Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2023, 12(2), 107-114
Published Online May 2023 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2023.122014

Hans汉斯

基于域特定批量归一化的对抗域适应图像分类

范博文，徐志洁

北京建筑大学理学院，北京

收稿日期：2023年4月12日；录用日期：2023年5月23日；发布日期：2023年5月31日

摘　要

无监督域适应(UDA)旨在将知识从带有大量标签的源域迁移到没有标签的目标域。目前的研究主要集中在统一两个域的特征分布上。然而，目标域通常具有更为复杂的背景信息，源域和目标域的全局特征分布并不相同，在源域和目标域之间直接共享整个网络强制全局分布对齐会导致性能的下降。针对此问题，提出了一种新的基于域特定批量归一化的对抗域适应模型。首先，采用对抗性学习损失模块，综合考虑域对齐和类别对齐，从对抗学习获得的混淆矩阵中自动构建一个新的损失函数来矫正自训练中的伪标签；其次，在卷积神经网络(CNN)的编码器架构中引入域特定批量归一化模块(DSBN)，通过分离批量归一化层来分别适应源域和目标域。将域特定信息与域不变信息分离，更好地学习域不变特征表示，来获得更好的泛化性能。最后，本文的方法在Office-Home数据集和Office-31数据集的准确率分别达到67.4%和89.4%，验证了模型的有效性。

关键词

无监督域适应，图像分类，域特定批量归一化

Domain Specific Batch Normalization Based on Adversarial Domain Adaptation Image Classification

Bowen Fan, Zhijie Xu

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Apr. 12th, 2023; accepted: May 23rd, 2023; published: May 31st, 2023

Abstract

Unsupervised domain adaptation (UDA) aims to transfer knowledge from a source domain with many labels to a target domain without labels. Current research mainly focuses on unifying the

文章引用: 范博文, 徐志洁. 基于域特定批量归一化的对抗域适应图像分类[J]. 人工智能与机器人研究, 2023, 12(2): 107-114. DOI: 10.12677/airr.2023.122014

feature distributions of the two domains. However, the target domain usually has more complex background information, the global feature distributions of the source and target domains are different, and directly sharing the entire network between the source and target domains to enforce global distribution alignment will lead to performance degradation. In response to this issue, a novel adversarial domain adaptation model is proposed based on domain-specific batch normalization. First, using the adversarial learning loss module, considering domain alignment and class alignment, a new loss function is automatically constructed from the confusion matrix obtained by adversarial learning to correct the pseudo-labels in self-training; second, a domain-specific batch normalization module (DSBN) is introduced in the encoder architecture of a convolutional neural network (CNN), which adapts to the source and target domains separately by separating the batch normalization layers. Separate domain-specific information from domain-invariant information and learn domain-invariant feature representations to achieve better generalization performance. Finally, the accuracy of the method in this paper in the Office-Home dataset and Office-31 dataset reached 67.4% and 89.4%, respectively, which verified the model's effectiveness.

Keywords

Unsupervised Domain Adaptation, Image Classification, Domain Specific Batch Normalization

1. 引言

目前在域适应方面的工作主要基于两种技术：域对抗学习和自训练。然而，域对抗学习只对齐领域之间的特征分布，而不考虑目标特征是否具有判别性。另一方面，自训练利用模型预测来增强对目标特征的识别，但它无法明确地对齐域分布。为了将这两种方法的优势结合起来，ALDA [1]通过引入了混淆矩阵以减少差距并对齐特征分布，该矩阵是通过对抗性方法学习的。从学习的混淆矩阵中自动构建一个新的损失函数，用作未标记目标样本的损失来校正伪标签。该方法可以同时考虑域对齐和类别对齐。尽管如此，由于源域和目标域具有不同的特征分布，该方法共享整个网络会导致性能的下降，仍然没有将域不变特征和域特定特征很好地分离。而本章的方法通过针对不同的域使用不同的批量归一化层，在学习域不变特征表示的同时进行域特定特征分离，获得更好的泛化性能，有效避免了上述问题。

我们认为出现上述问题的原因主要是由于源域和目标域之间是具有很多相似性的，因此我们经常需要依靠源域中的信息来学习网络适应未标记的目标域数据，所以两个域共享组件是不可避免的。因此，本章在基于对抗性域适应的基础上，同时考虑了类别对齐，通过构建新的目标域的损失函数来矫正伪标签。在此基础上，为了更好地学习域不变特征，使得分类器更准确，我们也考虑了分别对源域和目标域使用特定的批量归一化层的参数来捕获域特定信息，并利用该参数将域特定信息转换为域不变表示，使得模型具有优异的性能。

