

# 基于自监督任务最优选择的无监督域自适应方法

吴 兰 王 涵 李斌全

河南工业大学电气工程学院 郑州 450001

**摘 要** 无监督域自适应方法通过源域标签数据学习到的知识对目标域无标签数据进行分类,成为目前迁移学习中解决两个域特征对齐的主流方法。针对现实中存在已标签数据量少且质量不高造成提取的特征不完备的情况,文中提出了基于自监督任务最优选择的无监督域自适应方法。为使特征具有更强的语义信息,在两个域未标记数据上使用了多个自监督任务;此外,针对进行自监督任务时的易混淆特征,提出了一种新的智能组合优化策略自适应地选择有效特征;最后通过两个域沿着任务相关方向靠近使得源域标记数据训练的分类器能够更好地推广到目标域。仿真实验在公开的6个基准数据集上分别从分类精度、训练集数据使用量、自监督任务使用效果3个方面进行了对比分析。实验结果表明,所提方法在3个方面的表现都优于现有的先进方法,使用相同数据集时分类精度提高8%;在相同的分类精度要求下,所用数据量减少12%;与单个自监督任务对比时精度提高了11%。

**关键词**: 迁移学习; 自监督任务; 无监督域自适应; 语义信息; 智能组合优化策略

中图分类号 TP391.41

## Unsupervised Domain Adaptive Method Based on Optimal Selection of Self-supervised Tasks

WU Lan, WANG Han and LI Bin-quan

School of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China

**Abstract** The unsupervised domain adaptation method uses the knowledge learned from the source domain label data to classify the target domain unlabeled data, which has become the mainstream method to solve the feature alignment of the two domains in transfer learning. In view of the fact that the amount of labeled data is small and the quality is not high, the extracted features are incomplete, this paper proposes an unsupervised domain adaptation method based on the optimal selection of self-supervised tasks. In order to make the features have stronger semantic information, multiple self-supervised tasks are used on the unlabeled data in the two domains. In addition, a new intelligent combination optimization strategy is proposed to adaptively select effective features for self-supervised tasks. Finally, the two domains are approached along the task-related direction so that the classifier trained on the source domain label data can be better promoted to the target domain. The simulation experiment conducts a comparative analysis on the six public benchmark datasets from three aspects: classification accuracy, training data volume, and self-supervised task use effect. Experimental results show that the proposed method outperforms the existing advanced methods in three aspects. The classification accuracy is improved by 8% when using the same datasets, and the amount of data used is reduced by 12% under the same classification accuracy requirements. Compared with a single self-supervised task, the accuracy is improved by 11%.

**Keywords** Transfer learning, Self-supervised task, Unsupervised domain adaptation, Semantic information, Intelligent combination optimization strategy

## 1 引言

域自适应能够有效地解决域间样本概率分布不一致的问题,已在计算机视觉领域取得了广泛的关注。域自适应具有重要的研究意义,它利用大型数据集的优势训练出具有良好泛化性能的分类器,提高了模型在不同的场景下的适应力。因此在域自适应中,通过减小不同数据集间的分布差异,模型能够更好地推广到新的数据集和任务中。

在基于域自适应的目标分类中,包含半监督域自适应和

无监督域自适应。在半监督域自适应的场景中,关键是利用源域和目标域中的标签数据中含有的信息构建两个域之间的关系。Liu等<sup>[1]</sup>采用由粗到细的加权机制来构建两个域之间的关系;Saito等<sup>[2]</sup>通过交替地对未标记目标数据相对于分类器的条件熵和其相对于特征编码器的条件熵进行最大最小化,以充分利用两个域的标签信息;Tan等<sup>[3]</sup>提出利用两个域的标签信息进行协作分布对齐,以对偶映射进行知识迁移。但在现实场景中,目标域含有标签的场景并不多见。

在无监督域自适应中,通过学习两个域的域不变特征,将

基金项目:国家自然科学基金(61973103);河南省中原青年拔尖人才计划(19A120002)

This work was supported by the National Natural Science Foundation of China(61973103) and Province Central Plains Youth Top Talents Plan Henan University(19A120002).

通信作者:吴兰(wulan@haut.edu.cn)

含有标签信息的源域分类器推广到没有标签信息的目标域。对抗性自适应通过两个网络相互对抗来减少域间差异;Bousmalis等<sup>[4]</sup>基于生成对抗网络的模型使源域图像看起来像是从目标域中绘制的一样,将像素空间从一个域迁移到另一个域;Tzeng等<sup>[5]</sup>提出了一个不附带条件权重共享和GAN损失的对抗域自适应框架;Hoffman等<sup>[6]</sup>提出了一种利用生成图像空间对齐和潜在表示空间对齐的域间对抗自适应模型;Satio等<sup>[7]</sup>提出了对抗 dropout 正则化,鼓励生成器能够生成更好的判别特征;Shu等<sup>[8]</sup>提出了两种相关模型,虚拟对抗域适应模型 VADA 和以 VADA 模型为初始化的 DIRT-T 模型;Bousmalis等<sup>[9]</sup>基于对抗思想将特征提取划分为两个子空间的图像表示,一个对每个域私有,一个跨域共享。部分研究期望通过学习目标域有用的信息,能够保证分类器在目标域中的准确性。Sun等<sup>[10]</sup>通过学习在两个域上执行辅助的自监督任务进行对齐;Ghifary等<sup>[11]</sup>通过两个任务,即对带有标签信息的源域数据进行分类和对无标记的目标域数据进行重建,来学习到它们的共享编码表示;Satio等<sup>[12]</sup>提出了一种无监督域自适应的非对称训练,将伪标签分配给未标记的样本,通过不对称,一个网络被用来标记无标签的目标样本,一个网络进行样本训练。Haeusser等<sup>[13]</sup>通过在嵌入空间中源和目标数据之间的关联有效增强,从而导出目标域的分类标签,同时最小化标记源域上的分类错误。虽然上述方法取得了不错的成绩,但是它们没有进一步进行特征选择提取更有效的特征。

