

# 用于迁移学习的多尺度领域对抗网络

林佳伟<sup>1</sup>, 王士同<sup>2</sup>

(1. 江南大学人工智能与计算机学院, 无锡 214122; 2. 江苏省物联网应用技术重点建设实验室, 无锡 214122)

**摘要:** 深度学习算法的有效性依赖于大量的带有标签的数据, 迁移学习的目的是利用已知标签的数据集(源域)来对未知标签的数据集(目标域)进行分类, 因此深度迁移学习的研究成为了热门。针对训练数据标签不足的问题, 提出了一种基于多尺度特征融合的领域对抗网络(Multi-scale domain adversarial network, MSDAN)模型, 该方法利用生成对抗网络以及多尺度特征融合的思想, 得到了源域数据和目标域数据在高维特征空间中的特征表示, 该特征表示提取到了源域数据和目标域数据的公共几何特征和公共语义特征。将源域数据的特征表示和源域标签输入到分类器中进行分类, 最终在目标域数据集的测试上得到了较为先进的效果。

**关键词:** 深度学习; 迁移学习; 多尺度特征融合; 对抗域适应; 公共语义特征

**中图分类号:** TP181

**文献标志码:** A

## Multi-scale Domain Adversarial Network for Transfer Learning

LIN Jiawei<sup>1</sup>, WANG Shitong<sup>2</sup>

(1. College of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; 2. Internet of Things Application Technology Key Construction Laboratory of Jiangsu Province, Wuxi 214122, China)

**Abstract:** The effectiveness of deep learning algorithms depends on a large amount of labeled data. The purpose of transfer learning is to use a dataset with known labels (source domain) to classify a dataset with unknown labels (target domain), so the research of deep transfer learning has become a hotspot. For the problem of insufficient training data labels, a model of multi-scale domain adversarial network (MSDAN) based on multi-scale feature fusion is proposed. This method uses the idea of generating adversarial networks and multi-scale feature fusion to obtain the feature representation of the domain data and the target domain data in a high-dimensional feature space. The feature representation extracts common geometric features and common semantic features of the source domain data and the target domain data. The feature representation of the source domain data and the source domain label are input into the classifier for classification, and finally more advanced effect is obtained in the test of the target domain dataset.

**Key words:** deep learning; transfer learning; multi-scale feature fusion; adversarial domain adaptation; common semantic feature

## 引 言

深度学习已经被证明在图像识别<sup>[1]</sup>、推荐搜索<sup>[2]</sup>等领域取得良好的效果。尤其是卷积神经网络(Convolutional neural networks, CNN)在ImageNet数据集的图像识别上大放异彩<sup>[3]</sup>,使得深度学习的研究愈加火热,成为人工智能领域的重要分支。虽然深度学习方法能够自动地提取原始数据的高级语义特征,并且能够满足实际应用中的端到端的需求,但是传统的深度学习模型属于标准的监督式学习。监督式学习的有效性有两个必要条件,一是依赖于大量的带有标签的训练数据,二是需要假设训练数据和测试数据服从相同的分布,二者缺一不可。然而在实际应用场景中这两个条件并不一定成立,一种情况是如果需要在一个新的场景下实现分类任务,这时需要收集大量的训练数据并且要标注数据,数据标注的过程必然会浪费大量的人力资源,这是制约深度学习向前进一步发展的最大阻碍。另一种情况是受到外界环境的影响,数据的分布会随着时间、天气和视点的不同而发生变化,造成了数据集偏差,使得训练数据和测试数据之间的分布不匹配,这样的网络即使在训练集上表现良好,但在测试集上的分类性能会大大降低。因此,解决标签不足和数据集偏差的问题成为了深度学习领域的又一挑战。

针对标签不足和数据集偏差问题,迁移学习是一种强有力的解决方法。迁移学习是一种新颖的学习策略,近年来引起了许多人工智能学者的广泛关注。它的主要任务是通过缩小源域数据和目标域数据之间的分布差异来将相关知识从数据标签丰富的源域迁移到数据标签匮乏的目标域,利用源域的知识来辅助目标域的数据进行训练。由于迁移学习的任务不需要对目标域数据进行标注,因此在一个全新的应用场景下应用迁移学习的思想会使得模型在精度和泛化性能上具有较大的优势。根据迁移的内容不同,可以将迁移学习的方法分为如下4类:基于样本的迁移学习方法<sup>[4-5]</sup>、基于特征的迁移学习方法<sup>[6-7]</sup>、基于参数的迁移学习方法<sup>[8-9]</sup>和基于关系-知识的迁移学习方法。本文主要关注的是基于特征的迁移学习方法,主要思想是最小化两个域的数据在共享子空间的距离,也就是说通过一个特征变换使得源域数据和目标域数据在某个特征空间上表现出相似的性质。基于特征的迁移学习方法的主要区别在于特征空间的构造方式不同以及在特征空间上不同域之间的距离度量不同。通过迁移学习方法得到源域数据和目标域数据在特征子空间的表示后,便可以利用传统的机器学习算法对源域数据训练,从而实现了对缺乏标注的目标域数据进行分类。

