

DOI:10.11918/202305016

# 伪标签校正和优化的跨场景带钢缺陷识别

刘坤<sup>1</sup>, 王静凯<sup>1</sup>, 毛经坤<sup>2</sup>, 刘卫朋<sup>1</sup>, 陈海永<sup>1</sup>

(1. 河北工业大学 人工智能与数据科学学院, 天津 300130;

2. 天津理工大学 电气工程与自动化学院, 天津 300382)

**摘要:** 由于摄像机类型、参数和环境光照等因素影响,在跨场景情况下采集的带钢图像数据分布存在较大差异,基于深度学习的缺陷识别模型泛化性能差。为此,提出一种伪标签校正和优化的带钢缺陷识别域适应模型(pseudo label correction and optimization domain adaptation, PLCODA)。首先,设计一种基于最大熵和亮度约束的Retinex图像增强模块,用于生成一个与源域标签信息一致且不同于两域数据分布的中间域。其次,设计一个双预测对抗耦合模型,通过源域和中间域分别与目标域的对抗学习,实现对目标域样本的初始伪标签生成。最后,设计一种伪标签校正与迭代提纯策略,通过改进噪声矩阵对伪标签校正,并利用高置信度预测的迭代提纯、低置信度预测的自我惩罚和设计的标签差异度量函数减小伪标签与真值标签的差距。在邯郸钢铁集团和公开的谢维尔钢铁集团的带钢缺陷数据集上进行了验证,实验结果表明,提出的方法针对跨场景缺陷识别问题优于现有的域适应方法。

**关键词:** 无监督;域适应;跨场景;带钢表面缺陷识别;伪标签

中图分类号: TP391

文献标志码: A

文章编号: 0367-6234(2025)10-0135-08

## Cross-scene strip defect recognition based on pseudo-label correction and optimization

LIU Kun<sup>1</sup>, WANG Jingkai<sup>1</sup>, MAO Jingkun<sup>2</sup>, LIU Weipeng<sup>1</sup>, CHEN Haiyong<sup>1</sup>

(1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300130, China;

2. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300382, China)

**Abstract:** The data distributions of cross-scenario images of strip steel defects vary considerably due to imaging factors such as camera type, parameters, and environmental illumination, resulting in poor generalization performance of defect recognition models based on deep learning. To address this issue, we propose a pseudo label correction and optimization domain adaptation (PLCODA) model for strip steel defect recognition. Firstly, a Retinex image enhancement module based on maximum entropy and brightness constraints was designed to generate an intermediate domain that is consistent with the label information in the source domain while different from the data distribution in the two domains. Second, we develop a dual-prediction adversarial coupling architecture that performs adversarial learning between the target domain and each of the source and intermediate domains to generate initial pseudo-labels for target-domain samples. Finally, we propose a pseudo-label correction and iterative purification strategy: we correct pseudo-labels via an improved noise matrix, then iteratively purify them by reinforcing high-confidence predictions, self-punishing low-confidence predictions, and reducing the discrepancy between pseudo-labels and ground-truth labels using a designed label-difference metric. The method was validated on steel-strip defect datasets from Handan Iron & Steel Group and the publicly Severstal Steel Defect Dataset. Experimental results show that the proposed method is superior to the existing domain adaptation methods for cross-scenario defect recognition.

**Keywords:** unsupervised; domain adaptation; cross scene; identification of strip surface defects; pseudo label

带钢作为生产生活中必不可少的材料,在航空航天、土地建筑、汽车制造等领域备受关注。带钢在生产过程中,由于生产工艺或环境等因素的限制,表

面容易出现裂纹、划痕等缺陷,这些缺陷不仅影响产品的外观<sup>[1]</sup>,而且会使产品更容易锈蚀和破裂。视觉检测技术是带钢表面质量检测的常用手段之一,

收稿日期: 2023-05-06; 录用日期: 2023-06-30; 网络首发日期: 2025-09-02

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250902.0835.002>

基金项目: 国家自然科学基金(62173124); 京津冀基础研究合作专项(F2024202102); 河北省基础研究重大项目(A2023202049)