基于此,本文使用了3个自监督任务,并设置了智能组合优化策略去提取可鉴别特征。主要贡献如下:

(1)为了解决无监督域自适应场景中已标签数据量少且质量不高带来的特征提取不完备的问题,本文提出了一种新的框架。该框架包含两部分,一部分在源域标签数据上进行分类器训练,一部分在源域和目标域中未标记部分使用多个辅助自监督任务,并通过智能组合优化策略提取有效特征来进行两个域对齐工作,最后将源域分类器推广到目标域进行分类。

(2)为了得到更高级的语义特征,使其具有更强的分辨能力,在两个域未标记数据上使用了图像水平翻转、图像旋转、图像位置预测3个自监督任务进行表征学习。为了进一步优化自监督任务的易混淆特征,将各自监督任务进行随机组合,并且基于自监督任务分配伪标签的不确定性,设定阈值,用来选择最优的任务组合得到更完备的特征。

(3)本文在6种公开的基准数据集上分别从分类精度、训练集数据使用量、自监督任务使用效果3个方面进行了对比分析。在使用相同数据集下,与其他8种先进方法相比,本文方法的精度提高8%;在分类精度相同的情况下,与其他两种先进方法相比本文方法使用的数据集减少12%;与不同自监督任务效果相比,本文方法的精度提高11%。

## 2 相关工作

### 2.1 无监督域自适应

在无监督域自适应中,只有源域有标签,而目标域没有标签信息<sup>[4-23]</sup>。其思想大致可以归为基于样本和基于特征两类。基于样本的无监督域自适应大致可以分为两类。一类是以加权的方式筛选出源域中满足目标域概率分布的样本训练分类器,重点是求解加权系数。Zadrozny等<sup>[14]</sup>通过计算源领域与目标领域样本密度函数之比去获得权重。另一类是通过

刻画源域到目标域的迁移过程,使得源样本渐进地投影到目标域中。Caseiror等<sup>[15]</sup>提出对测地线在Spline流形上实施子空间的采样。在基于特征的无监督域自适应中,把两个域的样本共同映射到隐藏的特征空间,在这个空间中,两个域的特征概率分布差异减小的解决方案可以分为两种:1)基于特征空间的角度解决问题,Huberttyh等<sup>[16]</sup>利用跨域样本选择过程,构建域不变特征子空间;2)学习源域和目标域映射到概率空间的变换关系进行域自适应,Dai等<sup>[17]</sup>提出通过利用多个视角的数据构建成的翻译器去实现源域与目标域特征空间的映射。

### 2.2 自监督任务

自监督学习设置辅助任务从大规模的无监督数据中挖掘自身的监督信息对网络进行训练,从而学习到更多的语义特征。给灰度图像上色<sup>[24-26]</sup>、将图像分块并根据分块后的中心一块来预测位置<sup>[27-28]</sup>、对破坏的图像进行分类<sup>[29]</sup>、图像旋转<sup>[30-31]</sup>、图像修复<sup>[31]</sup>等大量辅助任务被提出并不断应用于机器视觉的数据处理,但是大多数方法仍没有将这些辅助自监督任务与无监督域自适应有效结合起来去解决视觉分类问题。我们的方法不仅将二者结合起来,并且设置了智能组合优化策略,充分利用自监督对数据的预处理去实现域自适应。

在训练过程当中,数据需要进行多个自监督任务,为使模型能够保持一定的训练速度和准确率,并且通过自监督任务所得到的不应该是域对齐无意义的特征,它应该能够对下一步进行域对齐起到一定的促进作用。图像的像素着色或者修复任务取决于多个因素,例如光照强度、四周景色,所提取到的特征并不是关于更高水平的视觉分类特征。因此预测结构标签的分类任务比预测像素或修补的重建任务更适用于分类工作的进行。