近年来,人工智能领域已经涌现出许多关于深度迁移学习方法的研究<sup>[10-11]</sup>,主要应用在对象识别领域。深度迁移学习方法的基本思想是通过将源域分类误差和不同领域之间差异的凸组合最小化来预测目标域数据集分类误差的误差上限<sup>[12]</sup>。实际上,深度迁移学习网络大多数是通过在网络的特征提取器和分类器之间增加一个适配层,该适配层计算在特征子空间中源域和目标域数据之间的距离,将该距离加入损失函数中进行训练,通过最小化该距离来拉近源域和目标域数据的分布,使得提取到的不同域之间的深层表示有着极为相似的语义特征。Tzeng等<sup>[13]</sup>提出了深度域混淆网络(Deep domain confusion, DDC),DDC在AlexNet的网络架构上引入了最大均值差异(Maximum mean discrepancy, MMD)作为不同域之间的距离度量,以此来减少不同数据域之间的分布不匹配问题,通过联合优化源域损失和MMD来减少目标域数据的损失。Long等<sup>[14]</sup>提出了深度适配网络(Deep adaptation networks, DAN),该网络在特征提取器和分类器之间添加了3个适配层,并且在每个适配层采用了表征能力更好的多核MMD(Multi kernel-MMD, MK-MMD)来度量不同数据域之间的分布差异,DAN网络的训练时间并没有增加,但是在多个任务上却取得

了比DDC更好的效果。Tzeng等<sup>[15]</sup>认为传统的深度域适应网络只在不同数据集之间进行了边缘分布的适配,没有考虑到类别之间的信息,因此提出了一个软标签损失来进行不同数据域的标签适配,最终的损失函数应包括源域的分类损失、域适应损失和软标签损失3部分。Ganin等<sup>[16]</sup>提出的领域对抗神经网络(Domain-adversarial neural network, DANN)首次利用对抗机制的思想来进行深度迁移,将特征提取器看成生成对抗网络(Generative adversarial networks, GAN)<sup>[17]</sup>的生成器,将域适配器看成生成对抗网络的判别器,以此提取出源域和目标域的共同特征来实现知识的迁移。Bousmalis等<sup>[18]</sup>在领域分离网络(Domain separation networks, DSN)中提出了公有特征和私有特征的全新概念,DSN认为源域和目标域都由两部分组成:公有部分和私有部分。公有部分用来学习源域和目标域之间的公有特征,私有部分用来保持不同域独立的特性。DSN认为深度域适应网络的损失函数应该由4部分组成:分类损失、私有部分的重构损失、私有部分与公有部分的差异性损失以及源域和目标域公有部分的相似性损失。Ghifary等<sup>[19]</sup>在深度重构分类网络(Deep reconstruction-classification networks, DRCN)中利用自动编码器的思想,在传统的CNN上添加了一个新的分支,该分支利用上采样的思想将数据的深层特征映射回原始数据,这样保证在提取源域的特征进行分类的同时添加了目标域数据的部分特征。针对样本标签空间不同的问题,Cao等<sup>[20]</sup>提出了选择性迁移网络(Selective adversarial networks, SAN),SAN认为在域适应的过程中,源域的类别往往很多,而目标域的标签类别通常只会包含一小部分,直接用源域所有数据来进行域适应不可避免地会产生负迁移的效果。基于此问题,SAN提出了两种约束,一种约束是基于样本级别的概率权重判别器,该约束可以选择源域中和目标域数据更为相似的样本参与迁移,另一种约束是基于类别级别的约束,该约束避免了不在目标域标签空间中的类别参与迁移。周晶雨等<sup>[21]</sup>提出了目标域不平衡场景下的在线迁移学习,设计了在目标域输入空间和特征空间过采样的方法有效地增强了目标域的分类性能。Long等<sup>[22]</sup>提出的条件对抗域适应(Conditional adversarial domain adaptation, CDAN)不仅对齐了源域和目标域的特征信息而且也对齐了标签信息,并且利用了多线性调整的方法来捕获多模态结构。Zhu等<sup>[23]</sup>认为以往的域适应都是基于全局的域对齐方法,提出的深度子域适应网络基于最大均值差异的思想提出了局部最大均值差异来对齐相关子域的分布,局部最大均值差异中源域数据的类别用真实的one-hot向量,而目标域数据的类别用网络分类得到的伪标签。针对开放场景下的域适应,分离适配(Separate to adapt, STA)<sup>[24]</sup>采用渐进式分离的方法逐步分离目标域中的已知类和未知类,STA认为目标域中的已知类和源域仅仅具有分布上的偏移,而未知类和源域不仅有分布间隙,而且还存在着语义间隙。STA首先利用源域数据训练多二元分类器以估计目标域中的数据和源域类别的相似性,将相似性高的作为已知类数据,相似性低的作为未知类数据,之后利用它们训练出一个细粒度的分类器来对目标域样本进行精细的分离。

