

基于多任务学习的有限样本多视角三维形状识别算法



周子钦 严华

四川大学电子信息学院 成都 610065

(zizhou@stu.scu.edu.cn)

摘要 随着三维扫描技术的快速发展,三维形状分析得到了学术界的广泛关注;尤其是深度学习在计算机视觉上取得的显著成功,使得基于多视图的三维形状识别方法成为了目前三维模型识别的主流方式。已有研究表明,三维数据集的数量对于最终的分类精度是一个非常重要的影响条件。然而,由于专业三维扫描设备的限制,三维形状数据难以采集。实际上,现有的公共基准三维数据集的规模远远小于二维数据集,三维形状分析的发展因此受到阻碍。为了解决这一问题,文中主要研究在极少数数据样本情况下,三维形状识别问题的优化策略。受多任务学习的启发,搭建了多分支的网络结构,并引入基于度量学习的辅助比较模块,用于挖掘类内和类间的相似性和差异性信息。网络模型包括主支路与辅助支路,分别使用不同的损失函数对应不同的训练任务,并使用权重超参数平衡多项损失。主支路获得预测分类,使用交叉熵损失函数进行更新;辅助支路得到不同样本间的相似性得分,使用均方差损失函数进行更新。为保证特征向量被投影到同一个空间中,主、辅助支路共享相同的特征提取模块,在训练阶段共同更新参数,在测试阶段仅使用主支路获得的分类结果。在两个公开的三维形状基准数据集上的大量实验结果表明,所提网络结构与训练策略相比传统方法,在少样本的情况下可以显著提高特征模块对不同类别的区分能力,获得更优的识别结果。

关键词: 多视图三维形状;有限样本;辅助支路;多任务学习;三维形状识别

中图分类号 TP391.413

3D Shape Recognition Based on Multi-task Learning with Limited Multi-view Data

ZHOU Zi-qin and YAN Hua

College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

Abstract With the rapid development of 3D scanning technology, 3D shape analysis has been widely concerned by researchers. Especially with the significant success of deep learning in computer vision, the approaches of 3D shape recognition based on multi-view have become the dominant methods. In the previous work, we notice that the amount of 3D shapes is essential for the recognition accuracy. However, due to the limitation of professional 3D scanning equipment, the 3D shape data is hard to collect. In fact, the scale of existing benchmark datasets is far smaller than that of 2D datasets which impedes the development of 3D shape analysis. In order to solve this problem, we mainly develop an optimal strategy of 3D shape recognition with limited data. Inspired by multi-task learning, we develop a novel network with multiple branches and construct an auxiliary comparison module based on metric learning to exploit the similarity and discrepancy between different samples intra-class and inter-class. The proposed network mainly includes a primary branch and an auxiliary branch, which respectively use disparate loss functions with different training tasks and hyper-parameter to balance different loss items. The primary branch aims to obtain the prediction of classification and uses Cross Entropy Loss function to train it. While the similarity scores of different samples are calculated by the auxiliary module, and the Mean Square Error is used to update this branch. Both two branches share the same feature extractor to project all samples into the same representation space and train the structure jointly in training phase, while the primary branch would be used in testing phrase to calculate the accuracy. Extensive experimental results have reported on two public 3D shape benchmark datasets which demonstrate the effectiveness of our proposed architecture to enhance the discriminative power and achieve better performance compared with traditional methods, especially in the situation where merely has limited multi-view data.

Keywords Multi-view 3D shape, Limit data, Auxiliary branch, Multi-task learning, 3D shape recognition

1 引言

随着人们对三维传感技术的日益关注,三维形状识别成

为了三维分析领域中非常重要的课题之一。尽管三维形状识别的发展前景广阔,但仍存在一些亟需解决的问题,其中三维形状数据的缺乏是三维分析领域的核心问题之一。例如,三

收稿日期:2019-07-23 返修日期:2019-11-06 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:四川省重点研发项目(2019YFG0409)

This work was supported by the Key Research and Development Program of Sichuan Province (2019YFG0409).

