

联合监督损失在深度图像分类中的应用

童程¹
TONG Cheng

摘要

在深度图像分类任务中，大多数都是采用卷积神经网络模型与 SoftMax 损失监督训练完成的。但在实际的实验过程中会发现，单独使用 SoftMax 损失并不能很好地解决分类问题。为了进一步增强网络的分类性能，在使用 SoftMax 损失对网络训练的同时添加额外的联合监督损失函数，对比中心损失（contrastive center loss），其加强了网络的学习能力。该损失在训练过程中会给每个类学习一个深度特征中心，中心参数与模型参数一起进行优化，并通过对比学习的思想加强网络对类别特征信息的学习，在 LFW 数据集与 CUB 数据集的实验结果表明，通过联合监督损失的训练方法可以提高网络的分类性能。

关键词

图像分类；神经网络；损失函数；联合监督；参数优化

doi: 10.3969/j.issn.1672-9528.2024.05.019

0 引言

随着计算机科技的发展，神经网络^[1]在深度学习分类领域中表现出卓越的性能，显示了学习不同类别特征的强大力量。面对现有的视觉分类任务，深度学习模型能否高效运行主要取决于三个因素：丰富的训练数据、模型的构建及其从头到尾运行的完整学习框架。图1演示了使用单一损失函数^[2]进行监督训练深度网络模型的过程，该过程适用于大多数深度模型学习。输入图像经过数据预处理后经过网络加工与处理，得到类别的信息，最后计算最终损失函数并用于梯度回传来修改权重参数。



图1 深度图像分类模型训练流程

本文在上述流程的基础上，结合额外损失函数的添加与引用，将多损失联合监督方法应用在传统神经网络分类模型中，旨在进一步加强网络对特征信息的学习，帮助网络的参数优化与训练。最终，通过具体实验验证了多损失联合监督训练可以达到增强网络性能的目的。

1 网络模型

1.1 ResNet 模型

在深度学习的发展过程中，非常深的网络模型在训练优化过程中会遇到与梯度相关的问题，即反向传播过程中梯度逐渐趋近于零或者趋于无穷，这导致无法有效更新网络参数。ResNet^[3]的提出很好地解决了上述问题，为构建更深层次的

网络提供了可行性。该网络由微软亚洲研究院于2015年提出，是深度残差学习的里程碑，主要借助残差学习思想，解决了训练过程中的部分梯度问题，为更深层次的网络设计提供了可行的解决方案。

残差学习核心概念是通过使用残差块来进行深度残差函数学习。这样的设计使得网络能够直接学习残差，而非目标映射，有助于减轻梯度消失问题，使得更深层次的网络训练变得更加稳定。残差块通过跳跃连接将输入直接传递到输出，形成了一条捷径，使得网络能够更容易学习恒等映射。这种结构有助于保留和传播梯度，进一步缓解了梯度消失问题。可以假想，如果不做非线性变换，那么F就是恒等映射，这层网络有和没有是一样的，这就相当于将网络在某种意义上变浅了，该方案有效地过滤了深层网络中没有效果的结构层次。ResNet网络核心的模块结构如图2所示。

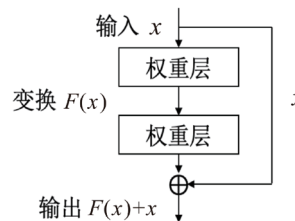


图2 ResNet 基础模块

此外，该网络还采用了批次归一化操作（batch normalization）^[4]代替以往网络中使用的标准化操作，这一操作有效地缓解了模型过拟合的不良现象。BN是深度学习中一种常用的正则化技术，旨在提高神经网络的训练速度和稳定性。在一般网络结构中，随着深度的逐渐增加，内部协变量的适当转移成为一个挑战。内部协变量转移是指网络中每一层输