作者简介: 刘坤(1980—),女,教授,博士生导师; 刘卫朋(1979—),男,教授,博士生导师; 陈海永(1980—),男,教授,博士生导师

通信作者: 毛经坤, jingkun@email.tjut.edu.cn

包括基于特征工程的传统缺陷识别方法和基于深度学习的缺陷识别方法。传统的缺陷识别方法主要依赖人工手动设计的视觉特征耗时、耗力,包括光谱法、统计法、模型法等。基于深度学习的缺陷识别方法在封闭场景中取得了良好的效果,但是需要大规模高质量的标注数据,在跨场景下难以获得充足的标注数据。

域适应是一种背景条件约束更加严格的迁移学习问题。以域适应为代表的迁移学习方法通过对齐源域和目标域的数据分布,减小源域与目标域的数据分布差异,学习具有领域不变的特征表示,实现领域不变知识的跨域迁移。Zhang 等<sup>[2]</sup>提出了一种基于差异的域适应框架,通过共享特征提取层学习两域共享特征和判别层匹配条件分布,实现数据分布对齐。Ganin 等<sup>[3]</sup>提出了一种对抗域适应模型,通过提出的梯度反转层使两个域的特征分布相似,实现提取域不变特征。Ghifary 等<sup>[4]</sup>提出了一种基于重构的域适应模型,通过有监督的源域标签预测和无监督的目标数据重构,实现对齐两域数据分布。Liu 等<sup>[5]</sup>提出了一种域适应方法,通过聚合每个域特征空间中的 3 个尺度特征实现对道路裂缝的检测。现有的域适应方法大多将从源域学到的知识直接应用于目标域,以实现跨场景的带钢表面缺陷识别,但两域之间较大的域间隙导致难以对齐两域数据分布,降低了模型的泛化性能。

最近,使用基于伪标签的域适应方法取得了一定的进展。伪标签方法利用源域数据所训练的模型在目标域数据上进行预测,根据预测结果对目标域样本进行标注以获得目标样本的标签。Ham 等<sup>[6]</sup>提出了一种新的伪标签域适应框架,通过提出伪标签修剪策略比较模型输出识别混淆的样本并对齐特征分布,有效缓解标签噪声问题。Jiang 等<sup>[7]</sup>提出了一种基于采样的隐式对齐域适应方法,其中,目标域样本选择过程由伪标签隐式指导,以获得域不变特征。Li 等<sup>[8]</sup>提出了一种领域自适应语义集中模型,目标域样本通过伪标签结果的成对对抗性来对齐最主要的特征。现有使用伪标签的域适应方法根据模型的预测对目标样本标注,再将目标样本及标签输入到模型中训练,但样本数量的不平衡导致样本数量少的类容易产生错误的伪标签进行迭代,从而进一步降低模型的泛化性能。

为了应对现有域适应模型泛化性能差的问题,设计了一种基于最大熵和亮度约束的 Retinex 图像增强模块,用于生成一个数据分布接近目标域的中间域,该中间域和源域分别与目标域之间对抗学习以构建双预测对抗耦合模型。该域作为源域和目标

域之间的桥接,有效缓解两域数据分布差异并增强两域之间的数据分布对齐。为了应对错误伪标签导致模型泛化性能差的问题,设计了一种伪标签校正与迭代提纯策略,用于提升目标域样本标签的置信度。初始伪标签通过改进的噪声矩阵进行校正,目标域样本及校正伪标签在双模型中相互迭代训练,使一个模型使用高置信度预测教授另一个模型,降低样本不足类错误标签的生成,实现高置信度预测的迭代提纯和低置信度预测的自我惩罚,并利用设计的标签差异度量函数共同减小伪标签与真值标签的差距。总之,提出了一种伪标签校正和优化的带钢缺陷识别域适应方法(PLCODA),有效缓解两域的数据分布差异并对齐了两域数据分布,有效提升了模型的泛化能力。

## 1 本文方法

本文提出的 PLCODA 模型如图 1 所示,由图像增强模块  $M$ ,特征提取器  $F_1$ 、 $F_2$ ,分类器  $C_1$ 、 $C_2$  和域鉴别器  $D$  组成。首先,针对给定的源域数据  $D_s = \{X_i, Y\}_{i=1}^{n_s}$  和目标域数据  $D_t = \{X_i\}_{i=1}^{n_t}$ ,通过本文设计的基于最大熵和亮度约束的 Retinex 图像增强模块,生成一个亮度分布不同于两域数据且与源域标签信息一致的中间域。其次,设计了一个内容耦合的双预测对抗模型,中间域和源域分别与目标域对抗学习,以获得域不变表示和初始目标域样本伪标签。最后,双模型利用伪标签校正策略对初始伪标签进行校正,缓解初始伪标签中的混淆信息。再利用伪标签迭代提纯策略和设计的标签差异度量函数,实现高置信度预测的迭代提纯和低置信度预测的自我惩罚,提升目标样本伪标签的置信度。