为了保持模型的高效性和准确率,选择了具体简单性和高性能的3个自监督任务。

(1)旋转预测:输入图像进行 $0^\circ, 90^\circ, 180^\circ$ 和 $270^\circ$ 的旋转。任务是预测图像的旋转角度。

(2)翻转预测:输入图像随机水平翻转。任务是预测图像是否翻转。

(3)位置预测:将图片裁剪到4个象限,每部分都赋予一个标签,从输入图像中随机裁剪出一块补丁。任务是预测补丁位置。

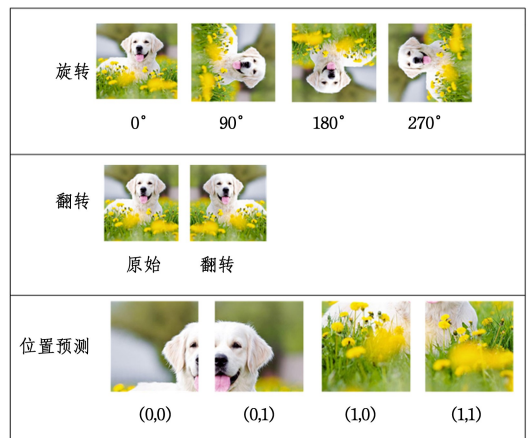


图1 3个自监督任务示例图

Fig. 1 Three examples of self-supervised tasks

### 3 最优选择的无监督域自适应方法

为了处理无监督域自适应的问题,期望利用源域中标签数据学到的知识对目标域中无标签数据进行分类。将源域中标签数据定义  $D_s = \{(x_i^s, y_i^s), i=1, 2, 3, \dots, n\}$ , 给定源域无标签数据为  $\bar{D}_s = \{\bar{x}_i^s, i=1, 2, 3, \dots, n\}$  和目标域无标签数据为  $D_t = \{x_i^t, i=1, 2, 3, \dots, m\}$ , 并且将其两个域中所有未标记数据定义为  $D_u = \bar{D}_s \cup D_t$ 。

#### 3.1 网络框架

本节将详细介绍所提出的解决自适应问题的网络框架。框架可分为两个部分,即主监督任务分类器和 3 个辅助自监督任务

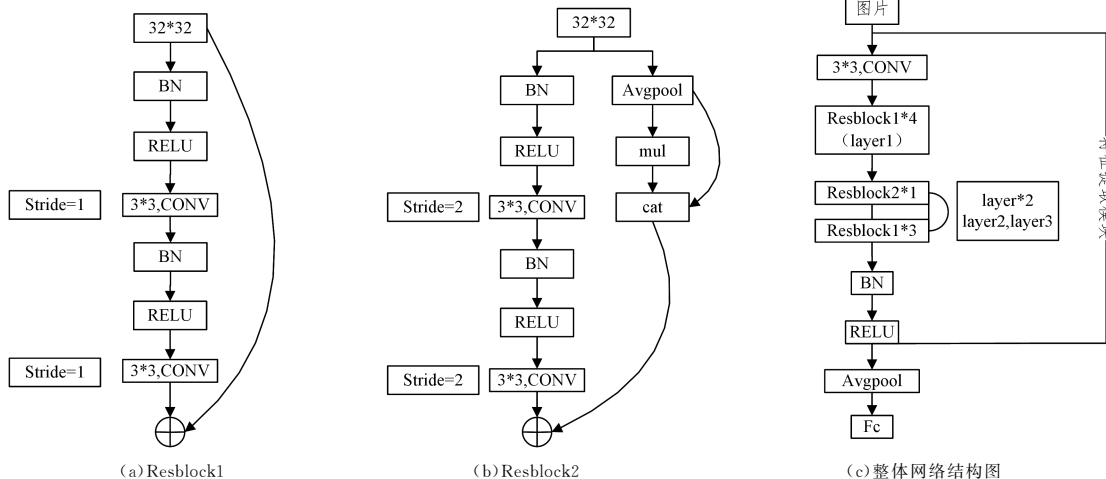


图 3 网络结构图

Fig. 3 Network structure diagram

在主监督任务分类器中,将从公共特征提取器  $F$  得到的关于标签信息的源域数据送入到特定的线性层  $\Phi_0$ , 进行分类训练,即源域标签数据分类损失  $L_0$  可表示为:

$$\ell_{cls}(D_s; F(x_i^s), \Phi_0) = \sum_{(x_i^s, y_i^s) \in D_s} L_0(\Phi_0(F(x_i^s)), y_i^s) \quad (1)$$

在辅助自监督任务训练阶段,首先将特征提取器  $F$  中得到的两个域未标记数据  $D_u$  中所有特征送入到 3 个辅助自监督任务的线性层  $\Phi_k (k=1, 2, 3)$  进行数据转换  $t_k (k=1, 2, 3)$ , 并将其视为分类问题。由此可以得到每个自监督任务分配伪标签的预测概率,然后通过智能组合优化策略得到有效特征,将源域和目标域沿着该组合策略相关方向对齐,从而完成最终分类任务。

#### 3.2 智能组合优化策略

为了提取到更完备的特征,本文首先使用了 3 个辅助自监督任务,每个自监督任务都有它们各自的线性层。不同的自监督任务,输出维度各有不同。对于图像水平翻转任务,视其输出层为二维度;对于图像旋转和位置预测任务,视其输出层为四维度。每个自监督辅助任务中经过数据转换  $t_k$  得到的带有伪标签的源域数据表示为:

$$T_k(s) = \{(t_k(\bar{x}_i^s), y_i), i=1, 2, 3, \dots, n\} \quad (2)$$

得到的带有伪标签的目标域数据表示为:

$$T_k(t) = \{(t_k(x_i^t), y_i), i=1, 2, 3, \dots, m\} \quad (3)$$

其中,  $t_k$  是每个自监督任务的数据转换过程,即指翻转、旋转、位置预测等 3 个自监督任务。  $y_i$  是两个域无标签数据进行自监督任务以后得到的伪标签信息。关于源域和目标域的无