2. 相关知识

批量归一化层

批量归一化层(BN)是卷积神经网络中广泛使用的模块，用 $X \in \mathbb{R}^{H\times W\times N}$ 表示每个通道的激活，则 BN 可以表示为：

$$BN\left(x[i,j,n];\gamma,\beta\right)=\gamma\cdot\hat{x}[i,j,n]+\beta \tag{1}$$

其中：

$$\hat{x}[i,j,n]=\frac{x[i,j,n]-\mu}{\sqrt{\sigma^2+\in}} \tag{2}$$

其中$\in$是一个很小的常数，小批量内激活的平均值和方差μ和σ通过以下公式计算：

$$\mu=\frac{\sum_n\sum_{i,j}x[i,j,n]}{N\cdot H\cdot W} \tag{3}$$

$$\sigma^2=\frac{\sum_n\sum_{i,j}\left(x[i,j,n]-\mu\right)^2}{N\cdot H\cdot W} \tag{4}$$

在训练过程中，BN 通过更新因子为α的指数移动平均来估计整个激活的均值和方差，记为$\bar{\mu}$和$\bar{\sigma}$。形式上，给定第t个小批量，其均值和方差为：

$$\bar{\mu}^{t+1}=\left(1-\alpha\right)\bar{\mu}^t+\alpha\mu^t \tag{5}$$

$$\left(\bar{\sigma}^{t+1}\right)^2=\left(1-\alpha\right)\left(\bar{\sigma}^t\right)^2+\alpha\left(\sigma^t\right)^2 \tag{6}$$

请注意，如果域的分布之间存在明显差异，则共享源域和目标域的均值和方差是不合适的。

3. 基于特定批量归一化的对抗域适应模型

3.1. 网络架构

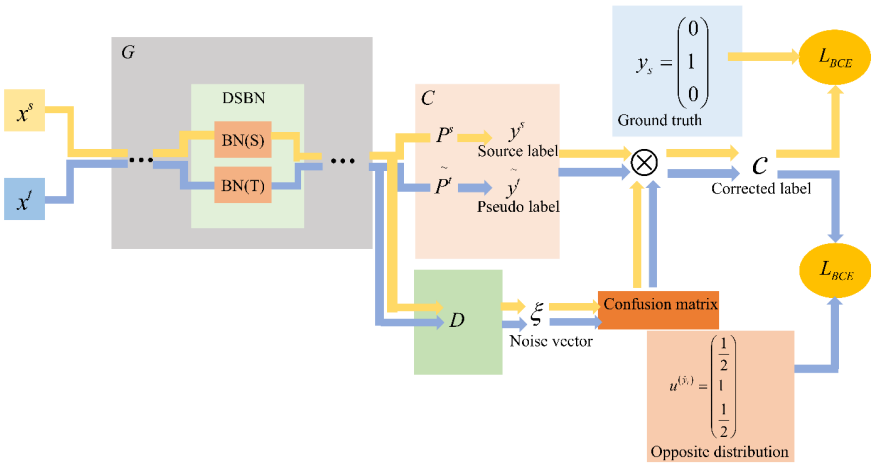


Figure 1. Network architecture diagram
图 1. 网络架构图

网络架构如图 1 所示，网络的输入为源域样本x^s和目标域样本x^t，通过特征提取器G提取特征，将传统的 BN 层针对不同的域替换为特定批量归一化层(DSBN)来捕获两个域的特定信息，DSBN 层由 BN 层中的两个分支组成，一个用于源域，另一个用于目标域。每个输入示例根据其域选择其中一个分支。除了 DSBN 的参数外，所有参数都在两个域之间共享，并有效地学习两个域中的共同特征，而域特定的信息则通过 DSBN 层的域特定参数有效地捕获。经标签分类器C得到源域的标签和目标域的伪标签，通过构造混淆矩阵来校正生成的伪标签。注意本章提出的模型中的域判别器D的输出为分类向量，并用该分类向量构造出混淆矩阵来优化特征提取器G、判别器D和标签分类器C。

