用训练集：MS1MV2 对模型进行训练，并在 LFW，YTF，CFP-FP，AgeDB-30，CALFW，CPLFW 五个数据集对论文实验进行了验证，并得到了相似的结果。针对 SFace 削弱了 hard samples 的学习，导致模型泛化较差的问题，在 Sface 的基础上进行优化，使分类器鼓励高质量 hard 样本学习，不鼓励低质量 hard 样本学习，提高模型泛化能力。

参考文献

- [1] LIU W, WEN Y, YU Z, et al. Large-margin softmax loss for convolutional neural networks[J]. IEEE Transactions on Image Processing, 2016.
- [2] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 815-823.
- [3] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//European conference on computer vision. 2016: 499-515.
- [4] LIU W, WEN Y, YU Z, et al. Sphreface: Deep hypersphere embedding for face recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 212-220.
- [5] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large margin cosine loss for deep face recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5265-5274.
- [6] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4690-4699.
- [7] ZHONG Y, DENG W, HU J, et al. SFace: Sigmoid-constrained hypersphere loss for robust face recognition[J]. IEEE Transactions on Image Processing, 2021, 30: 2587-2598.