1. 南京审计大学 江苏南京 211815

入分布的变化,这可能导致梯度消失或爆炸,使得网络训练变得困难。BN操作通过在每个训练最小批次上对网络的输入进行归一化,缓解了内部协变量转移的问题。首先,批归一化通过计算每个最小批次的均值和方差,然后对输入进行归一化,使其数据符合一般的标准正态分布,最后通过可学习的缩放和平移参数进行调整。这使得网络对输入数据的变化更加鲁棒,加速了训练过程,并且允许网络使用更高的学习率,从而提高了性能。批归一化不仅在提高训练速度和稳定性方面表现出色,还有助于防止过拟合,减少对精心调整超参数的需求。对于每一层 d 维的输入 $X=(x^1, x^2, \dots, x^d)$,BN会归一化它的每个维度,具体的归一化公式为:

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}[x^{(k)}]}} \quad (1)$$

总的来说,该网络的结构细节描述如下:该网络每个残差块都以一系列的卷积操作开始。对于每个卷积层,都会在卷积操作之后紧跟BN层。这有助于规范化卷积层的输入,提高网络的训练稳定性。在BN之后,通常会应用激活函数ReLU^[5]来引入非线性函数变换。这有助于网络学习更复杂的表示。该网络在深度学习发展史中有着重要意义,主要做出了以下创新:引入残差学习和跳跃连接的设计很好地解决了梯度方面的相关隐患,使模型的训练更加深入和稳定;网络深度的可扩展性强,该网络的设计使得网络深度可扩展性更好,可以轻松地构建具有更复杂结构的网络,有助于提取更关键的特征表示。

1.2 MobileNet 模型

传统的深度学习模型参数量较大,存储代价较高,训练时间与推理时间较长。MobileNet^[6]的提出在一定程度上缓解了上述问题。该模型是由谷歌研究团队在2017年提出,是一种轻量级网络模型,该设计在一定程度上保证模型性能的同时显著减小模型的运算成本和参数量,是轻量级深度学习的开山之作。该网络成功的主要原因得益于引入了深度可分离卷积,该结构不同以往,将标准卷积分解为深度卷积(depthwise conv)和逐点卷积(pointwise conv)两个步骤,这一操作显著地减少了网络整体的推理量和参数量。

不妨假设输入特征图为 F ,它的维度信息尺寸大小为 $D_{in} \times D_{in} \times M$,其中 D_{in} 表示输入正方形矩阵信息的高和宽, M 表示输入信息的通道数。经过卷积操作后得到的特征图为 F' ,它的尺寸大小为 $D_{out} \times D_{out} \times N$,其中 D_{out} 为输出矩阵的高和宽, N 为输出通道数。假设卷积核为 $D_s \times D_s$ 的大小,那么传统卷积的计算代价则为:

$$D_s \cdot D_s \cdot M \cdot N \cdot D_{in} \cdot D_{in} \quad (2)$$

上述计算成本表现为输入输出通道数的乘积,以及输入矩阵大小和卷积核大小的乘积。为了很好地降低上述计算成本,使用深度可分离卷积打破输入输出通道数与卷积数之间的固有关系。与标准卷积相比,该新型卷积在保证尺寸维度信息不变的前提下,通过两个步骤代替标准卷积,分别为深度卷积和逐点卷积。首先深度卷积核的通道数为1,卷积核数量与输入通道数量一致,然后通过 1×1 的逐点卷积将维度调整为与标准卷积一致的输出通道数量。如此一来,该卷积操作的计算成本可以表示为:

$$D_s \cdot D_s \cdot M \cdot D_{in} \cdot D_{in} + M \cdot N \cdot D_{in} \cdot D_{in} \quad (3)$$

将上述计算成本与标准卷积计算成本进行对比,可以大致估计出减少了多少计算量,具体计算为:

$$\frac{D_s \cdot D_s \cdot M \cdot D_{in} \cdot D_{in} + M \cdot N \cdot D_{in} \cdot D_{in}}{D_s \cdot D_s \cdot M \cdot N \cdot D_{in} \cdot D_{in}} = \frac{1}{N} + \frac{1}{D_s^2} \quad (4)$$