3.2.损失函数

我们使用了噪声校正域判别器 D 来学习向量 $\xi^{(x_t)}$。如图 2 所示，校正噪声的域判别器 D 以特征 $G(x)$ 为输入，输出多类的得分向量 $D\left(G(x)\right)\in\mathbb{R}^K$，经过一个 sigmoid 层后输出为：

$$\xi^{(x)}=\sigma\left(D\left(G(x)\right)\right) \tag{7}$$

其中 $\xi^{(x)}$ 的每个分量表示伪标签与正确标签相同的概率，即：

$$\xi_k^{(x)}=p\left(y=k\left|\hat{y}=k,x\right.\right) \tag{8}$$

采用对抗学习的思想，使判别器和生成器进行极大极小博弈。我们不让判别器执行域分类任务，而是让判别器为源域和目标域生成不同的噪声向量。如图 2 所示，对于源特征 $G(x_s)$，判别器的目的是最小化校正后的标签向量 $c^{(x_s)}$。则源域的对抗性损失为：

$$\mathcal{L}_{Adv}\left(x_s,y_s\right)=\mathcal{L}_{BCE}\left(c^{(x_s)},y_s\right)=\sum_k -y_{sk}\log c_k^{(x_s)}-\left(1-y_{sk}\right)\log\left(1-c_k^{(x_s)}\right) \tag{9}$$

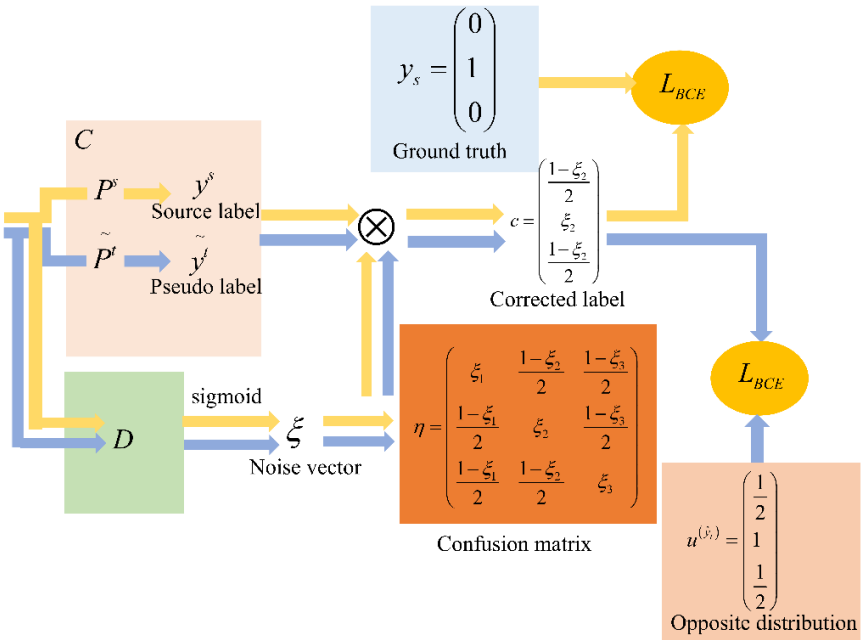


Figure 2. The illustration of noise-correcting domain discrimination (K = 3)
图 2. 噪声校正域判别器的示意图(K = 3)

对于目标域特征 $G(x_t)$，判别器以相反的方式进行操作，判别器将伪标签校正为相反的分布 $u_k^{(\hat{y}_t)}\in\mathbb{R}^K$，其中：

$$u_k^{(\hat{y}_t)}=\begin{cases}0 & k=\hat{y}_t\\ \dfrac{1}{K-1} & k\neq\hat{y}_t\end{cases} \tag{10}$$

则目标域的对抗性损失为：

$$\mathcal{L}_{Adv}\left(x_t\right)=\mathcal{L}_{BCE}\left(c^{(x_t)},u^{(\hat{y}_t)}\right) \tag{11}$$

因此对抗性总损失为：

$$\mathcal{L}_{Adv}\left(x_s,y_s,x_t\right)=\mathcal{L}_{Adv}\left(x_s,y_s\right)+\mathcal{L}_{Adv}\left(x_t\right) \tag{12}$$

判别器需要最小化损失函数来区分源域特征和目标域特征，生成器必须最大化损失函数来欺骗判别器，通过这种方式，我们的噪声校正判别器可以实现特征对齐。

由于对抗性学习的过程可能是不稳定的，我们将正则化项添加到判别器的损失中：

$$\mathcal{L}_{\mathrm{Re}g}\left(x_s,y_s\right)=\mathcal{L}_{CE}\left(p_D^{(x_s)},y_s\right) \tag{13}$$