通信作者:严华(yanhua@scu.edu.cn)

维形状数据库 ModelNet^[1]共包含 127915 个 CAD 三维模型,而二维数据库 ImageNet^[2]具有数千万张标注后的 2D 图像。三维数据缺乏的原因在于,二维图片的采集相对便捷,而采集三维形状数据时需要专业的三维扫描设备。然而,传统的神经网络训练方法需要大量的标记数据才能取得良好的训练效果,在三维形状分析领域也是如此。

三维形状数据获取的困难性,使得现有的三维数据集包含的种类和样本数目都远远不足,这在一定程度上阻碍了三维形状识别的发展。大量的实验结果表明,在样本较少的情况下,三维形状识别的精度明显下降。然而,受类脑研究的启发,人类可以通过几个样本甚至一个样本去学习新类别,学者认为深度学习也有潜力完成此任务。

因此,我们考虑一种具有实际意义但极富挑战性的情况,即每个类别仅包含极为有限的三维形状样本。本文研究的目的在于训练出一种有效的方法,利用有限的三维形状数据充分挖掘特征信息,从而获得出色的识别性能。

受多任务学习的启发,我们在分类任务的基础上构造了一个基于度量学习的额外分支,并对数据输入策略进行了改进,同时提出了一种新颖的损失函数求公式用于网络训练。所提算法在学习新的三维形状类别时,仅须收集极少三维样本即可获得良好的识别效果;并且通过训练有限的三维形状,可以充分降低运算代价,有利于三维分析的实际应用与普及。

本文的主要贡献如下:

- (1)对输入数据策略进行了重构。在训练过程中,将每一个小批量训练样本随机划分为了支撑集和查询集。
- (2)特征提取器由主支路与辅助支路联合学习,获得了更强的特征表达能力。
- (3)在三维形状数据有限的情况下,采用带辅助支路的多任务学习方法。主支路基于交叉熵损失函数,来获得正确的识别结果;辅助支路基于均方误差损失函数的度量学习,通过比较两个子集之间的特征向量得到相似得分,进而挖掘出样本间的相似性与差异性信息。

2 相关工作

本文方法基于多支路的多任务学习,解决了多视图三维形状数据有限情况下的识别问题。

2.1 三维形状识别

随着对三维形状分析和计算要求的提高,计算机视觉领域提出了大量的三维形状识别、检索、匹配方法^[3]。根据输入数据的类型,可以将三维形状识别方法大致分为 3 类:基于点云的方法^[4-5]、基于体素的方法^[1,6]和基于多视图的方法^[7-8]。相比基于点云和体素的表征方式,基于多视图的表征方式具有运算复杂度低、存储开销小等优势,在现有的三维模型识别方法中取得了更高的精度。这种表征的优势在于能够将三维模型二维化,使其能够直接利用现有二维图像领域的诸多处理技术。

基于多视图的方法将三维形状投射到一系列的二维图像中,即通过固定度放置一个“虚拟相机”,从不同的角度表达三维形状的视觉外观。Su 等首次提出一种新颖的多视图卷积神经网络(Multi-view Convolutional Neural Networks, MVC-NN)^[9],用于从不同角度的 3D 形状视图中学习识别特征;然

后使用最大池化操作将不同视角获得的特性融合成一个紧凑的特征描述子。本文基于 MVCNN 网络结构模型,使用传统交叉熵损失函数指导训练作为对比实验结果基准 1 (baseline1)。

2.2 三元组损失函数

为了提取富有表现力的深度特征,Schroff 等^[10]提出了一种新的三元组损失函数(triplet loss)来进一步增强网络的区分能力。Wang 等^[11]已证实,当样本数量非常少时,使用三元组损失函数可以提高识别准确度。He 等^[12]将三元组损失函数引入多视角三维形状识别中,相比传统交叉熵损失函数取得了更好的预测精度。

为了比较单一支路与多支路在构建成对样本时的效果差异,在实验部分构建三元组对,并引入 triplet loss 指导训练作为实验结果基准 2(baseline2)。此方案同样使用 MVCNN 网络模型作为基础结构,其三元组损失函数可公式化表示为:

$$\mathcal{L}_t(\mathbf{a}, \mathbf{p}, \mathbf{n}) = \max\{d(\mathbf{a}, \mathbf{p}) - d(\mathbf{a}, \mathbf{n}) + m, 0\} \quad (1)$$