在计算机视觉领域,浅层网络提取的特征往往是图像的边缘和纹理等几何特征,深层网络提取的特征往往是具有语义信息的更高级别的特征。传统的深度域适应方法都是基于单个特征图的特征来进行域适应,如果特征提取网络较浅,域适应过程只注重了几何特征信息的对齐而忽略了语义特征信息的对齐;如果特征提取网络较深,域适应过程只注重了语义特征信息的对齐而忽略了几何特征信息的对齐。现有的基于卷积神经网络的深度域适应方法都是基于单一特征结构来对齐不同数据域之间的分布,这样的方法会导致只有部分特征信息能够从源域迁移到目标域。为了解决这一问题,本文提出基于多尺度特征融合的领域对抗网络(Multi-scale domain adversarial network, MSDAN)模型。具体来说,MSDAN通过对齐多个深层表示之间的分布来学习基于多结构的域不变的表示以迁移更多的信

息。通过实验表明,该算法能够有效地提高图像领域的跨域识别精度。本文的主要贡献可以归纳如下:

(1) 利用GAN的思想提出了一个基于多尺度特征融合的领域对抗网络模型来提取基于多结构的域不变表示。

(2) 通过在传统的数字数据集上的分类和迁移学习经典数据集 Office-Caltech10 上进行多次实验证明了MSDAN在迁移学习的任务上与其他算法相比有着明显的优越性。

## 1 相关工作

### 1.1 生成对抗网络

2014年,GAN作为一种全新的非监督式深度学习架构得到了快速的发展。GAN利用生成器 $G$ 和判别器 $D$ 互相对抗的机制来模拟数据的真实分布,在图像生成、风格迁移以及人体姿态识别等多个领域具有较为广泛的应用。其中生成器 $G$ 用来模拟数据的真实分布,生成以假乱真的数据;判别器 $D$ 用来判别输入到网络的数据是来自于真实分布还是生成的分布。模型训练的目的在于通过优化生成器 $G$ 使得判别器的误差函数达到最大,优化判别器 $D$ 使得判别器的误差函数达到最小。训练过程中二者互相对抗以得到最优的生成器和判别器。用 $x$ 表示真实数据, $z$ 表示随机噪声,GAN网络的目标函数最终表示为

$$V(G, D) = E_{x \in P_{\text{data}}} [\lg D(x)] + E_{z \in P_z} [\lg (1 - D(G(z)))] \quad (1)$$

$$G^* = \arg \min_G \max_D V(G, D) \quad (2)$$

式(1,2)表明GAN的训练过程是一个最小最大优化问题,在网络训练过程中通过交替优化生成器 $G$ 和判别器 $D$ ,最终达到生成器 $G$ 模拟出的数据分布尽量接近真实分布的目的。深度域适应问题要求输入到分类器中的深层特征表示不包含域的判别性信息,利用这一特性将GAN的判别器作为域适应组件来提取源域数据和目标域数据的公共特征,当判别器无法判别输入的数据来自于哪个数据域时即表明提取的特征是源域数据和目标域数据的共同特征。

### 1.2 计算机视觉中的多尺度问题

在卷积神经网络提取特征的过程中,随着网络深度的不断加深,特征图所表示的语义特征越来越抽象,下采样的倍数越来越高,感受野也会随之变大,但是特征图的分辨率会随之降低,特征图所表示的空间几何信息也会越来越差。因此浅层特征图的分辨率较高,几何特征信息表征能力强,感受野较小,适合处理小目标对象;深层特征图的分辨率较低,语义特征信息表征能力强,感受野较大,适合处理大目标对象。实验表明利用多尺度模型融合不同特征图包含的信息作用于下游任务能够增强模型的鲁棒性和准确率。现有的多尺度模型结构大致分为如下4类:多尺度输入、多尺度特征融合、多尺度特征预测融合以及以上方法的组合。MSDAN模型利用的是多尺度特征融合的思想,利用深度网络多个尺度下的融合特征进行深度域适应。

## 2 MSDAN

本文利用并行分支网络结构(Inception结构)<sup>[23]</sup>的思想提出了基于多尺度融合的领域对抗网络。Inception结构的想法是从网络宽度上对CNN进行改进,把经过不同尺度大小卷积核的特征图进行拼接,经过较大卷积核得到的特征图提取了图像的深层特征,经过较小卷积核得到的特征图提取了图像的浅层特征。最终得到的特征图融合了不同尺度的特征,如图1所示。

将输入的带标签的源域数据和不带标签的目标域数据得到了多尺度特征后,利用判别器来提取源

域数据和目标域数据的共同特征,之后在共同特征的基础上可以利用带标签的源域损失来代替不带标签的目标域损失。

### 2.1 算法原理

为完整地描述基于并行分支的MSDAN模型,该模型的数据输入分别从联合概率分布 $P(x^s, y^s)$ 和边缘概率分布 $Q(x^t)$ 中采样得到, $P$ 和 $Q$ 不相等。从 $P$ 中采样得到的数据组成源域 $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ ,从 $Q$ 中采样得