### 1.1 基于最大熵和亮度约束的 Retinex 图像增强模块

Ben-David 等<sup>[9]</sup>认为模型在目标域上的泛化误差能够衡量模型在目标域中的性能。源域泛化误差和领域间差异是影响目标域任务性能的重要因素,其中,源域泛化误差表示模型在源域中学习知识的质量,领域间差异表示源域和目标域数据分布之间的相似性。针对两域的数据分布差异,域适应方法通过对齐两域数据分布,降低两域的域间差异,从而获得更小的目标域泛化误差。域间差异的数学表示如下:

$$d_{H\Delta H}(S, T) = \sup_{h, h' \in H} \|E_{x \sim P(x^s)} [h(x) \neq h'(x)] - E_{x \sim P(x^t)} [h(x) \neq h'(x)]\| \quad (1)$$

式中: $h(x)$ 表示分类器, $P(x^s)$ 和 $P(x^t)$ 分别表示源域和目标域数据分布。

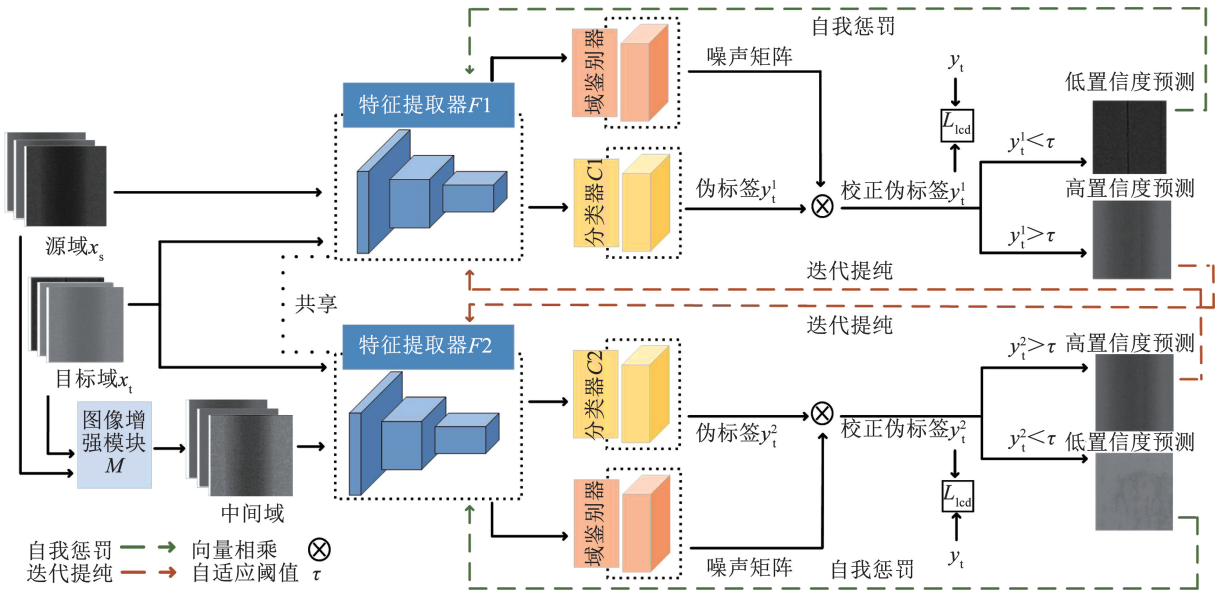


图 1 本文提出的 PLCODA 模型结构

Fig. 1 PLCODA model structure diagram proposed in this article

源域和目标域的域差异越小,模型的泛化性能越好。域间差异和分类器  $h(x)$  的误差相关,分类器的分类效果越好,则域间差异越小及两个域的分布越相似。本文设计了一种基于信息熵理论和亮度约束的 Retinex 模型,旨在生成一个数据分布不同于两域数据且与源域标签信息一致的中间域,该域和源域分别与目标域之间对抗学习以构建双预测对抗耦合模型。双模型利用提出的伪标签校正和优化策略,使一个模型生成的高置信预测伪标签教授另一个模型学习,目标样本及伪标签在两个模型的相互迭代训练可以降低样本不足类错误标签的生成,并持续生成高置信预测伪标签,提高分类器的分类效