其中, $(\mathbf{a}, \mathbf{p}, \mathbf{n})$ 即为三元组,表示 anchor, positive, negative 样本所得特征向量; $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$ 表示使用欧氏距离计算向量距离; m 代表 margin 值(设为 1.0)。此损失函数使得经过学习后,同类样本的 positive 样本更靠近 anchor 样本,而不同类的 negative 样本远离 anchor 样本。

2.3 多任务学习

多任务学习(Multi-task Learning, MTL)是迁移学习(Transfer Learning)的一种,从自然语言处理、语音识别到计算机视觉,在深度学习方法的许多应用中都取得了巨大的成功。大量的经验表明,当不同任务间彼此存在相关性时,相比仅仅使用原始单一任务,MTL 可以通过在多任务上引入诱导偏置来改善神经网络,提高模型的泛化能力,提升目标任务的效果。

MTL 方法大多采用硬参数共享或软参数共享隐藏层两种方法。基于硬参数共享的方法在多个任务上直接共享神经网络的同一部分,避免了对原始任务的过拟合。基于软参数共享的方法中,不同任务使用不同的网络模型,但是网络参数将距离(L1, L2 等)作为正则化约束。Long 等^[13]提出的基于硬参数共享卷积层和具有矩阵先验信息的特殊任务层的深度关系网络,增强了不同任务之间关系的学习能力。

3 有限三维样本识别算法的优化

受多任务学习的启发^[14-15],我们在三维形状分类任务的基础上构造了一个基于度量学习的额外分支,称为辅助比较支路。完整的网络结构如图 1 所示。

本文所提框架主要包含 3 个模块:特征提取模块、主分类模块和辅助比较模块。我们将每一次迭代时输入的小批量训练样本数据集划分为支撑集和查询集,同时引入基于辅助支路的多任务学习方法。主支路获得三维形状的预测分类结果,并使用交叉熵损失函数指导训练;辅助支路构建同类与不同类的特征组合作为额外引入的约束条件,挖掘类内与类间不同三维形状样本间的相似性与差异性,通过均方误差指导训练。

在训练阶段,为了共同训练整个网络模型参数,主支路与辅助支路的损失函数通过参数 λ 进行加权求和操作,并共同

训练特征提取模块,以提高特征提取模块的表达能

力。在测试阶段,则仅由主支路获得分类预测结果。

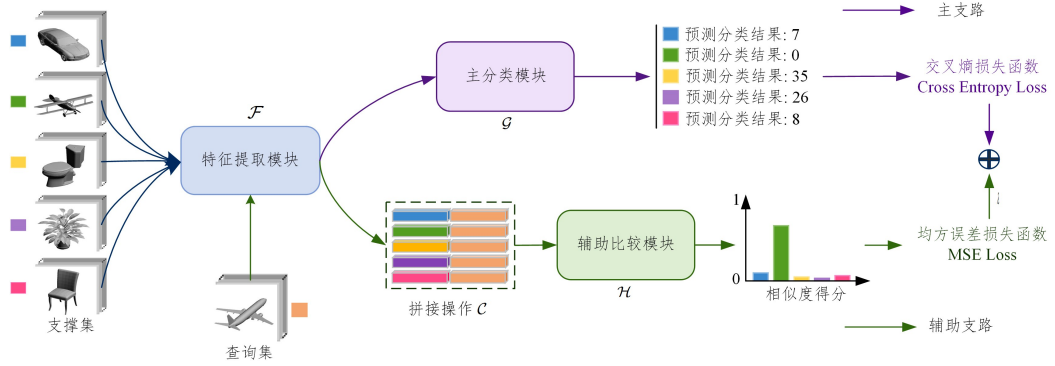


图1 多任务学习的有限样本多视角三维形状识别算法概述(电子版为彩色)

Fig. 1 Overview of algorithm of multi-task learning with limited multi-view data

3.1 问题定义与训练样本输入策略

与实际三维世界相比,现有的三维形状数据集的类别极其不完整;同时,在公共基准数据集中,三维形状的规模也远远小于二维图像数据库的规模。此外,三维形状比二维图像的维数更高,导致了较大的计算复杂度,不可避免地给三维形状识别带来了挑战。通过大量的对比实验可以看到,在极少量多视图三维形状数据的情况下,在测试集上的三维形状识别精度大幅下降。

本方案基于多任务学习,在辅助任务中需要构建同类与不同类样本间成对的组合特征。因此,本方案对输入数据策略进行了改进,将每一个小批量训练样本划分为支撑集(Support Set)与查询集(Query Set)。