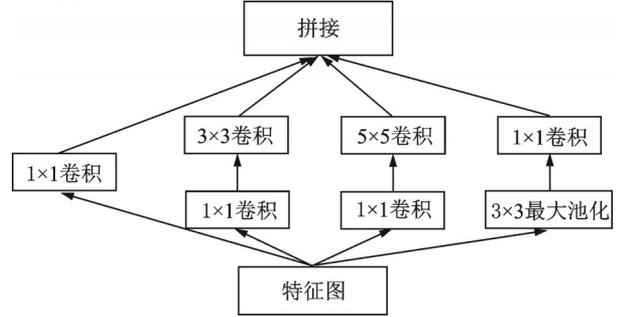


图1 Inception块结构图  
Fig.1 Architecture of Inception block

到数据组成目标域 $\mathcal{D}_t = \{x^t\}_{j=1}^{n_t}$ 。MSDAN网络的目的是计算得到条件概率 $P(Y|X = x_t)$ ,即通过带有标签的源域数据和不带标签的目标域数据训练网络来预测目标域数据的类别。从图2可以看出,MSDAN模型主要由以下4个模块组成:卷积特征提取模块 $F_c(\cdot)$ ,多尺度特征融合模块 $F_i(\cdot)$ ,分类器模块 $F_y(\cdot)$ 和判别器模块 $F_d(\cdot)$ ,其对应参数分别为 $\theta_c, \theta_i, \theta_y, \theta_d$ 。网络的域适应过程是利用GAN的博弈思想实现的, $F_c, F_i$ 起到GAN中生成器的作用, $\theta_c, \theta_i$ 优化的目的是使得目标域生成的特征尽量接近源域的特征,使得判别器的输出损失达到最大,判别器无法判别出输入的特征是来自于源域还是目标域; $F_d$ 起到GAN中判别器的作用, $\theta_d$ 优化的目的是使得判别器的输出损失达到最小,判别器能够有效地判断出输入的数据是来自于源域还是目标域。通过 $(\theta_c, \theta_i)$ 和 $\theta_d$ 的互相对抗得到最优的生成器和判别器,这样源域和目标域数据通过 $F_c$ 和 $F_i$ 得到的特征分布是最为相似的。

对于输入到MSDAN中的第 $i$ 个数据 $(x_i, y_i)$ 或 $x_i$ ,如式(3)和式(4)所示,MSDAN的分类损失和域损失分别采用标准的交叉熵损失。其中 $d_i$ 是自行构建的输入数据的域标签,如果输入到网络中的数据来自于 $\mathcal{D}_s, d_i = 0$ ;如果输入到网络中的数据来自于 $\mathcal{D}_t, d_i = 1$ 。 $e(y_i)$ 为数据真实类别标签的one-hot向量。

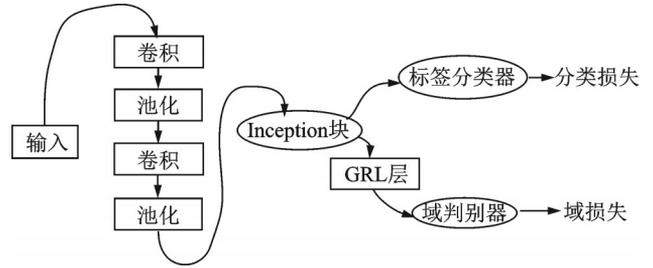


图2 基于Inception块的MSDAN架构图  
Fig.2 Architecture of MSDAN based on Inception block

$$\mathcal{L}_y(F_y(F_i(F_c(x_i))), y_i) = -e(y_i)^T \lg [F_y(F_i(F_c(x_i)))] \quad (3)$$

$$\mathcal{L}_d(F_d(F_i(F_c(x_i))), d_i) = -d_i \lg [F_d(F_i(F_c(x_i)))] - (1 - d_i) \lg [1 - F_d(F_i(F_c(x_i)))] \quad (4)$$

为了方便表示,将式(3,4)表示为

$$\mathcal{L}_y(\theta_y, \theta_i, \theta_c) = \mathcal{L}_y(F_y(F_i(F_c(x_i))), y_i) \quad (5)$$

$$\mathcal{L}_d(\theta_d, \theta_i, \theta_c) = \mathcal{L}_d(F_d(F_i(F_c(x_i))), d_i) \quad (6)$$

将源域数据 $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ 和目标域数据 $\mathcal{D}_t = \{x^t\}_{j=1}^{n_t}$ 输入到MSDAN可得到完整的优化目标为

$$\mathcal{E}(\theta_y, \theta_d, \theta_i, \theta_c) = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{L}_y^{\times 1}(\theta_y, \theta_i, \theta_c) - \lambda \left[ \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{L}_d^i(\theta_d, \theta_i, \theta_c) + \frac{1}{n_d} \sum_{i=n_s+1}^{n_s+n_t} \mathcal{L}_d^i(\theta_d, \theta_i, \theta_c) \right] \quad (7)$$

通过使目标函数  $\mathcal{E}(\theta_y, \theta_d, \theta_i, \theta_c)$  达到最小值可以得到该模型的最优解为  $(\theta_y^*, \theta_d^*, \theta_i^*, \theta_c^*)$ , 使用交替优化的思想来寻求最优解。

$$(\hat{\theta}_y, \hat{\theta}_i, \hat{\theta}_c) = \underset{\theta_y, \theta_i, \theta_c}{\operatorname{argmin}} \mathcal{E}(\theta_y, \hat{\theta}_d, \theta_i, \theta_c) \quad (8)$$

$$(\hat{\theta}_d) = \underset{\theta_d}{\operatorname{argmax}} \mathcal{E}(\hat{\theta}_y, \theta_d, \hat{\theta}_i, \hat{\theta}_c) \quad (9)$$