果,使域间差异最小化。中间域起到源域和目标域数据的桥接作用,有助于两域在特征空间中匹配更多相似特征,实现源域到目标域的逐步适应并对齐数据分布。

本模块针对图像数据施加亮度信息约束来增加先验知识的前置过程,以此减少深度学习网络训练中学习到的无用信息特征对处理图像任务的模型性能的影响。本模块再引入一个约束条件,即反射率的最大通道与照明图的最大通道一致且其信息熵最大,使增强后的图像有足够的信息与源域标签信息保持一致。本文提出的图像增强模块由分解网络、光照调节网络和重建网络构成,整体结构如图 2 所示。

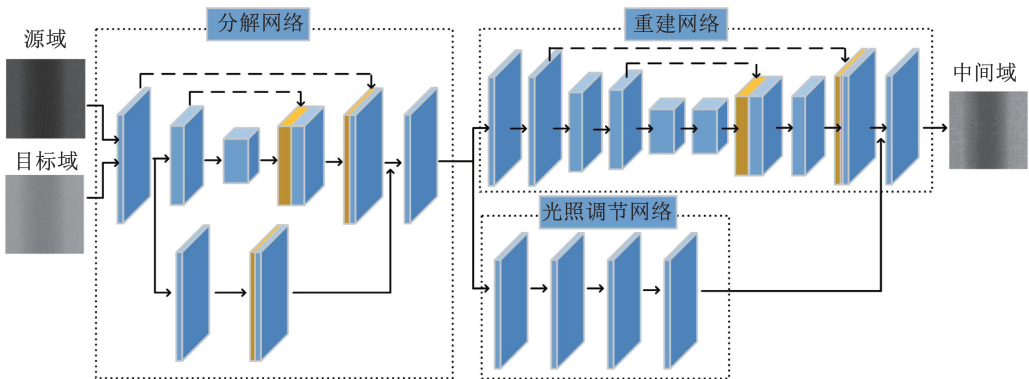


图 2 基于最大熵和亮度约束的 Retinex 模型结构

Fig. 2 Retinex model structure diagram based on maximum entropy and brightness constraints

本模块在分解网络以 Retinex 方式分解图像为两个分量,既反射率  $R$  和照明图  $I$ ,具体为

$$S = R \circ I \quad (2)$$

图像增强问题可以转换为 3 个距离项加亮度约束的形式,具体为

$$\min_{R,S} L_{\text{recon}} + \lambda_1 L_R + \lambda_2 l_1 + \lambda_3 l_B \quad (3)$$

式中:  $L_{\text{recon}}$  表示重建损失,  $L_R$  表示反射损失,  $l_1$  表示光照损失,  $l_B$  表示中间亮度值,  $\lambda_1$ 、 $\lambda_2$  和  $\lambda_3$  为权重参数。

在光照调节网络提取源域和目标域图像的像素亮度信息,并利用亮度约束函数得到中间亮度信息

$l_B$ , 作为源域图像在重建时的增强信息。亮度约束函数作为先验知识, 使一种光照条件转换成另一种, 具体为

$$l_B = \frac{2 \sum_{i=1}^i R_i * 0.299 + G_i * 0.587 + B_i * 0.114}{i} \quad (4)$$

式中:  $R, G, B$  为图像的 3 个通道,  $i$  为数据集样本的总数。在重建网络中采用  $L_1$  范数约束所有损失, 对此最小化重建损失  $l_{\text{recon}}$ 。为了保证增强图像有足够的信息, 用直方图均衡化的结果作参照可以提高图像的信息熵, 使反射率的最大通道与图像的最大通道一致并具有最大信息熵。