在每一个小批量训练集中,从缩减后的训练集中随机选择C类三维形状(每一个形状具有K个标记样本)作为输入数据。然后,在每个类别中随机选择S(S<K)个样本组成成支撑集 $S = \{(X_i, y_i)\}_{i=1}^m$ ($m = S \times C$);每个类别其余的Q(Q=K-S)个样本组合构成查询集 $Q = \{(X_j, y_j)\}_{j=1}^n$ ($n = Q \times C$)。

在小批量训练样本中划分支撑集和查询集旨在与网络模型结构和训练方案结合,构造样本对来充分挖掘类内与类间的相似性和差异性。

3.2 三维形状特征的提取

为了将所有训练样本(支撑集和查询集)映射到同一空间中,从而挖掘不同样本提取的特征向量间的相似性和差异性,本文采用硬参数共享的方案,即主、辅助支路共享同一个特征提取模块的结构与参数。为了公平地与已有方案进行对比,本文沿用了与MVCNN相似的网络结构来构造特征提取模块F。其中,骨干网络结构由预训练好的经典网络模型构成,再通过最大池化操作获得融合特征。特征提取模块的具体结构如图2所示。

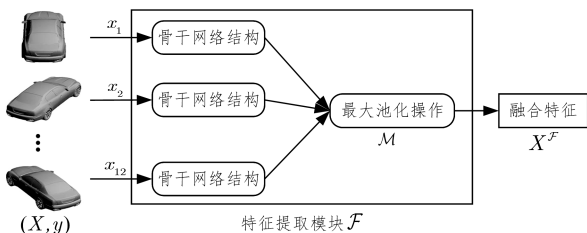


图2 特征提取模块流程示意图

Fig. 2 Architecture of feature extractor module

每一个三维形状X都由一系列不同视角的二维图像 $\{x_1, x_2, \dots, x_{12}\}$ 构成,y代表其真实类别标签。三维形状数据经过特征提取模块得到每个二维图像对应的特征向量,并通过最大池化操作 $\mathcal{M}(\cdot)$ 进行融合,最终得到融合特征 X^F 。

$$\begin{aligned} X^F &= \mathcal{M}(\mathcal{F}(X)) \\ &= \mathcal{M}((W_F)^T x_1 + b_F, (W_F)^T x_2 + b_F, \dots, (W_F)^T x_{12} + b_F) \end{aligned} \quad (2)$$

其中, x_r ($r \in \{1, 2, \dots, 12\}$) 表示其中一个视角的二维图像; $\{W_F, b_F\}$ 表示特征提取模块F的网络模型参数与偏置值;最大池化操作 $\mathcal{M}(\cdot)$ 对每一个二维图像得到的原始特征进行比较,在每一个元素位置上选择最大的值,从而得到最后的融合特征。将支撑集中三维形状和查询集中三维形状的特征分别记为 X_i^F 与 X_j^F 。

3.3 多支路多任务学习

在此任务中,由于训练集中每个类含有的三维形状极其有限,仅仅通过分类层进行预测从而更新网络参数并不能充分挖掘到不同样本间的关系。为了提高特征提取模块的表达能,使得输出的相同类别的不同样本的特征向量距离较小,而不同类别的样本特征距离较大,我们引入基于多支路的多任务学习来协同训练特征提取模块。

本文提出的网络结构包括主支路与辅助支路,两条支路共享特征提取模块F,详情请见第3.2节。主支路的任务在于提高预测精度;辅助支路用于得到待查询的三维形状与支撑集中所有三维形状的相似度得分,其任务在于挖掘类内、类间的信息。主、辅助任务由不同的损失函数加权联合指导网络训练,最终能够提高特征提取模块的表达能。