选择机制。其总体框架图如图 2 所示,网络结构如图 3 所示。

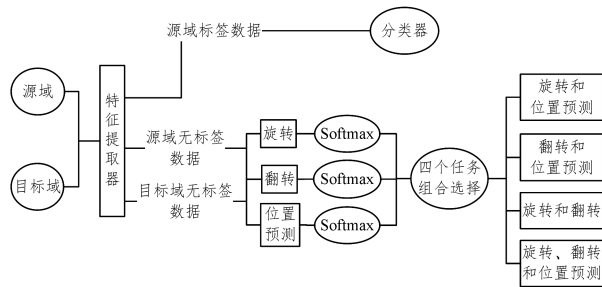


图 2 模型总体框架图

Fig. 2 Overall framework of model

标签数据的伪标签分类损失,即  $L_{T_k^s}$  和  $L_{T_k^t}$ , 两个损失函数分别可以表示为:

$$L_{T_k^s} = \sum_{(x_i^s, y_i) \in T_k(s)} L_k(\Phi_k(F(t_k(\bar{x}_i^s))), y_i) \quad (4)$$

$$L_{T_k^t} = \sum_{(x_i^t, y_i) \in T_k(t)} L_k(\Phi_k(F(t_k(x_i^t))), y_i) \quad (5)$$

每个辅助监督任务中的损失函数即为上述两式之和,可表示为:

$$\ell_k = L_{T_k^s} + L_{T_k^t} \quad (6)$$

此外,不同的数据集下,数据间在进行不同的自监督任务时提取的特征易混淆,可能会产生负面的迁移效果,如图 4 所示。

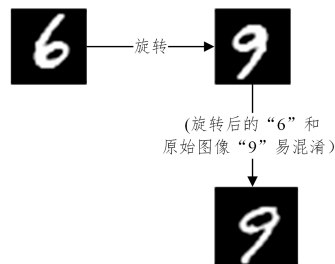


图 4 易混淆数据解析图

Fig. 4 Analysis chart of confusing data

为了避免以上数据间易混淆的情况造成负迁移的效果,本文提出了一种有效的优化策略。本文以上述 3 个辅助自监督任务为基底,通过对自监督任务进行随机组合,同时为了保证实验的便捷性,设置了 4 种任务组合自适应地选择有效特

征。其主要步骤如图 5 所示。

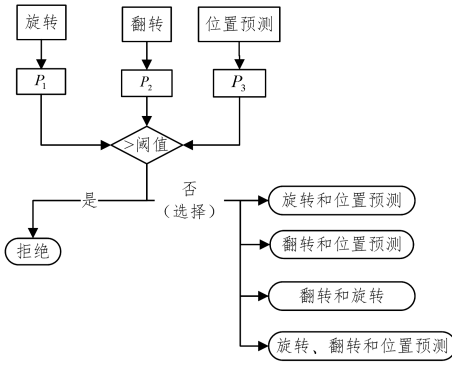


图 5 智能组合优化策略图示

Fig. 5 Diagram of intelligent combination optimization strategy

本文通过熵量化各项任务预测伪标签的不确定性,较小的熵意味着更有信心的预测,较高的熵意味着概率分布更稀疏,这表明该预测更具有不确定性。各任务预测伪标签的不确定性的公式可表示为:

$$H_k = - \sum_{x_i \in D_n} P(x_i) \log(P(x_i)), k=1, 2, 3 \quad (7)$$

其中,  $p(x_i)$  表示类别中第  $i$  个样本在辅助自监督任务后经过 Softmax 函数后得到的所分配伪标签的概率。

选择任务组合时,为了能够更确定地舍弃掉概率分布最稀疏的辅助自监督任务,通过阈值  $\rho$  在各个熵之间设置一个边界:

$$\rho = \frac{1}{3} \sum_{k=1}^n H_k, n=1, 2, 3 \dots \quad (8)$$

并根据各任务预测的不确定性与阈值  $\rho$  的数值比较完成最终的任务组合策略。为了保证熵与阈值之间能得到较准确的比较结果,各数值精确到小数点后两位:

$$H_k = \begin{cases} > \rho, & \text{reject the } k\text{th task} \\ \leq \rho, & \text{others} \end{cases} \quad (9)$$

由此,可得到最终经过优化策略后的各个任务组合损失函数  $\ell_K'$  ( $K=1, 2, 3, 4$ ):

$$\ell_1' = L_{T_1} + L_{T_1'} + L_{T_3} + L_{T_3'} \quad (10)$$

$$\ell_2' = L_{T_2} + L_{T_2'} + L_{T_3} + L_{T_3'} \quad (11)$$

$$\ell_3' = L_{T_1} + L_{T_1'} + L_{T_2} + L_{T_2'} \quad (12)$$

$$\ell_4' = L_{T_1} + L_{T_1'} + L_{T_2} + L_{T_2'} + L_{T_3} + L_{T_3'} \quad (13)$$