### 1.2 伪标签校正策略

通过域鉴别器对抗学习构造了噪声矩阵, 以减少伪标签与真值标签的差异并对齐特征分布。域鉴别器  $D$  以深度特征  $F(x)$  为输入, 输出得到一个多分类的向量  $D(F(x)) \in \mathbf{R}^K$ 。经过一个 sigmoid 层后, 鉴别器产生噪声向量  $\varepsilon(x)$ 。利用鉴别器产生的噪声向量构造噪声校正矩阵  $\eta_{kl}^{(x_i)}$ 。生成的伪标签与噪声校正矩阵相乘后, 伪标签向量转化为校正之后的标签向量作为目标域数据训练时的标签。

$$\eta_{kl}^{(x_i)} = \frac{1 - \varepsilon_l^{(x_i)}}{K} \quad (5)$$

式中:  $k, l$  表示噪声校正矩阵第  $(k, l)$  个分量,  $K$  表示类别。

### 1.3 基于双模型预测的伪标签迭代提纯策略

为了进一步提升目标域样本伪标签的置信度, 设计了基于双模型的伪标签迭代提纯策略。目标域样本及伪标签通过在双模型中相互循环训练, 实现持续生成高置信度预测伪标签并提高目标样本分类的性能。设置了一个自适应阈值  $\tau$ , 当输出的类别概率在阈值  $\tau$  之上时称为高置信度预测, 而类别概率低于阈值  $\tau$  时称为低置信度预测。当特征提取器  $F1$  模型的校正伪标签为高置信度预测时, 该目标样本与标签输入到特征提取器  $F2$  的模型中参与训练。反之同理, 特征提取器  $F2$  模型中高置信预测伪标签的目标样本也输入到特征提取器  $F1$  的模型中训练。而低置信度预测目标样本在原有模型中迭代训练。实现一个模型使用高置信度预测教授另一个模型, 或使用低置信度预测进行自我惩罚。

### 1.4 设计的损失函数

#### 1.4.1 图像增强损失函数 $L_{\text{img}}$

本文设计的图像增强损失函数  $L_{\text{img}}$  由式(3)3 个距离项和 1 个约束函数组成, 具体为

$$L_{\text{img}} = \min_{R,S} l_{\text{recon}} + \lambda_1 l_R + \lambda_2 l_1 + \lambda_3 l_B \quad (6)$$

对于重建损失  $l_{\text{recon}}$ , 采用  $L_1$  范数约束可以表示为

$$l_{\text{recon}} = \|S - R \circ I\|_1 \quad (7)$$

对于反射损失, 与现有的仅使用  $L_1$  范数的方法不同。对于图像增强任务, 处理后的图像应具有足够的信息, 且处理后的图像应与原始图像信息一致, 而直方图均衡化可以提高图像的信息熵。基于上述考虑反射损失为

$$l_R = \left\| \max_{c \in R, G, B} R^c - F\left(\max_{c \in R, G, B} S^c\right) \right\|_1 + \lambda_4 \|\Delta R\|_1 \quad (8)$$

式中:  $F(x)$  表示图像  $X$  的直方图均衡化算子,  $\lambda_4$  为权重参数。该损失函数是图像增强模块的一个约束即反射率的最大通道与图像的最大通道一致, 且信息熵最大。

对于光照损失, 采用文献[10]中提出的结构感知平滑度损失, 即

$$l_1 = \|\Delta I \circ \exp(-\lambda_5 \Delta R)\|_1 \quad (9)$$

在图像增强模块中, 还加入中间亮度值的约束, 即

$$l_B = \frac{2 \sum_{i=1}^i R_i * 0.299 + G_i * 0.587 + B_i * 0.114}{i} \quad (10)$$

由式(7) ~ (10) 得到了图像增强的损失函数  $L_{\text{img}}$ , 即

$$L_{\text{img}} = \|S - R \circ I\|_1 + \lambda_1 \left\| \max_{c \in R, G, B} R^c - F\left(\max_{c \in R, G, B} S^c\right) \right\|_1 + \lambda_4 \|\Delta R\|_1 + \lambda_2 \|\Delta I \circ \exp(-\lambda_5 \Delta R)\|_1 + \lambda_3 \frac{2 \sum_{i=1}^i R_i * 0.299 + G_i * 0.587 + B_i * 0.114}{i} \quad (11)$$