3.3.1 主支路学习

主支路由共享的特征提取模块与主分类模块组成。主分类模块由多层全连接层(Fully Connected Layer, FC)构成。将所有的支撑集样本送入主分类模块,得到分类输出向量 x_i^G ,其维度为数据集类别个数:

$$x_i^G = G(X_i^F) = (W_G)^T X_i^F + b_G \quad (3)$$

其中, $\{W_G, b_G\}$ 表示特征提取模块G的网络模型参数与偏置值, x_i^G 最大值对应的索引值即为预测分类结果 \tilde{y}_i 。分类输出向量 x_i^G 与真实类别标签 y。主支路的损失函数的计算公式如下:

$$\mathcal{L}_p = \frac{1}{m} \sum_{i \in S} CE(x_i^G, y_i)$$

$$= \frac{1}{m} \sum_{i \in S} (-\mathbf{x}_i^g[y_i] + \log(\sum_j \exp(\mathbf{x}_i^g[y_j]))) \quad (4)$$

3.3.2 辅助支路学习

辅助支路由共享的特征提取模块与辅助分类模块组成。已有研究已证实,当辅助支路任务与主支路任务相关时,多支路学习对原始任务的学习有帮助作用。辅助分类模块的结构如图3所示。

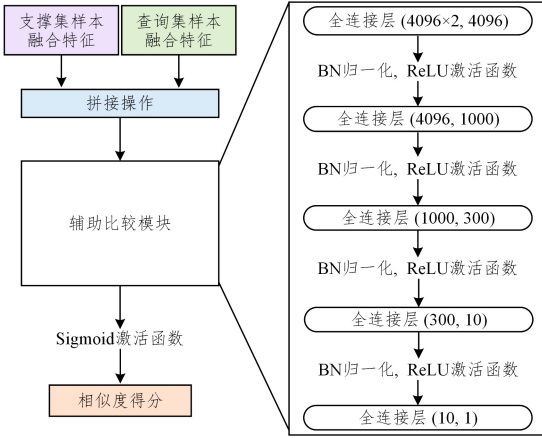


图3 辅助比较模块流程示意图

Fig. 3 Architecture of auxiliary comparison module

首先,我们定义拼接操作 $C(\cdot, \cdot)$,将支撑集与查询集的特征向量进行拼接,其操作的具体示例如图1所示。在小批量训练集中,每一个查询集样本都与支撑集中的每一个类的特征向量进行拼接,得到 C 个组合后的特征向量。然后,通过基于度量学习的辅助分类模块得到每一个组合后的特征向量的相似度值 $h_{i,j}$:

$$h_{i,j} = \mathcal{H}(C(\mathbf{X}_i^s, \mathbf{X}_j^q)) = (\mathbf{W}_\mathcal{H})^T \mathcal{C}(\mathbf{X}_i^s, \mathbf{X}_j^q) + \mathbf{b}_\mathcal{H} \quad (5)$$

其中, $\{\mathbf{W}_\mathcal{H}, \mathbf{b}_\mathcal{H}\}$ 表示特征提取模块 \mathcal{H} 的网络模型参数与偏置值。利用“Sigmoid”激活函数将 $h_{i,j}$ 归一至 $0 \sim 1$,得到最终的相似度 $\bar{h}_{i,j}$ 。

$$\bar{h}_{i,j} = \frac{1}{1 + e^{-h_{i,j}}} \quad (6)$$

其中, $\bar{h}_{i,j}$ 表示类内和类间的不同样本之间的相似程度。辅助分类模块的训练目标在于,当查询集样本与支撑集样本为同一个类别时,相似度得分 $\bar{h}_{i,j}$ 趋近于 1;当此三维形状数据不为同一个类别时,相似的得分趋近于 0。因此,我们使用均方误差 (Mean Square Error, MSE) 损失函数来指导辅助支路进行训练,在反向过程中对辅助分类模块和特征提取模块进行更新。辅助支路的损失函数的计算式如下:

$$\mathcal{L}_a = \arg \min_{\mathcal{F}, \mathcal{H}} \sum_{i=1}^m \sum_{j=1}^m \| \bar{h}_{i,j} - 1(y_i = y_j) \|_2^2 \quad (7)$$