其中 $p_D^{(x_s)}=soft\max\left(D\left(G\left(x_s\right)\right)\right)$，$\mathcal{L}_{CE}$ 是交叉熵损失。因此判别器的总损失为：

$$\min_D E_{(x_s,y_s),x_t}\mathcal{L}_{Adv}\left(x_s,y_s,x_t\right)+\mathcal{L}_{\mathrm{Re}g}\left(x_s,y_s\right) \tag{14}$$

在对混淆矩阵进行对抗性学习后，我们为目标域数据构建一个损失函数为：

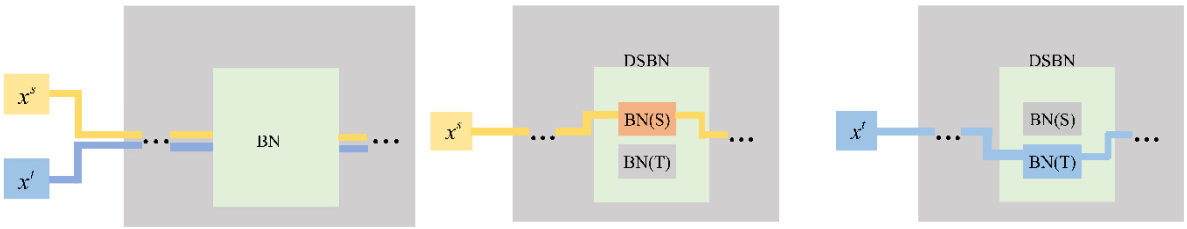
$$\mathcal{L}_T\left(x_t,\mathcal{L}_{unh}\right)=\sum_{k,l}\eta_{kl}^{(x_t)}p\left(\hat{y}_t=l\middle|x_t\right)\mathcal{L}_{unh}\left(p_t,k\right)=\sum_k c_k^{(x_t)}\mathcal{L}_{unh}\left(p_t,k\right) \tag{15}$$

其中 k,l 表示矩阵的行列数，因此总损失为：

$$\min_C E_{(x_s,y_s),x_t}\left(\mathcal{L}_{CE}\left(p_s,y_s\right)+\lambda\mathcal{L}_T\left(x_t,\mathcal{L}_{unh}\right)\right) \tag{16}$$

$$\min_G E_{(x_s,y_s),x_t}\left(\mathcal{L}_{CE}\left(p_s,y_s\right)+\lambda\mathcal{L}_T\left(x_t,\mathcal{L}_{unh}\right)-\lambda\mathcal{L}_{Adv}\left(x_s,y_s,x_t\right)\right) \tag{17}$$

其中 $\lambda\in[0,1]$ 是超参数。



a) BN　　b) DSBN

图 3. Illustration of difference between BN and DSBN
Figure 3. BN 和 DSBN 之间的差异说明

3.3. 特定批量归一化层

由于源域和目标域的特征分布存在差异，共享整个网络会导致性能下降，因此为了更好地学习域不变特征表示，将域特定信息分离，引入特定批量归一化层(DSBN)来为每个域生成不同的分布。图 3 说明了 BN 和 DSBN 之间的差异，形式上，DSBN 为每个域标签 $d\in\{S,T\}$ 分配域特定的参数 γ_d 和 β_d，用 $X_d\in\mathbb{R}^{H\times W\times N}$ 表示属于域标签 d 的每个通道的激活，则 DSBN 可以表示为：

$$\mathrm{DSBN}\left(x_d\left[i,j,n\right];\gamma_d,\beta_d\right)=\gamma_d\cdot\hat{x}_d\left[i,j,n\right]+\beta_d \tag{18}$$

其中

$$\hat{x}_d\left[i,j,n\right]=\frac{x_d\left[i,j,n\right]-\mu}{\sqrt{{\sigma_d}^2+\in}} \tag{19}$$

$$\mu_d=\frac{\sum_n\sum_{i,j}x_d\left[i,j,n\right]}{N\cdot H\cdot W} \tag{20}$$

$$\sigma_d^{\ 2} = \frac{\sum_n \sum_{i,j} \left(x_d[i,j,n] - \mu_d \right)^2}{N \cdot H \cdot W} \tag{21}$$

在训练过程中，DSBN 通过更新因子为α的指数移动平均来估计整个激活的均值和方差，其均值和方差为

$$\bar{\mu}_d^{\ t+1} = (1-\alpha)\bar{\mu}_d^{\ t} + \alpha\mu_d^{\ t} \tag{22}$$

$$\left(\bar{\sigma}_d^{\ t+1}\right)^2 = (1-\alpha)\left(\bar{\sigma}_d^{\ t}\right)^2 + \alpha\left(\sigma_d^{\ t}\right)^2 \tag{23}$$