可以清晰地看出,该网络使用深度可分离卷积极大地降低了参数的运算量,其计算量比传统卷积少8至9倍,这是令人兴奋的,因为在节省了数倍的参数条件下,网络精度也得到了保证。

该网络结构主要基于上述独特卷积进行搭建,但是第一层依然还是基本卷积,详细的具体结构搭建则是由研究人员经过大量探索得出的。一般来说,所有层后会紧接着一个BN标准化层和ReLU非线性激活函数层,只有最终的线性分类全连接层不使用激活函数,而是直接经过SoftMax层开展分类任务,得出类别系数。

2 损失函数

2.1 SoftMax 损失

给定 N 个最小批次训练样本 $D_N = \{(X_i, l_i) | i=1, 2, \dots, N\}$,其中 X_i 是输入图像, $l_i \in \{1, 2, \dots, K\}$ 是其 K 类分类问题的对应标签。考虑网络模型特征提取层的输出,在这种情况下, $x_i \in \mathbb{R}^d$ 可以表示为网络提取得到的深度特征向量。传统的SoftMax损失^[7]是一个复合架构,总共三个部分分别为线性层、激活层与损失层。计算所有类别的概率估计,并随后衡量预测误差。具体来说,深度特征 x_i 作为全连接层的输入,经过线性转换,从而根据以下表达式将其映射到原始系数向量 z_i :

$$z_i = W^T x_i + B \quad (5)$$

式中: $W = \{w_1, w_2, \dots, w_k\} \in \mathbb{R}^{d \times K}$ 表示全连接层的权重, $B = \{b_1, b_2, \dots, b_k\} \in \mathbb{R}^{K \times 1}$ 偏置参数。然后,推导出概率分布 $p(l=j|x_i)$,并通过SoftMax激活函数处理规范所有类别概率。最后,使用交叉熵损失来计算预测类别标签和实际类别标签 l_i 的损失值,从而最终表述SoftMax损失函数,表示为 L_S ,方程为:

$$L_s = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K l_i \log p(l=j|x_i) \\ = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_i^T x_i + b_i}}{\sum_{j=1}^K e^{w_j^T x_i + b_j}} \quad (6)$$

上述损失最大的优点就是适用范围较广，并且拥有可观的计算代价且易于实施。但是在部分分类任务中，仅仅使用该损失是不够的，所以使用其他损失协助该损失进行训练优化是有意义的。

2.2 对比中心损失 (contrastive center loss)

网络模型的应用显著提高了视觉处理中图像分类识别任务的性能。SoftMax 损失通常作为监督信号，但它只会惩罚分类损失。对比中心损失^[8]作为一种辅助监督信号，它可以进一步增强特征的判别力，因为它为每个类学习一个深度类别中心。对比中心损失同时考虑了类内差异与类间的可分离性质。该方法的主要目的是惩罚以下两者之间的比值：训练样本到相应类别中心的距离，以及训练样本到非对应类别中心的距离之和。具体的损失方程为：

$$L_{cl-c} = \frac{1}{2} \sum_{i=1}^N \frac{\|x_i - c_{l_i}\|_2^2}{\left(\sum_{j=1, j \neq l_i}^K \|x_i - c_j\|_2^2\right) + \delta} \quad (7)$$

为了与 SoftMax 损失的部分参数进行统一， N 表示最小批次的样本数， x_i 表示经过网络处理得到的深度特征， c_j 表示该损失为每个类别学习的深度特征中心， K 表示待分类问题的类别总数。 δ 是防止分母为 0 的参数，默认设置为 1。显然，对比中心损失只需要对网络稍加修改就可以直接用于深度神经网络。在整个网络训练过程中，类别中心 c_j 与网络参数将一起进行优化与更新。