3.3.3 多任务学习

第 3.3.1 节与第 3.3.2 节详细描述了主支路与辅助支路的训练方法和损失函数。辅助比较模块的引入给特征提取模块提供了更强的指导约束,使得输出的特征向量类内样本差异更小而类间样本的特征向量差异更大。为了使主支路、辅助支路共同指导特征提取模块进行学习,将主支路、辅助支路的学习过程转化为多任务学习。总损失函数通过加权联立主支路、辅助支路的损失函数,来共同指导模型训练更新。

$$\mathcal{L} = \mathcal{L}_p + \lambda \cdot \mathcal{L}_a \quad (8)$$

其中, \mathcal{L}_p 为主支路损失, \mathcal{L}_a 为辅助支路损失, λ 为辅助支路损失函数的权重。

主支路的损失函数用于更新特征提取模块与主分类模块的网络参数;辅助支路的损失函数用于共同更新特征提取模块与辅助分类模块的网络参数。

4 实验及结果分析

4.1 实验数据库与评测标准

4.1.1 数据集

为了证明本文方法的有效性,在公开的普林斯顿 ModelNet 数据集^[1]上对其进行了广泛的实验。CAD ModelNet 数据集是一种用于三维形状识别和跨模态检索的大型三维 CAD 模型数据集,包含 662 个类别,总计 127 915 个三维合成 CAD 模型。本文使用了两个公共的三维形状基准数据集,即 ModelNet10 和 ModelNet40,它们是 ModelNet 数据集的两个子集。

ModelNet10 由 10 个类别总计 4 899 个模型组成,其中训练集包含 3 991 个三维形状,测试集包含 908 个三维形状。ModelNet40 由 40 个类别总计 12 311 个模型组成,其中包括 ModelNet10 中的类。ModelNet40 的训练集包含 9 843 个三维形状,而测试集包含 2 468 个三维形状。

为了评估所提方法的有效性,即其在有限三维形状数据情况下的表现,除了使用原始的数据集进行训练与测试外,我们额外从 ModelNet10 和 ModelNet40 的训练集的每一个类别中随机选择部分样本作为新的训练集。为了比较的公平性,使用完整的原始测试集计算识别精度。采用第 3.1 节中描述的改进后的小批量训练样本输入策略,其输入的三维形状数据示例如图 4 所示。

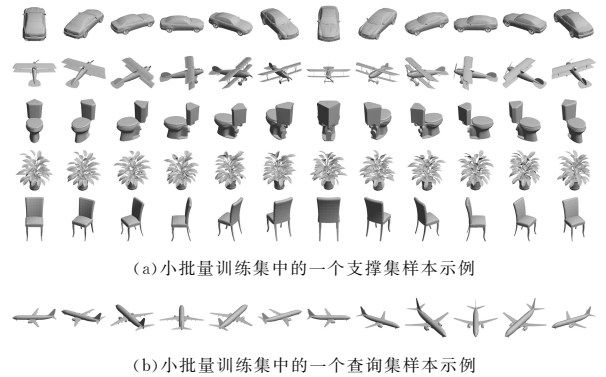


图4 小批量训练集输入数据示例

Fig. 4 Some samples of 3D shapes in a mini-batch training set

4.1.2 评价指标

我们将识别精度(单位为%)作为评价指标,在基准数据集上对所有实验方案进行评估。

4.2 实验配置

本文方法基于开源的 Pytorch 0.4.1 工具箱,使用 python 3.6 平台实现。在一台配备一块英伟达 1080Ti GPU (12G 内存)和一块 E5-1650 v4 CPU 的工作站上对提出的深度模型进行训练和测试。

4.3 实验细节

(1)数据处理。基于多视角的 3D 形状由一组 2D 图像表

示。本实验中,每 30 度放置一个“虚拟相机”,设置二维视图的数量为 12。此外,每一张二维图像的大小被统一缩放到 $224 \times 224 \times 3$ 像素,并对图片做归一化处理,即对 RGB 的每一个通道减去 ImageNet 的平均值^[2]。

(2)网络结构。对于特征提取模块 \mathcal{F} ,我们采用了预训练好的“fc7”层(含)之前的 AlexNet^[16]与 VGG11^[17]网络结构与初始化权值。为了公平比较,我们遵循 MVCNN 中的最大池化操作来处理不同视图的特性向量,然后再将融合后的特征输入主分类模块 \mathcal{G} 与辅助比较模块 \mathcal{H} 。主分类模块 \mathcal{G} 由一个具有随机初始化权值的完全连接层(即 4096-1000-10/40)组成,辅助比较模块 \mathcal{H} 的结构如图 3 所示。主分类模块和辅助比较模块中的权重使用“msra”方法初始化^[18]。所有层都采用“ReLU”激活函数和批处理标准化(Batch Normalization),在训练阶段使用标准 SGD 优化器更新网络模型参数。辅助比较模块的最后一层,采用“Sigmoid”激活函数将得到的相似度得分归一化到 0~1。