期望 DSBN 通过分别估计每个域的特定参数来捕获特定域的信息。DSBN 允许网络更好地学习域不变特征，因为通过利用从给定域捕获的统计信息和学习参数，可以有效地去除网络中的域特定信息。

4. 实验结果与分析

4.1. 实验设置

在训练过程中，使用预训练好的 ResNet-50 作为生成器网络，使用动量为 0.9 的随机梯度下降法(SGD)来训练模型。初始学习率$\eta_0 = 0.01$，学习率调整按照公式$\eta_p = \frac{\eta_0}{(1+\alpha q)^{\beta}}$，其中$\alpha = 10$，$\beta = 0.75$。使用 PyTorch 框架来实现。在本章的方法中有两个超参数，伪标签的阈值δ和超参数λ。用阈值δ对目标样本进行选择，如果目标域的预测低于该阈值，则我们在训练时忽视这些样本。设置$\delta = 0.9$，λ从 0 增加到 1，增加幅度为$\frac{2}{1+\exp(-10 \cdot q)} - 1$。

比较的方法包括 ResNet-50 [2]、DANN [3]、ADDA [4]、MADA [5]、ALDA [1]、JAN [6]。对于所有上述方法均遵循原论文的实验结果。

4.2. 数据集

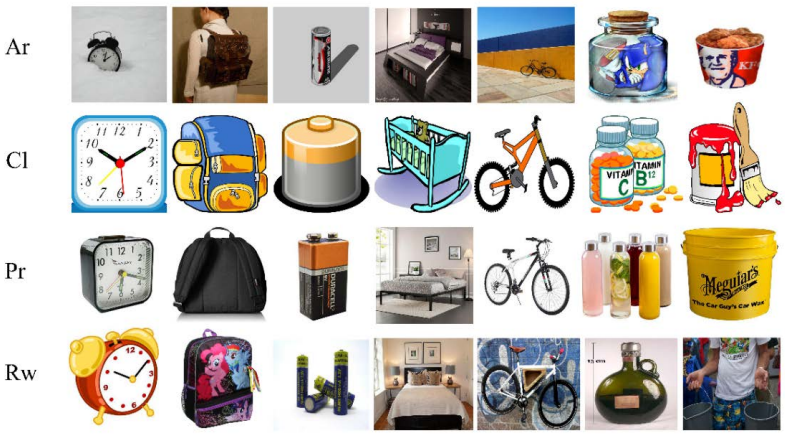


Figure 4. Office-Home 数据集的示例图像
图 4. Example images of Office-Home dataset

实验中使用了 Office-Home 数据集[7]和 Office-31 [8]数据集。其中 Office-Home 聚集包含了 65 类 15500 张图片。这些图片总共来自四个领域：以绘画等艺术描绘为主的艺术图像(Artistic Images, Ar)、剪贴画(Clip Art, Cl)、无背景的真实物品(Product, Pr)和用相机拍的真实世界中的物品照片(Real-World, Rw)。这些图像

如图 4 所示。Office-31 数据集共有 31 个类别的 4110 张图像，这些图像来自三个域，包括亚马逊(A)、网络摄像头(W)和数码单反(D)。其中，其中 A 是 2817 幅从网上商家下载的图像；W 包括 795 幅从网络摄像头获取的低分辨率图像；D 包括 498 幅数码单反的高分辨率图像。部分示例图像如图 5 所示。

图 5. Example images of Office-31 dataset
Figure 5. Office-31 数据集的示例图像

4.3. 实验结果与分析

我们在中等规模的 Office-Home 数据集上使用 ResNet-50 的结果如表 1 所示，本章提出的方法显著优于其他方法。由于 ALDA 的方法将对抗学习和自训练的方法相结合来学习，因此该方法比基于域对抗学习的方法如 DANN、JAN 取得更好的结果。然而，在域对齐方面，由于源域和目标域具有不同的特征分布，ALDA 共享整个网络会导致性能的下降，仍然没有很好地学习域不变特征。而本章的方法通过捕获域特定信息，取得了更好的实验结果。

Table 1. Accuracies (%) on Office-Home
表 1. Office-Home 数据集的精度(%)