3 实验

为了使用联合监督损失对 ResNet34 网络与 MobileNet 网络进行训练优化，首先整合了损失函数方程，将其称为联合监督损失，具体公式为：

$$\text{Loss} = L_s + \lambda L_{cl-c} \quad (8)$$

参数 λ 控制对比中心损失的系数，用来平衡整体网络损失。使用联合监督损失进行卷积神经网络参数优化与训练的具体流程如图 3 所示。

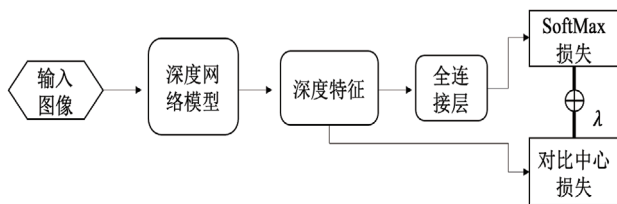


图 3 联合监督训练流程图

在图 3 中可以清晰看出联合监督损失计算的整体流程。

首先，网络的主干分支不会受到影响，会正常计算 SoftMax 损失。此外，为了联合监督损失，对比中心损失的正常实施会对所采用的网络模型结构中的深度特征层添加一个捷径分支，该分支就是用来专门计算额外监督损失的。最终，将两个损失合并，一起对网络参数与中心参数进行优化与更新。需要注意的是，在后续的实验中除非具体说明，否则中心对比损失的系数一直设置为 1，并且无论网络参数还是中心参数，都使用 SGD^[9] 优化器进行优化。更重要的是，联合监督损失只在训练过程中使用，后续的推理只使用 SoftMax 损失。

3.1 LFW 数据集实验

为了评估联合监督方法在人脸识别领域的有效性，使用 LFW^[10] 数据集进行了实验试验。LFW 数据集是公认的面部分类任务存储库，该数据集总共包括 13 233 张面部图像，包括 5749 个不同的个体。数据集中的每个图像都用各自的姓名进行精心注释。为了方便视觉分类任务，选择了拥有最丰富样本表示前 20 名个人进行实验。随后，将数据集划分为训练与测试集，保持 8:2 的比例。值得注意的是，数据集中的每张图像都是彩色图片，具体尺寸为 250×250 像素。

实验详细设置：实验在两个常用的模型上进行。数据预处理方面，首先对训练数据进行随机裁剪和缩放操作，产生 224×224 像素大小的三通道图像，这有助于增强网络的训练过程。此外，随机水平翻转程序应用于图像，从而在数据集中引入了更高的可变性。这种刻意的多样化作为一种机制，使模型能够捕捉面向各个方向的特征。相比之下，对于测试数据集，故意省略采用随机翻转，只采用中心裁剪技术。选择这种方法是为了避免应用数据增强。随后的网络训练方案总共跨越了 300 个 epoch，网络的学习率微调为 0.01，批量大小设置为 64。对比中心损失的学习率设置为 0.5。

表 1 LFW 数据集实验的 top-1 与 top-5 准确率

损失方法	模型	
	MobileNetV2	ResNet34
SoftMax Loss	91.15/99.19	93.16/99.60
Joint Loss	92.49/99.46	95.17/99.73

实验结果分析：通过实验的结果可以清晰地发现，无论哪种网络模型，使用联合监督损失方法都使基准损失的准确率得到了提升。在 MobileNetV2 模型中，top-1 精度提高了 1.3%，top-5 提高较少，只有约 0.3%。在 ResNet34 模型中提升更加明显，top-1 提高了 2%，但是 top-5 准确率提升非常低，因为该准确率已经趋于 100% 了。

3.2 CUB 数据集实验

为了进一步验证联合监督损失方法在深度分类领域的有效性, 选择了一个更复杂的任务进行实验。CUB-200-2011^[11]数据集是细粒度图像分类广泛采用的基准。该实验数据有共计 11 788 张图像, 包含 200 个类别, 每个类别代表不同的鸟类物种。具体来说, 该数据集被划分为两组, 包括 5994 张指定用于训练的图像和 5794 张用于评估分类模型的图像。在开展实验前, 对 CUB 数据集进行了筛选操作, 降低了其规模。具体来说, 重点关注了数据集中最初的 100 种鸟类。为了方便分类任务, 数据集遵守 8:2 的比例将其划分为两个子集, 分别作为训练数据集与测试数据集。