(3)参数设置。设置最大数据库遍历次数为 200,每一个小批量训练样本的类别数 $C=5$,每个类别的支撑集样本数 $S=1$,查询集样本数 $Q=1$ 。本文将总损失函数公式中的权重超参数设置为 $\lambda=10$ 。特征提取模块的初始学习率设为 1×10^{-4} ,主分类模块与辅助比较模块的初始学习率为 1×10^{-3} 。学习率衰减规律满足每 50 个迭代衰减为原始的 1/10。优化器使用朴素 SGD,动量设置为 0.9。

4.4 实验结果

4.4.1 实验方案概述

本文方法基于多任务学习,能在有限的三维形状数据下提高网络的识别能力。传统训练方法也使用相同的有限训练集与原始数据集,通过相同的参数设置来训练网络并将其作为基线。

baseline1:在每次迭代中,将小批量训练集中的所有样本直接送入主支路,得到的最终分类特征向量仅用于预测分类结果,并使用交叉熵损失函数进行网络训练,而不使用辅助支路。

baseline2:在每次迭代中,将小批量训练集中的所有样本直接送入主支路,得到的最终分类特征向量使用交叉熵损失函数和三元组损失函数(Triplet Margin Loss)进行网络训练,而不使用辅助支路。三元组损失函数如式(1)所示。

本文方案:在每次迭代中,根据改进的输入策略将小批量训练集中的所有样本引入网络模型,使用主支路和辅助支路的不同损失函数进行联合学习。

4.4.2 实验结果分析

为了进一步评估所提方法的有效性,在不同训练集样本的规模大小下对其进行了测试,即每个类别从原始数据到随机选择其中的 1/2, 1/4, 1/8, ..., 1/16, 依次减小。在两个基准数据库 ModelNet10 与 ModelNet40 上进行测试的结果如表 1 和表 2 所列。从表中可以看出,在不同样本数下,无论是使用 AlexNet 还是 VGG11 作为骨干结构网络,本文方法得到的精度相比基准水平平均有提高。在样本非常少,尤其是在 1/16 与 1/8 的情况下,多支路多任务的学习方法较 baseline1 的精确度有较为明显的提高,同时也优于引入三元组损失函数而对网络不做修改的 baseline2。

表 1 ModelNet10 测试集的识别精度定量比较

Table 1 Quantitative comparison on testing set of ModelNet10 in terms of recognition accuracy

实验方法	骨干网络结构	ModelNet10 训练集大小				
		1/16	1/8	1/4	1/2	1
		(单位:%)				
baseline1	AlexNet	85.13	89.54	90.42	91.30	93.06
	VGG11	86.67	90.97	91.41	92.84	93.61
baseline2	AlexNet	85.90	90.31	91.08	91.74	93.28
	VGG11	88.33	91.85	91.52	92.84	94.05
本文方案	AlexNet	87.56	90.75	91.41	92.07	93.50
	VGG11	89.20	92.51	92.84	93.28	93.94

表 2 ModelNet40 测试集的识别精度定量比较

Table 2 Quantitative comparison on testing set of ModelNet40 in terms of recognition accuracy

实验方法	骨干网络结构	ModelNet10 训练集大小				
		1/16	1/8	1/4	1/2	1
		(单位:%)				
baseline1	AlexNet	61.10	70.91	79.74	81.60	85.37
	VGG11	66.13	74.59	82.50	84.12	87.70
baseline2	AlexNet	64.14	72.00	81.48	82.74	85.74
	VGG11	69.37	75.12	82.25	85.01	88.37
本文方案	AlexNet	66.61	73.38	82.13	83.39	85.62
	VGG11	70.87	75.77	82.74	87.24	88.25

表 1 和表 2 证明了本文方法在样本数极为有限的情况下的有效性,该方案仅仅增加了一个简单的辅助比较模块来构建辅助支路,而不以额外增加网络深度为代价;并且证明了当样本极少时,使用不同损失项对不同支路的学习任务进行联合更新,比直接使用联合损失函数训练同一个网络的效果更佳。