根据所选的任务组合,最终确定的任务组合损失函数为上面 4 个公式之一。

### 3.3 域自适应

为了防止模型的过拟合,学习 Sun 等<sup>[10]</sup>的思想计算了两个域验证集(未标记)数据的均值,将其和源域验证集上的主任务分类损失结合起来作为最终验证集的主要损失函数,当验证损失函数达到最小时完成早期停止。其中,均值可以表示为:

$$D(D_s', D_t'; F) = \left\| \frac{1}{n} \sum_{x_i' \in D_s'} F(x_i') - \frac{1}{m} \sum_{x_i' \in D_t'} F(x_i') \right\|_2 \quad (14)$$

其中,  $D_s'$  和  $D_t'$  是两个域上未标记部分的验证集。

为了完成两个域的对齐并进行分类任务,得到总的目标优化函数可表示为:

$$\ell_{at} = \min \ell_{ds}(D_s; F(x_i^s), \Phi_0) + \ell_K', K=1, 2, 3, 4 \quad (15)$$

其中,  $K$  指任务组合,目标优化函数只取其中一个任务组合,

即式(10)一式(13)之间任意一个。

## 4 仿真实验

### 4.1 数据集、实验设置和实验细节

数据集:为了验证本文模型的有效性和优越性,使用了 MNIST, USPS, SVHN, MNISTM 4 种手写数据集和 STL-10 和 CIFAR-10 两种风景数据集。其中, MNIST 是手写数字灰度图像,有数字 0-9 共 10 个类别,其中训练集为 60 000 张,测试集为 10 000 张。USPS 是手写数字灰度图像,与 MNIST 数据集有所不同,有数字 0-9 共 10 个类别,其中训练集为 7 291 张,测试集为 2 097 张。SVHN 是从谷歌街景中裁剪出的门牌号码集,为彩色图像,有数字 0-9 共 10 个类别,其中训练集为 73 257 张,测试集为 26 032 张。MNISTM 是由 MNIST 数据集和 BSD500 数据集随机色块混合而成,其中训练集为 60 000 张,测试集为 10 000 张。CIFAR-10 是 10 类中心自然场景对象,其中 50 000 个样本在训练集中,10 000 在测试集中。STL-10 是物体的彩色图像,与 CIFAR-10 略有不同,其中训练集中有 5 000 个样本,测试集中有 8 000 个样本。

实验设置:采用 Sun 等<sup>[10]</sup>的 26 层预激活 ResNet 作为模型,其中  $\Phi_0$  是作为预测的最后一个线性层,  $F$  是一个公共特征提取器。在训练时,每个辅助的自监督任务有一个简单的线性层  $\Phi_k$  ( $k=1, 2, 3$ ) 连接到特征提取器  $F$  的末端。本文使用随机梯度下降优化模型,权重衰减为  $5 \times 10^{-4}$ , 动量为 0.9, 批量大小为 128, 初始学习率为 0.1。使用两个阶段学习率,其中学习率在每个阶段下降至上一阶段学习中的 1/10。通过源域分类器  $\Phi_0$  和辅助自监督任务分类器  $\Phi_k$  ( $k=1, 2, 3$ ) 的收敛性判断各参数取值。对于优化,按照式(15)来完成。

实验细节:在 4 种手写数字数据集上,本文分别和 AD-DA<sup>[5]</sup>, VADA<sup>[8]</sup>, DIRT-T<sup>[8]</sup>, ATT<sup>[12]</sup>, DANN<sup>[22]</sup>, KNN-ad<sup>[23]</sup> 等主流方法进行了对比分析。在两种风景数据集中,分别和 VADA<sup>[8]</sup>, DIRT-T<sup>[8]</sup> 等方法进行了对比分析。以上方法均在引言或相关部分有相关介绍。ResNet 作为图像识别较为常用的网络结构,实验均是在此基础上进行实验。这样做一方面为了方便与其他方法进行比较;另一方面,该网络结构可以在图像特征上获得更好的提取结果。

### 4.2 实验结果和性能对比分析

#### 4.2.1 与不同现有方法对比分析

为了验证方法的有效性,实验分别在风景数据集中和手写数字数据集中与当前最先进的方法进行仿真实验比较,本文显然得到了更好的分类性能。手写数据集实验比较结果如表 1 所列,风景数据集实验比较结果如表 2 所列。在表 1 和表 2 中,在多个任务中文本方法优于其他现有的方法。例如,在 MNIST-MNISTM 中,本文方法的精度比 VADA<sup>[8]</sup> 高了 1.2%; 在 MNIST-USPS 中,精度比 ADDA<sup>[5]</sup> 提高了 8%; 在 MNIST-SVHN 这一组任务中,精度比 ATT<sup>[12]</sup> 提高了 13%; 在 USPS-MNIST 中,精度比 ADDA<sup>[5]</sup> 提高了 0.58%; 在 STL-CIFAR 中,精度比 VADA<sup>[8]</sup> 提高了 6.35%; 在 CIFAR-STL 中,精度比 VADA<sup>[8]</sup> 提高了 3.19%。其原因可能如下: 1) 本文设置了 3 个自监督任务,基于这些自监督任务的任务组合不仅能够增加最终提取的特征数量,也能够使得特征具有更强的语义信息; 2) 本文提出了智能组合优化策略,能够筛