从式(8,9)可以看出,  $\theta_i, \theta_c$  的更新方向会使得模型的分类损失减少而域损失增大;  $\theta_d$  的更新方向会使得模型的域损失增大。在训练过程中, 这种对抗机制的实现方式是在判别器模块前面添加梯度反转层(Gradient reversal layer, GRL)实现的。在网络训练过程中该层没有需要更新的参数, 在反向传播时, 流经该层的梯度会乘以一个负数。通过Adam优化器<sup>[24]</sup>的作用, 域适应损失在  $\theta_d$  上的作用方式表现为梯度下降, 在  $\theta_c, \theta_i$  上的作用方式表现为梯度上升。网络优化的结果是求出一组参数  $(\theta_c^*, \theta_i^*, \theta_y^*, \theta_d^*)$ , 使得MSDAN的分类损失和域损失同时达到最小值。

## 2.2 算法过程

输入: 样本集  $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ ,  $\mathcal{D}_t = \{x^t\}_{j=1}^{n_t}$ ;

网络参数  $\theta = (\theta_c, \theta_i, \theta_y, \theta_d)$ ;

平衡系数  $\lambda$ , 迭代次数 epochs, 学习率  $\alpha$ ;

Adam优化器参数:  $\delta = 10^{-8}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $m_0 = 0$ ,  $v_0 = 0$ 。

输出: 最优参数  $\theta^* = (\theta_c^*, \theta_i^*, \theta_y^*, \theta_d^*)$

算法步骤

步骤1 网络参数、超参数的初始化。

步骤2 输入第  $i$  个源域数据  $(x_i^s, y_i^s)$  和第  $i$  个目标域数据  $x_i^t$ ;

步骤3 源域数据的前向传播

$$\begin{aligned} \mathcal{L}_y(\theta_c, \theta_i, \theta_y) &= -e^{(y_i^s)^T} \lg [F_y(F_i(F_c(x_i^s)))] \\ \mathcal{L}_d^s(\theta_c, \theta_i, \theta_d) &= -\lg [1 - F_d(\text{GRL}(F_i(F_c(x_i^s))))] \end{aligned}$$

步骤4 目标域数据的前向传播

$$\mathcal{L}_d^t(\theta_c, \theta_i, \theta_d) = -\lg [F_d(\text{GRL}(F_i(F_c(x_i^t))))]$$

步骤5 目标函数

$$\begin{aligned} \mathcal{E}(\theta) &= \mathcal{E}(\theta_c, \theta_i, \theta_y, \theta_d) = \\ &\mathcal{L}_y(\theta_c, \theta_i, \theta_y) + \lambda(\mathcal{L}_d^s(\theta_c, \theta_i, \theta_d) + \mathcal{L}_d^t(\theta_c, \theta_i, \theta_d)) \end{aligned}$$

步骤6 反向传播求梯度

$$\Delta\theta = (\Delta\theta_c, \Delta\theta_i, \Delta\theta_y, \Delta\theta_d) = (\nabla_{\theta_c} \mathcal{E}(\theta), \nabla_{\theta_i} \mathcal{E}(\theta), \nabla_{\theta_y} \mathcal{E}(\theta), \nabla_{\theta_d} \mathcal{E}(\theta))$$

步骤7 Adam参数更新

$$\begin{aligned} m &\leftarrow \beta_1 \cdot m + (1 - \beta_1) \cdot \Delta\theta \\ v &\leftarrow \beta_2 \cdot v + (1 - \beta_2) \cdot \Delta\theta \otimes \Delta\theta \\ &\wedge \\ m &\leftarrow m / (1 - \beta_1^i) \\ &\wedge \\ v &\leftarrow v / (1 - \beta_2^i) \\ (\theta_c, \theta_i, \theta_y, \theta_d) &= \theta \leftarrow \theta - \alpha \cdot \hat{m} / (\sqrt{\hat{v}} + \delta) \end{aligned}$$

步骤8 返回  $\theta^* = (\theta_c, \theta_i, \theta_y, \theta_d)$

### 3 实验和分析

实验所用的数据集包括两部分,一部分是经典的数字图像分类数据集,另一部分是迁移学习经典数据集 Office-Caltech10。本实验将所提出的 MSDAN 模型与其他经典算法进行对比,证明多尺度融合的有效性。所有实验均在 Python3.6 以及 Pytorch1.0 的框架下实现,计算机处理器配置为 Intel® Core™ i5-4210 CPU@2.6 GHz,内存 8 GB,在 Windows10 环境下配置 Python 环境以及 Pytorch 框架。

#### 3.1 数字分类数据集下的实验

本组实验研究了 MSDAN 在多个经典手写数字数据集上的域适应性能。所用数据集包括 MNIST 数据集, MNIST-M 数据集, SVHN 数据集和 USPS 数据集。实验的目的是用一个带标签的数据集和另一个不带标签的数据集进行训练以对不带标签的数据集进行分类。