#### 1.4.2 对抗损失函数 $L_{\text{adv}}$

在本文中由噪声校正域鉴别器为中间域、源域和目标域生成不同的噪声向量, 之后构建噪声校正矩阵, 与伪标签相乘后得到校正伪标签。为了得到高置信度预测的伪标签, 设计了一种标签确定性差异(LCD)度量函数。假设  $C_1$  是源域样本生成的伪标签,  $y_s$  是源域真值标签。使用相关矩阵  $A$  来研究标签差异, 即

$$A = C_1 y_s^T \quad (12)$$

式中:  $A$  是大小为  $K \times K$  的方阵。  $A$  可以有效评估标签预测的相关性。利用  $A$  的特性, 使两个标签保持一致性和相关性, 即最大化  $A$  的对角元素。

对于源域特征  $F(x_s)$ , 旨在最小化生成的源域标签向量  $c^{(x_s)}$  和源域真值标签  $y_s$  之间的差异。源域数据的对抗损失

$$L_{\text{lcd}}(x_s, y_s) = \sum_{m,n=1}^K A_{mn} - \sum_{m=1}^K A_{mm} = \sum_{m \neq n}^K A_{mn} \quad (13)$$

对于目标域特征  $F(x_t)$ , 旨在最小化生成的源域标签向量  $c^{(x_s)}$  和目标域标签向量  $c^{(x_t)}$  之间差异。目标域数据的对抗损失

$$L_{\text{led}}(x_t) = \sum_{m,n=1}^K A_{mn} - \sum_{m=1}^K A_{mn} = \sum_{m \neq n}^K A_{mn} \quad (14)$$

对抗损失函数如下:

$$L_{\text{adv}}(x_s, y_s, x_t) = L_{\text{led}}(x_s, y_s) + L_{\text{led}}(x_t) \quad (15)$$

### 1.4.3 分类损失函数 $L_{\text{cls}}$

为了将分类任务嵌入到训练中,在噪声校正域鉴别器中添加了一个正则化项,即

$$L_{\text{reg}}(x_s, y_s) = L_{\text{CE}}(P_D^{(x_s)}, y_s) \quad (16)$$

式中:  $P_D^{(x_s)} = \text{softmax}(D(F(x_s)))$ ,  $L_{\text{CE}}$  为交叉熵损失。

使用高置信度预测进行迭代提纯,使一个模型可以学习另一个模型。循环迭代的定义如下:

$$L_{\text{bim}} = \frac{\sum_{i=1}^B \max(p(c^{x_i}) > \tau) \hat{y}_i \log(q(y | x_i))}{B} \quad (17)$$

式中:  $B$  为最小批量,  $p$  和  $q$  为两个模型的分布概率。

使用低置信度预测进行自我惩罚。自我惩罚的定义如下:

$$L_{\text{sp}} = \frac{\sum_{i=1}^B \max(p(c^{x_i}) < \tau) \hat{y}_i \log(1 - p(y | x_i))}{B} \quad (18)$$

分类损失函数如下:

$$L_{\text{cls}} = L_{\text{reg}} + L_{\text{bim}} + L_{\text{sp}} \quad (19)$$

综上,本文设计的整体损失函数如下:

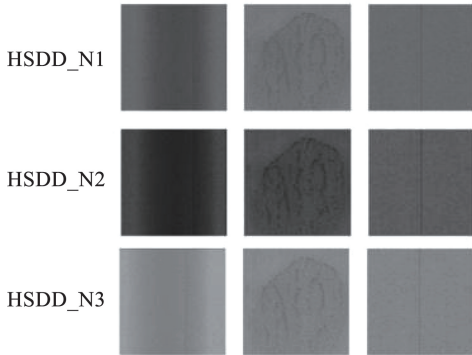
$$L_{\text{all}} = L_{\text{img}} + L_{\text{adv}} + L_{\text{cls}} \quad (20)$$