选出具有鉴别性的更高质量的特征促进分类工作。

表 1 在手写数据集中与现有方法比较

Table 1 Comparison with existing methods in handwritten datasets

Source Target	MNIST MNISTM	MNIST USPS	MNIST SVHN	USPS MNIST	SVHN MNIST
ADDA <sup>[5]</sup>	—	89.4	—	90.1	76.0
VADA <sup>[8]</sup>	97.7	—	47.5	—	97.9
DIRT-T <sup>[8]</sup>	98.9	—	54.5	—	99.4
ATT <sup>[12]</sup>	94.2	—	52.8	—	86.2
DANN <sup>[22]</sup>	81.5	—	35.7	—	73.6
KNN-ad <sup>[23]</sup>	86.7	—	40.3	—	78.8
OURS	98.9	97.63	66.10	90.68	88.32

表 2 在风景数据集中与现有方法比较

Table 2 Comparison with existing methods in the landscape datasets

Source Target	STL-10 CIFAR-10	CIFAR-10 STL-10
VADA <sup>[8]</sup>	73.5	80.0
DIRT-T <sup>[8]</sup>	75.3	—
OURS	79.85	83.19

4.2.2 不同辅助自监督任务效果的对比分析

本文共使用了 3 种自监督辅助任务,在手写数字数据集上的效果如图 6 所示,在风景数据集中的效果如表 7 所列。其中,R 指只有旋转预测任务,L 指只有位置预测任务,F 指只有水平翻转任务。

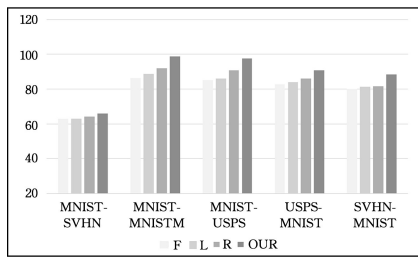


图 6 数字数据集中不同自监督任务的效果对比

Fig. 6 Comparison of effects of different self-supervised tasks in digital datasets

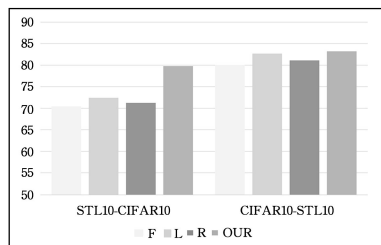


图 7 风景数据集中不同自监督任务的效果对比

Fig. 7 Comparison of the effects of different self-supervised tasks in the landscape datasets

从图 6、图 7 中可以看出,在不同数据集上,各个任务的效果不同。通过进行单个自监督任务与使用智能组合以后的精度对比,可以观察到使用智能组合以后的精度明显提高。例如,在 MNIST-MNISTM 这一组任务中,本文方法和其他单个自监督任务相比精度提高了 10%;在 MNIST-USPS 中本文方法和其他单个自监督任务相比精度提高了 11%;在 USPS-MNIST 中本文方法和其他单个自监督任务相比精度提高了 8%;在 SVHN-MNIST 中本文方法和其他单个自监督任务相比精度提高了 6%;在 STL-CIFAR 中本文方法和其他单个自监督任务相比精度提高了 7%。这样的结果证明这

些自监督任务能够很好地与域自适应结合起来,实现两个域之间的对齐,又证明了在优化易混淆特征方面,智能组合起到了很大的作用,优化后使得模型能够自适应选择得到更有效的特征,有利于精度的提高。

4.2.3 训练集使用量的对比分析

在实验当中,为了验证模型在训练数据量较少时的性能,将其分别与 ADDA<sup>[5]</sup>,ATT<sup>[12]</sup>两种方法进行了训练集使用量的对比。在与 ADDA<sup>[5]</sup>的对比实验中,将分类的准确率分别保持在 89.4%,90.1%,76.0%;在与 ATT<sup>[12]</sup>的对比实验当中,将分类的准确率分别保持在 94.2%,52.8%,86.2%。通过改变目标域训练数据不断地迭代训练来拟合其他先进方法的准确率,将准确率精度范围控制在±0.5。数据最终使用量的对比分别如图 8、图 9 所示。