MNIST 数据集包括 250 个不同人手写的 10 分类灰度图像,有 6 万张的训练图像和 1 万张的测试图像,像素大小为  $28 \times 28$ 。MNIST-M 数据集由 MNIST 数据集和 BSDS500 数据集的随即色块混合而成,像素大小为  $32 \times 32$ ,包括 59 000 张图像。SVHN 是采自于 Google 街景图像中的门牌号,图像大小为  $32 \times 32$ ,包括 4 578 张训练图像和 1 627 张测试图像。USPS 数据集是美国邮政手写数字数据集,像素大小为  $32 \times 32$ ,包括 999 张训练图像和 250 张测试图像。在数据预处理过程中,所有输入图像通过随机裁剪、旋转和翻转的方式实现数据扩增,并把所有图像像素归一化到  $[0, 1]$  之间。

现将提出的 MSDAN 算法与以下算法进行比较来证明 MSDAN 算法的优越性:(1)CovNet<sub>src</sub>。仅在源域上训练在目标域上测试的卷积神经网络;(2)SDA<sup>[25]</sup>。堆叠式降噪自动编码器,已成功应用到情感分析领域的迁移学习算法;(3)DRCN。深度重构分类网络,通过自动编码器的原理将目标域属于重构来进行迁移;(4)SA<sup>[26]</sup>。子空间对齐方法,通过优化一个函数来将源域空间映射到目标域空间实现无监督的迁移学习;(5)DANN。域对抗神经网络,通过生成对抗网络的思想将特征提取器和域适配器互相对抗来实现深层特征的领域不变性;(6)CovNet<sub>tgt</sub>。通过带标签的目标域数据来训练网络模型并在目标域数据集上进行测试。

网络的详细配量以及复杂度分析如表 1 所示。在不同数字数据集的深度域适应网络中,网络的隐藏层均使用 RELU 激活函数,在标签分类模块和域适应模块的最后一层均使用交叉熵函数来作为损失函数。网络使用 Adam 作为优化函数,学习率大小设置为 0.01 并使用学习率衰减策略,其中  $\lambda$  是需要手动调节的超参数。

表 2 总结了在相同环境下的数字数据集独立运行 10 次的跨域识别精度。在 3 组跨域识别的任务上应用了 MSDAN 模型,从表 2 可以看出,在 MNIST 数据集到 MNIST-M 数据集的跨域识别任务上,MSDAN 的分类精度达到了 81.9%,较 DANN 算法提高了约 4%。在 SVHN 数据集到 MNIST 数据集的跨域识别任务上,分类精度达到了 70.8%,较 DANN 算法提高了约 4%。在 MNIST 数据集到 USPS 数据集上的跨域识别精度达到了 87.9%,比 DANN 算法提高了约 3%。综合 3 组数字数据集的跨域识别精度,MSDAN 在图像跨域识别任务上的精度比 DANN 算法平均提高了 3.8%,MSDAN 算法精度提升的关键是利用了多尺度特征融合的思想来提取出基于多结构的图像域不变表示,能够将源域的更多信息迁移到目标域,例如图像的色调、亮度以及饱和度,与基于单一结构的图像域不变表示的 DANN 算法相比,迁移的知识更加全面。因此,在无监督域适应的分类任务上,MSDAN 模型相较于其他域适应模

表1 数字数据集参数设置以及复杂度分析

Table 1 Parameter setting and complexity analysis of digital datasets

类型	大小	数量	填充	步长	输出	参数数量/ $10^3$	计算量/ $10^6$
convolution	$5 \times 5$	64	0	1	$24 \times 24 \times 64$	4.8	2.76
max pool	$2 \times 2$	1	0	2	$12 \times 12 \times 64$	0	0
inception	$1 \times 1, 3 \times 3, 5 \times 5$	64	0, 1, 2	1	$12 \times 12 \times 64$	6.66	71 831
convolution	$5 \times 5$	50	0	1	$8 \times 8 \times 50$	80	5.12
max pool	$2 \times 2$	1	0	2	$4 \times 4 \times 50$	0	0
linear	$800 \times 100$				100	80	0.160
linear	$100 \times 100$				100	10	0.020
linear	$100 \times 10$				10	1	0.002
linear	$800 \times 100$				100	80	0.160
linear	$100 \times 2$				2	0.2	0.000 4

型具有明显的优越性。

### 3.2 Office-Caltech10数据集下的实验

本组实验研究了MSDAN在Office-Caltech10数据集不同域上的域适应性。该数据集包括4个数据域: Amazon, Caltech, Dslr, Webcam, 每个数据域包括10个类别的图片。4个域的图片是在不同的环境下由单反相机和网络相机拍摄的图像, 拥有着不同的数据分布。训练前将图像像素值归一化到 $[0, 1]$ 之间。在每次迭代过程中, 每张图像随机裁剪为 $224 \times 224$ 的大小并且随机进行水平翻转。

在Office-Caltech10数据集上将MSDAN算法与以下算法进行了对比实验: (1)

CovNet<sub>src</sub>, 仅在源域上训练在目标域上测试的AlexNet网络; (2) JGSA<sup>[27]</sup>, 联合几何对齐和统计对齐, 同时减少域之间的分布差异和几何差异来进行域适应; (3) DDC, 深度域混淆网络, 通过在网络中添加适配层来减少不同域之间的最大均值差异以此来保持领域不变性; (4) DANN, 域对抗神经网络, 利用GAN的思想来实现源域和目标域的不可区分性; (5) DARN<sup>[28]</sup>, 深度对抗重构分类网络, 在生成对抗网络的基础上, 通过在优化目标中添加目标域的重构损失来进行目标域数据的信息保持; (6) CovNet<sub>tgt</sub>, 通过带标签的目标域数据来训练网络模型并在目标域数据集上进行测试。