4.5 消融实验

为了探究实验方案中超参数的作用,我们进行了更多的对比实验来评估其对最终效果的影响。其中,最为重要的参数是总损失函数(即式(7))中的平衡多项损失的权重 λ 。

如第 3.3 节所述,多任务学习的总损失包含两个部分的损失函数:主分类损失和辅助比较损失。这两个损失项代表了不同但相关的训练任务。超参数 λ 用于平衡两个损失项, $\lambda=0$ 表示只使用主分类任务,辅助任务不帮助网络训练,等同于 baseline1 的训练方案。

通过改变超参数 λ 的值(如 0, 0.1, 1, 10, 100),在 ModelNet10 数据集上进行了实验,以评估这两项的不同贡献对识别精度的影响。由于在极其有限的训练样本上的测试结果更具有代表性,图 5 给出了当将选取 1/16 的 ModelNet10 作为训练集训练 AlexNet 后得到的结果作为骨干网络结构的网络模型时, λ 取不同值时对测试精度的影响。

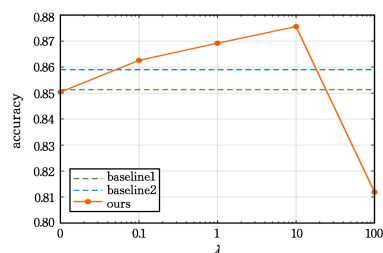


图 5 λ 取不同取值时在 ModelNet10 的 1/16 训练集上的测试集精度
Fig. 5 Recognition testing accuracy of proposed method versus λ with 1/16 of ModelNet10 training samples

从图5中可以看出,超参数 λ 的取值对最终的识别精度有一定影响。当 $\lambda=0$ 时,辅助支路不影响训练过程,等同于baseline1;当 $\lambda=10$ 时,取得最高的精度。

结束语 近年来,如何降低人工收集数据的代价,使用更少的样本获得更高的精度,已成为一个非常重要的研究课题。本文针对三维领域数据缺乏的问题,提出了新颖的训练方案。与传统的训练方法相比,所提方案获得了更好的识别性能,尤其在三维形状数据极其有限的情况下显著提升了精度。后续工作拟将此方案应用于基准效果更佳的三维形状识别模型,并解决计算机视觉下其他样本有限的识别问题。

参 考 文 献

- [1] WU Z, SONG S, KHOSLA A, et al. 3D shapenets: A deep representation for volumetric shapes[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1912-1920.
- [2] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009: 248-255.
- [3] YU T, YAN J, WANG Y, et al. Generalizing graph matching beyond quadratic assignment model[C]// Advances in Neural Information Processing Systems. 2018: 853-863.
- [4] YANG Y, FENG C, SHEN Y, et al. Foldingnet: Point cloud auto-encoder via deep grid deformation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 206-215.
- [5] SHEN Y, FENG C, YANG Y, et al. Mining point cloud local structures by kernel correlation and graph pooling[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4548-4557.
- [6] QI C R, SU H, NIEßNER M, et al. Volumetric and multi-view cnns for object classification on 3D data[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 5648-5656.
- [7] JOHNS E, LEUTENEGGER S, DAVISON A J. Pairwise decomposition of image sequences for active multi-view recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3813-3822.
- [8] HAN Z, SHANG M, LIU Y S, et al. View inter-prediction gan: Unsupervised representation learning for 3D shapes by learning global shape memories to support local view predictions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 8376-8384.
- [9] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 945-953.
- [10] 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.
- [11] WANG C, ZHANG X, LAN X. How to train triplet networks with 100k identities? [C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 1907-1915.
- [12] HE X, ZHOU Y, ZHOU Z, et al. Triplet-center loss for multi-view 3D object retrieval[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1945-1954.
- [13] LONG M, WANG J. Learning multiple tasks with deep relationship networks[J]. arXiv:1506.02117, 2015.
- [14] BINGEL J, SØGAARD A. Identifying beneficial task relations for multi-task learning in deep neural networks[J]. arXiv:1702.08303, 2017.
- [15] LU Y, KUMAR A, ZHAI S, et al. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5334-5343.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.
- [18] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]// Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.



ZHOU Zi-qin, born in 1996, postgraduate. Her main research interests include machine learning, deep learning, computer vision, 3D shape analysis, active learning and neural architecture search.



YAN Hua, born in 1971, professor. His main research interests include pattern recognition and intelligent system.