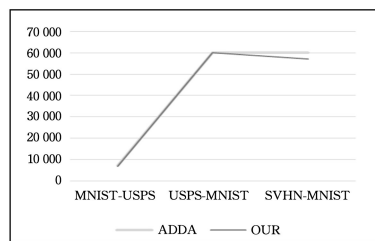


图 8 相同准确率下,与 ADDA 进行训练集使用量对比

Fig. 8 Comparison of training set usage with ADDA at the same accuracy rate

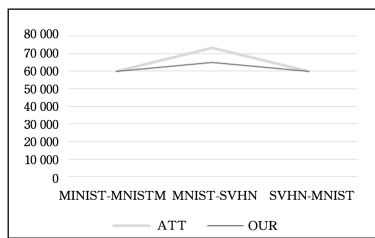


图 9 相同准确率下,与 ATT 进行训练集使用量对比

Fig. 9 Comparison of training set usage with ATT at the same accuracy rate

从图中可以观察到,在相同的准确率下,本文较于其他两种方法所使用的目标域训练集数据量明显较少。与不同方法比较时,若与本方法的分类精度相差较大,在实验中将有更多的数据删减。例如在与 ADDA<sup>[5]</sup>进行比较时,数据量删减了 8%;在与 ATT<sup>[12]</sup>方法进行比较时,由于准确率差距较大,所以目标域训练集数据删减了 12%。这是因为在自监督任务的 3 种组合选择下,每一种组合均至少包含一种数据增强任务,进行任务组合的预处理以后,不仅能够得到理想的特征,也能够增加一定的特征数量来提高分类任务的精度。

**结束语** 本文针对机器视觉中已标签数据量少且质量不好导致提取的特征不完备使分类性能不完善的问题,提出了基于自监督任务最优选择的无监督域自适应方法。该方法有效使用了 3 个自监督任务,且提出了一种新的智能组合优化策略对特征进行优化。首先在两个域无标签数据上进行辅助自监督任务;其次,利用智能组合优化策略进行有效的特征选择;最后通过任务组合将源域分类器推广到目标域进行分类任务。在多个数据集上进行仿真实验验证,从分类精度、训练集使用量和自监督任务使用效果 3 个方面对模型性能进行了分析证明了本文模型具有良好的性能,能够在智能优化策略

下得到更多的有效特征,促进机器视觉分类方面的任务。在未来的工作中,可以引进更多的自监督任务和无监督域自适应场景有效结合,同时也可以引进多个源数据和自监督任务相结合来处理无监督场景中的分类工作。

### 参 考 文 献

- [1] LIU H, CAO Z J, LONG M S, et al. Separate to Adapt: Open Set Domain Adaptation via Progressive Separation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019;2927-2936.
- [2] SAITO K, KIM D, SCLAROFF S, et al. Semi-Supervised Domain Adaptation via Minimax Entropy[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019;8050-8058.
- [3] TAN S H, JIAO J N, ZHENG W S. Weakly Supervised Open-set Domain Adaptation by Dual-domain Collaboration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019;5394-5403.
- [4] BOUSMALIS K, SILBERMAN N, DOHAN D, et al. Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; 3722-3731.
- [5] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial Discriminative Domain Adaptation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;7167-7176.
- [6] HOFFMAN J, TZENG E, PARK T, et al. CyCADA: Cycle-Consistent adversarial domain adaptation[J]. arXiv: 1711. 03213, 2017.
- [7] SAITO K, USHIKU Y, HARADA T, et al. Adversarial dropout regularization[J]. arXiv: 1711. 01575, 2017.
- [8] SHU R, BUI H H, NARUI H, et al. A DIRT-T approach to unsupervised domain adaptation[J]. arXiv:1802. 08735, 2018.
- [9] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, et al. Domain Separation Networks[C]// Advances in Neural Information Processing Systems. 2016.
- [10] SUN Y, TZENG E, DARRELL T, et al. Unsupervised domain adaptation through self-supervision[J]. arXiv: 1909. 11825, 2019.
- [11] GHIFARY M, KLEIJN W B, ZHANG M J, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[C]// European Conference on Computer Vision. 2016; 597-613.
- [12] SAITO K, USHIKU Y, HARADA T. Asymmetric tri-training for unsupervised domain adaptation[J]. arXiv: 1702. 08400, 2017.
- [13] HAEUSSER P, FRERIX T, MORDVINTSEV A, et al. Associative Domain Adaptation[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017; 2765-2773.
- [14] ZADROZNY B. Learning and Evaluating Classifiers Under Sample Selection Bias[C]// Proceedings of the International Conference on Machine Learning (ICML). Canada; ICML, 2004; 114-122.
- [15] CASEIRO R, HENRIQUES J F, MARTINS P, et al. Unsupervised Domain Adaptation by Sampling Subspaces Along the Spline Flow[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA; IEEE, 2015;3846-3854.
- [16] TSAI Y, YEH Y R, WANG F Y C. Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA; IEEE, 2016;5081-5090.
- [17] DAI W, CHEN Y, XUE G R, et al. Translated Learning: Transfer Learning Across Different Feature Spaces[C]// Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS). Canada; NIPS, 2008; 353-360.
- [18] KANG U L, JIANG L, YANG Y, et al. Contrastive Adaptation Network for Unsupervised Domain Adaptation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019;4893-4902.
- [19] PAN Y W, YAO T, LI Y H, et al. Transferrable Prototypical Networks for Unsupervised Domain Adaptation[J]. arXiv: 1904. 11227, 2019.
- [20] KANG G L, JIANG L, YANG Y, et al. Contrastive Adaptation Network for Unsupervised Domain Adaptation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019;4893-4902.
- [21] SENER O, SONG H O, SAXENA A, et al. Learning Transferrable Representations for Unsupervised Domain Adaptation. Domain Separation Networks[C]// Advances in Neural Information Processing Systems. 2016.
- [22] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-Adversarial Training of Neural Networks[C]// Journal of Machine Learning Research. 2016;1-35.
- [23] GANIN Y, LEMPISKY V. Unsupervised domain adaptation by backpropagation[C]// International Conference on Machine Learning. 2015;1180-1189.
- [24] ZHANG R, ISOLA P, EFROS A A. Colorful image colorization[J]. European Conference on Computer Vision, 2016, 9907: 649-666.
- [25] LARSSON G, MAIRE M, SHAKHAROVICH G. Learning representations for automatic colorization[J]. European Conference on Computer Vision, 2016, 9908: 577-593.
- [26] VONDRICK C, SHRIVASTAVA A, FATHI A, et al. Tracking Emerges by Colorizing Videos[C]// Proceedings of the European Conference on Computer Vision. 2018;391-408.
- [27] NOROOZI M, FAVARO P. Unsupervised learning of visual representations by solving jigsaw puzzles[J]. European Conference on Computer Vision, 2016, 9910: 69-84.
- [28] DOERSCH C, GUPTA A, EFROS A A. Unsupervised visual representation learning by context prediction[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015;1422-1430.
- [29] JENNI I, FAVARO P. Self-Supervised Feature Learning by