在Office-Caltech数据集上进行实验的网络参数如表3所示。在网络正向传播过程中使用RELU函数作为激活函数, 在标签分类模块和域适应模块的最后一层均使用Softmax函数将网络输出归一化到 $0 \sim 1$ 之间, 网络使用交叉熵函数作为损失函数。网络使用随机梯度下降法优化目标函数, 特征提取模块和域适应模块学习率为 $10^{-5}$ , 分类器学习率大小为 $10^{-3}$ 。

表4总结了相同环境下在Office-Caltech10数据集的不同域之间独立运行10次的跨域识别精度,

表2 数字数据集运行10次平均分类精度

Table 2 Average classification accuracy of 10 times on the digital datasets %

方法	MNIST	SVHN	MNIST	Average
	MNIST-M	MNIST	USPS	
CovNet <sub>src</sub>	52.2	54.9	85.5	64.2
SDA	42.6	55.2	43.1	46.9
DRCN	42.1	60.9	49.4	50.8
SA	56.7	59.3	85.9	67.3
DANN	77.5	66.6	85.2	76.4
MSDAN	81.9	70.8	87.9	80.2
CovNet <sub>tgt</sub>	95.9	99.4	96.1	97.1

表3 Office-Caltech10数据集参数设置以及复杂度分析

Table 3 Parameter setting and complexity analysis of Office-Caltech10 dataset

类型	大小	数量	填充	步长	输出	参数数量/ $10^3$	计算量/ $10^6$
convolution	$11 \times 11$	64	2	4	$55 \times 55 \times 64$	23.2	70.3
max pool	$3 \times 3$	1	0	2	$27 \times 27 \times 64$	0	0
convolution	$5 \times 5$	192	2	1	$27 \times 27 \times 192$	307.2	224
max pool	$3 \times 3$	1	0	2	$13 \times 13 \times 192$	0	0
convolution	$3 \times 3$	384	1	1	$13 \times 13 \times 384$	663.5	112.1
convolution	$3 \times 3$	256	1	1	$13 \times 13 \times 256$	884.7	149.5
convolution	$3 \times 3$	256	1	1	$13 \times 13 \times 256$	589.8	23
max pool	$3 \times 3$	1	0	2	$6 \times 6 \times 256$	0	0
inception	$1 \times 1, 3 \times 3, 5 \times 5$	256	0,1,2	1	$6 \times 6 \times 256$	176.6	63.6
linear	$9\ 216 \times 4\ 096$				4 096	37 748.7	755
linear	$4\ 096 \times 4\ 096$				4 096	16 777.7	335.5
linear	$4\ 096 \times 10$				10	40	0.8
linear	$9\ 216 \times 1\ 024$				1 024	9 437	188.7
linear	$1\ 024 \times 1\ 024$				1 024	1 048.6	20.1
linear	$1\ 024 \times 2$				2	2	0.04

从表4可以看出,从Webcam数据域到Amazon数据域上的跨域识别任务的精度达到63.9%,从Dslr数据域到Amazon数据域上的精度达到64.7%,从Dslr数据域到Webcam数据域上的精度达到90.0%,MSDAN算法比DANN算法的平均识别精度提高了2.6%,这是因为MSDAN通过将卷积神经网络的多个深层表示进行对齐,通过这种方法提取了源域和目标域的公共几何特征和公共语义特征,将源域的更多知识迁移到目标域。如若在更大的数据集上测试算法的性能,可以用ResNet网络模型代替AlexNet网络模型。由表4可见基于多尺度特征融合的领域对抗网络相较于其他算法在域适应的任务上有着明显的优越性。

#### 4 结束语

对于图像分类场景下的迁移学习问题,提出了一种基于多尺度融合的领域对抗网络模型。该模型减少了原始数据在多个子空间下的域适应偏差,有效地提取了多个尺度空间下的不同数据集之间的共享特征。利用基于并行分支的多尺度融合模型来实现不同特征空间下的领域自适应,实验证明该方法较单一尺度下的领域自适应相比有较大的提升。这种通过并行分支的多尺度融合思想也可以推广到其他深度域适应网络中。

表4 Office-Caltech10数据集运行10次平均分类精度

Table 4 Average classification accuracy of 10 times on the Office-Caltech10 dataset %

方法	Webcam Amazon	Dslr Amazon	Dslr Webcam	Average
CovNet <sub>src</sub>	41.2	35.2	67.2	47.9
JGSA	40.8	38.7	93.2	57.6
DDC	72.1	42.5	71.0	61.0
DANN	61.1	62.7	86.7	70.3
MSDAN	63.9	64.7	90.0	72.9
DARCN	65.8	66.0	89.8	73.9
CovNet <sub>gr</sub>	81.3	81.3	93.3	85.3