Domains	ResNet-50 [2]	DANN [3]	JAN [6]	ALDA [1]	Ours
Ar→Cl	34.9	45.6	45.9	**53.7**	53.1
Ar→Pr	50.0	59.3	61.2	**70.1**	68.5
Ar→Rw	58.0	70.1	68.9	**76.4**	76.1
Cl→Ar	37.4	47.0	50.4	60.2	**65.3**
Cl→Pr	41.9	58.5	59.7	72.6	**73.4**
Cl→Rw	46.2	60.9	61.0	71.5	**72.6**
Pr→Rw	38.5	46.1	45.8	**56.8**	55.5
Pr→Ar	31.2	43.7	43.4	51.9	**53.9**
Pr→Cl	60.4	68.5	70.3	77.1	77.8
Rw→Ar	53.9	63.2	63.9	70.2	**72.6**
Rw→Cl	41.2	51.8	52.4	**56.3**	56.2
Rw→Pr	59.9	76.8	76.8	82.1	**84.3**
Avg	46.1	57.6	58.3	66.6	**67.4**

为了更好地检测模型在真实世界的数据上进行领域自适应时的表现，这一部分使用 Office-31 数据集

进行实验验证，我们在 Office-31 数据集的结果如表 2 显示。所提出的方法总体上优于所有比较方法，并将先进的结果平均从 88.7%提高到 89.4%。在具有挑战性的迁移任务(如 A→W 和 A→D)上表现优异，本章的方法也显示出显著的改善。上述实验结果证明，本文提出的方法可以进行准确的域对齐，同时考虑了类别对齐，尤其是在域差异较明显的情况下，提供更好的性能。

Table 2. Accuracies (%) on 31
表 2. Office-31 数据集的精度(%)

Domains	ResNet-50 [2]	DANN [3]	ADDA [1]	JAN [6]	MADA [5]	ALDA [1]	Ours
A→W	68.4	82.0	86.2	85.4	90.0	95.6	**96.4**
D→W	96.7	96.9	96.2	97.4	97.4	**97.7**	97.5
W→D	99.3	99.1	98.4	99.8	99.6	**100.0**	**100.0**
A→D	68.9	79.7	77.8	84.7	87.8	94.0	**96.4**
D→A	62.5	68.2	69.5	68.6	70.3	**72.2**	72.1
W→A	60.7	67.4	68.9	70.0	66.4	72.5	**73.8**
Avg	76.1	82.2	82.9	84.3	85.2	88.7	**89.4**

5. 结论

本文提出一个基于特定批量归一化的对抗域适应模型，基于域对抗性学习和自我训练的优势。使用噪声校正域判别来学习混淆矩阵。然后利用校正后的损失函数对目标分类器进行优化。在此基础上，将批量归一化层替换为域特定批量归一化层。具有批归一化层的单独分支，每个域分配一个，同时跨域共享所有其他参数。通过同时考虑类别对齐和域对齐来学习域不变特征和域区分特征。实验结果报告了与其他方法相比显著改进的结果。

基金项目

北京市自然科学基金(No. 8202013)；2022 年北京建筑大学研究生创新项目(NO. PG2022145)。

参考文献

[1] Chen, M., Zhao, S., Liu, H., *et al*. (2020) Adversarial-Learned Loss for Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 3521-3528. https://doi.org/10.1609/aaai.v34i04.5757

[2] He, K., Zhang, X., Ren, S., *et al*. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, USA, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[3] Ganin, Y. and Lempitsky V. (2015) Unsupervised Domain Adaptation by Backpropagation. https://arxiv.org/abs/1409.7495

[4] Tzeng, E., Hoffman, J., Saenko, K., *et al*. (2016) Adversarial Discriminative Domain Adaptation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, USA, 21-26 July 2017, 7167-7176. https://doi.org/10.1109/CVPR.2017.316

[5] Pei, Z.Y., Cao Z.J., Long, M.S. and Wang, J.M. (2018) Multi-Adversarial Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 3934-3941. https://doi.org/10.1609/aaai.v32i1.11767

[6] Long, M., Zhu, H., Wang, J., *et al*. (2017) Deep Transfer Learning with Joint Adaptation Networks. https://arxiv.org/abs/1605.06636

[7] Venkateswara, H., Eusebio, J., Chakraborty, S., *et al*. (2017) Deep Hashing Network for Unsupervised Domain Adaptation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, USA, 21-26 July 2017, 5018-5027. https://doi.org/10.1109/CVPR.2017.572

[8] Saenko, K., Kulis, B., Fritz, M., *et al*. (2010) Adapting Visual Category Models to New Domains. In: Daniilidis, K., Maragos, P. and Paragios, N., eds., *Computer Vision—ECCV* 2010, Springer, Berlin, Heidelberg, 213-226. https://doi.org/10.1007/978-3-642-15561-1_16