Learning to Spot Artifacts[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2733-2742.

- [30] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised Representation learning by predicting image rotations [J]. arXiv: 1803.07728, 2018.
- [31] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2536-2544.
- [32] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised repre-

sentation learning by predicting image rotations [J]. arXiv: 1803.07728, 2018.



**WU Lan**, born in 1981, Ph.D, professor, master tutor. Her main research interests include information security, artificial intelligence, multisensor networked information fusion theory, fault diagnosis of complex systems and devices, intelligent information processing.

(上接第 330 页)

CQT 特征和梅尔频谱特征, 然后采用 CycleGAN 对 CQT 特征和梅尔频谱特征映射的图片做风格迁移, 最后通过 WaveNet 解码器生成音乐波形, 最终实现了带有人声的音乐的风格迁移。在实验训练的分类器上, 符合要求的音乐的平均风格迁移率达到了 94.07%。本文提取以上两种特征, 既保留了音乐的特征, 也保留了较多的人声的特征。与此同时, 这两种特征的重叠部分共同约束了波形, 使得 WaveNet 对波形的预测更加准确。

在以后的工作里, 可以考虑如何将巨量的特征数据输入网络和如何保留前后音乐片段之间的关联性。除此之外, 如何提高风格转换后音乐的质量也是需要着重解决的问题。

## 参 考 文 献

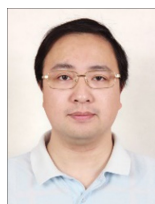
- [1] JING Y, YANG Y, FENG Z, et al. Neural style transfer: A review[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 26(11): 3365-3385.
- [2] DAI S, ZHANG Z, XIA G G. Music style transfer: A position paper[J]. arXiv: 1803.06841, 2018.
- [3] GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style[J]. arXiv: 1508.06576, 2015.
- [4] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 2223-2232.
- [5] YI Z, ZHANG H, TAN P, et al. Dualgan: Unsupervised dual learning for image-to-image translation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2849-2857.
- [6] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[J]. arXiv: 1703.05192, 2017.
- [7] BRUNNER G, WANG Y, WATTENHOFER R, et al. Symbolic music genre transfer with cyclegan[C]//2018 IEEE 30th International Conference on Tools with Artificial Intelligence (IC-TAI). IEEE, 2018: 786-793.
- [8] HUANG S, LI Q, ANIL C, et al. Timbretron: A wavenet (cy-

clegan (cqt (audio))) pipeline for musical timbre transfer[J]. arXiv: 1811.09620, 2018.

- [9] MOR N, WOLF L, POLYAK A, et al. A universal music translation network[J]. arXiv: 1805.07848, 2018.
- [10] GOSODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. 2014: 2672-2680.
- [11] OORD A, DIELEMAN S, ZEN H, et al. Wavenet: A generative model for raw audio[J]. arXiv: 1609.03499, 2016.
- [12] POLYAK A, WOLF L. Attention-based wavenet autoencoder for universal voice conversion[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 6800-6804.
- [13] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts[J]. Distill, 2016, 1(10): e3.
- [14] ENGEL J, RESNICK C, ROBERTS A, et al. Neural audio synthesis of musical notes with wavenet autoencoders[C]//International Conference on Machine Learning. PMLR, 2017: 1068-1077.
- [15] DEFFERRARD M, BENZI K, VANDEREGHEYNST P, et al. Fma: A dataset for music analysis[J]. arXiv: 1612.01840, 2016.
- [16] WU M, LIU X. A Double Weighted KNN Algorithm and Its Application in the Music Genre Classification[C]//2019 6th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2020: 335-340.



**YE Hong-liang**, born in 1999. His main research interests include deep learning and music processing.



**ZHU Wan-ning**, born in 1983, Ph.D. His main research interests include quantum information technology and quantum computing.