实验详细设置: 鉴于 CUB 数据集中图像尺寸的不固定性, 通过修改尺寸操作将所有图像统一调整为 224×224 的彩色三通道图像, 以便后续的训练与测试。随后, 使用随机裁剪, 并在裁剪区域周围应用 4 像素填充, 作为数据增强策略的一部分。此外, 通过引入一定概率执行的图像随机水平翻转来增强网络学习。同样上述操作仅适用于训练数据集, 包括尺寸标准化和增强技术, 而测试数据集仅进行尺寸修改。随后在设置网络学习优化器时, 初始学习率设定为 0.1, 在特定 epoch 学习率会降低, 特别是 50、150 和 250 时, 缩放到原来的 0.1 倍。同样, 对比中心损失的类别中心学习率设置为 0.5。

表 2 CUB 数据集实验的 top-1 与 top-5 准确率

损失方法	模型	
	MobileNetV2	ResNet34
SoftMax Loss	54.50/79.68	55.35/81.05
Joint Loss	55.93/80.95	56.45/82.07

实验结果分析: 通过上述表格可以看出, 该数据集的分类难度较大, 整体的分类准确率偏低, 但是也有效证明了联合监督训练的有效性。在 MobileNetV2 模型中, top-1 准确率提高了 1.4%, top-5 提高了约 1.2%。在 ResNet34 模型中, top-1 与 top-5 准确率分别提高了 1.1% 与 1%。

4 总结

在本文中提出了应用于神经网络模型的联合监督损失训练方法, 该方法将传统 SoftMax 损失与对比中心损失一起整合应用在深度网络模型的训练中。主损失按照正常流程对网络参数进行优化和训练, 额外的监督损失通过利用网络深度特征分支进行计算深度特征损失, 帮助网络训练。通过具体的模型与数据实验, 证明了联合监督损失训练方法可以提高网络的分类准确率。

参考文献:

- [1] 邓茗春, 李刚. 基于人工神经网络的智能数据分析方法研究[J]. 信息技术与信息化, 2008(5):50-51+33.
- [2] 陈聪. L1 正则化与 pinball 损失函数的极限学习机[J]. 信息技术与信息化, 2023(3):37-40.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway:IEEE, 2016: 770-778.
- [4] BJORCK N, GOMES C P, SELMAN B, et al. Understanding batch normalization[J]. Advances in neural information processing systems, 2018(12):7705-7716.
- [5] CHEN Y, DAI X, LIU M, et al.Dynamic relu[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 351-367.
- [6] 史俊冰, 师思汝, 王明宇, 等. 基于 MobileNet 的救护车识别系统设计与实现[J]. 信息技术与信息化, 2022(9):106-109.
- [7] PANG T, XU K, DONG Y, et al. Rethinking softmax cross-entropy loss for adversarial robustness[EB/OL].(2019-05-25)[2024-01-15].https://arxiv.org/abs/1905.10626.
- [8] QI C, SU F. Contrastive-center loss for deep neural networks[C]//2017 IEEE international conference on image processing (ICIP). Piscataway:IEEE,2017: 2851-2855.
- [9] ZHANG S, CHOROMANSKA A E, LECUN Y. Deep learning with elastic averaging SGD[C]//Advances in Neural Information Processing Systems 28,vol. 1. La Jolla:Neural Information Processing Systems,2015:685-693.
- [10] SENGUPTA S, CHEN J C, CASTILLO C, et al. Frontal to profile face verification in the wild[C]//2016 IEEE winter conference on applications of computer vision (WACV). Piscataway: IEEE,2016: 1-9.
- [11] KRAUSE J, JIN H, YANG J, et al. Fine-grained recognition without part annotations[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE, 2015: 5546-5555.

【作者简介】

章程(1999—), 男, 江苏扬州人, 硕士研究生, 研究方向: 计算机视觉、模式识别、深度学习。

(收稿日期: 2024-03-18)