## 参考文献:

- [1] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2016: 770-778.
- [2] 余永红,殷凯宇,王强,等.基于深度学习的欧几里得嵌入的推荐算法[J].南京航空航天大学学报,2020,52(5): 729-735.  
YU Yonghong, YIN Kaiyu, WANG Qiang, et al. Recommendation algorithm of Euclidean embedding based on deep learning [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 729-735.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, 1409(15): 1556-1563.
- [4] LONG Mingsheng, WANG Jianmin, DING Guiguang, et al. Transfer joint matching for unsupervised domain adaptation [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2014: 1410-1417.
- [5] ALJUNDI R, EMONET R, MUSELET D, et al. Landmarks-based kernelized subspace alignment for unsupervised domain adaptation[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2015: 56-63.
- [6] SUN Baochen, FENG Jiashi, SAENKO K. Return of frustratingly easy domain adaptation[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2016: 2058-2065.
- [7] YAN K, KOU L, ZHANG D. Learning domain-invariant subspace using domain features and independence maximization[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2018, 48(1): 288-299.
- [8] LONG Mingsheng, WANG Jianmin, DING Guiguang, et al. Adaptation regularization: A general framework for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26 (5): 1076-1089.
- [9] WANG Jindong. Visual domain adaptation with manifold embedded distribution alignment[J/OL]. ACM, 2018. <https://arxiv.org/abs/1807.07258>.
- [10] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2017: 7167-7176.
- [11] MOTIIAN S, PICCIRILLI M, ADJEROH D A, et al. Unified deep supervised domain adaptation and generalization[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice:IEEE, 2017: 5715-5725.
- [12] BEN D, BLITZER J, CRAMMER K, et al. Analysis of representations for domain adaptation[J]. Advances in Neural Information Processing Systems, 2006, 19: 137-144.
- [13] TZENG E, HOFFMAN J, ZHANG N, et al. Deep domain confusion: Maximizing for domain invariance[J/OL]. Computer Science, 2014. <https://arxiv.org/abs/1412.3474>.
- [14] LONG Mingsheng, CAO Yue, CAO Zhangjie, et al. Learning transferable features with deep adaptation networks[C]// Proceedings of International Conference on Machine Learning.[S.l.]:[s.n.], 2015: 97-105.
- [15] TZENG E, HOFFMAN J, DARRELL T, et al. Simultaneous deep transfer across domains and tasks[C]// Proceedings of IEEE International Conference on Computer Vision (ICCV). Santiago:IEEE, 2015 : 4068-4076.
- [16] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030.
- [17] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems.[S.l.]: NIPS, 2014: 2672-2680.
- [18] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, et al. Domain separation networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems.[S.l.]: NIPS, 2016: 343-351.
- [19] GHIFARY M, KLEIJN W B, ZHANG Mengjie, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[C]// Proceedings of European Conference on Computer Vision. [S.l.]:Springer, 2016: 597-613.
- [20] CAO Zhangjie, LONG Mingsheng, WANG Jianmin, et al. Partial transfer learning with selective adversarial networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition.[S.l.]:IEEE, 2018: 2724-2732.
- [21] 周晶雨,王士同.对不平衡目标域的多源在线迁移学习[J].智能系统学报,2021,17(2): 248-256.  
ZHOU Jingyu, WANG Shitong. Multi-source online transfer learning for imbalanced target domains [J]. CAAI Transactions on

- Intelligent Systems, 2021, 17(2): 248-256.
- [22] LONG Mingsheng, CAO Zhangjie, WANG Jianmin, et al. Conditional adversarial domain adaptation[C]// Proceedings of 32nd Conference on Neural Information Processing Systems. Montréal: NIPS, 2018.
- [23] ZHU Y, ZHUANG F, WANG J, et al. Deep subdomain adaptation network for image classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 99: 1-10.
- [24] LIU H, CAO Z, LONG M, et al. Separate to adapt: Open set domain adaptation via progressive separation[C]// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 2922-2931.
- [25] GLOROT X, BORDES A, BENGIO Y. Domain adaptation for large-scale sentiment classification: A deep learning approach [C]// Proceedings of ICML. New York: ACM, 2011: 513-520.
- [26] FERNANDO B, HABRARD A, SEBBAN M, et al. Subspace alignment for domain adaptation[J/OL]. Computer Science, 2014. <https://arxiv.org/abs/1409.5241>.
- [27] ZHANG Jing, LI Wanqing, OGUNBONA P. Joint geometrical and statistical alignment for visual domain adaptation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2017: 1859-1867.
- [28] 林佳伟, 王士同. 用于无监督域适应的深度对抗重构分类网络[J/OL]. 计算机科学与探索, 2021. <http://kns.cnki.net/kcms/detail/11.5602.tp.20210204.1711.026.html>.
- LIN Jiawei, WANG Shitong. Deep adversarial-reconstruction-classification networks for unsupervised domain adaptation[J/OL]. Journal of Frontiers of Computer Science and Technology, 2021. <http://kns.cnki.net/kcms/detail/11.5602.tp.20210204.1711.026.html>.

#### 作者简介:



林佳伟(1996-),男,硕士研究生,研究方向:人工智能和模式识别, E-mail: JNULJW@163.com。



王士同(1964-),通信作者,男,教授,研究方向:人工智能,模式识别等。

(编辑:夏道家)