## 2 实验验证

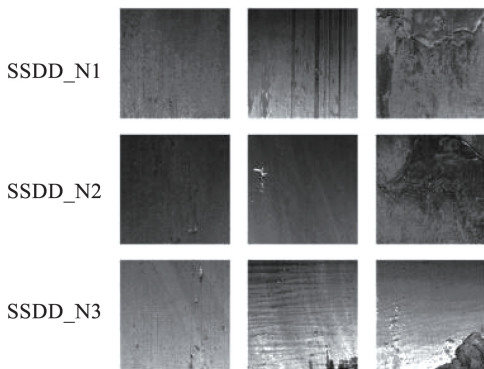
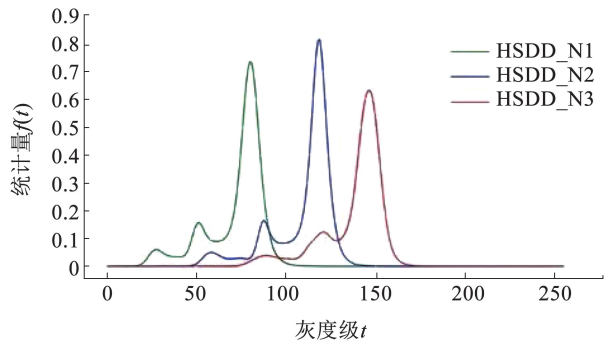
### 2.1 数据集

邯钢钢铁缺陷数据集 (HSDD) 是邯钢钢铁集团收集的带钢缺陷数据集。HSDD 共有 11 814 张图像,  $240 \times 240$  图像分辨率, 有 6 类缺陷, 包括渐变、划痕、油污、边缘、网纹和油点。将 HSDD 按照灰度高斯分布分为 3 个域 HSDD\_N1、HSDD\_N2 和 HSDD\_N3。

谢维尔带钢缺陷数据集 (SSDD) 是谢维尔钢铁集团公开的带钢缺陷数据集, 其光照分布比其他带钢缺陷更为复杂。SSDD 共有 12 568 张图像, 图像分辨率为  $256 \times 256$ 。有 3 类缺陷, 包括凹坑、划痕和补丁, 根据图像灰度的高斯分布, 将 SSDD 分为 3 个域 SSDD\_N1、SSDD\_N2 和 SSDD\_N3。HSDD 和 SSDD 亮度子数据集的示例情况及光照数据分布如图 3 所示。



(a) 邯钢钢铁集团带钢缺陷数据集及光照分布直方图



(b) 谢维尔钢铁集团带钢缺陷数据集及光照分布直方图

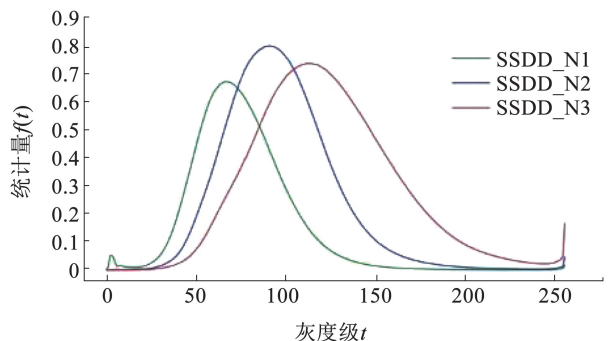


图 3 不同光照场景的带钢缺陷数据集及光照分布

Fig. 3 Strip steel defect datasets and light distribution under different lighting scenarios

## 2.2 实验环境与参数设置

本实验采用的计算机硬件配置为 I7-11700K, NVIDIA GeForce RTX 3090, 操作系统为 Win10。网络模型使用 PyTorch 框架搭建, epoch 设置为 100, batch size 设置为 64。使用 SGD 优化器对模型进行优化, 初始学习率为 0.000 3, 动量为 0.9, 衰减系数为 0.000 5。损失函数的参数分别设置为  $\lambda_1 = 0.1$ 、 $\lambda_2 = 0.1$ 、 $\lambda_3 = 0.6$ 、 $\lambda_4 = 0.01$ 、 $\lambda_5 = 10$ , 使用 ResNet50 作为特征提取器, 加载在 ImageNet 上的预训练分类权重作为初始权重。经验证,  $\lambda_3 = 0.6$  时, 模型性能最好。

## 2.3 评价指标

为了对提出方法的性能进行统计性能评估, 采用精确率评估不同方法在数据集上的有效性, 即

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (21)$$

式中: Accuracy 表示精确率, TP 表示实际阳性且预

测为阳性的样本数, TN 表示实际为阴性且预测为阴性的样本数, FP 表示实际为阴性且预测为阳性的样本数, FN 表示实际为阳性且预测为阴性的样本数。

## 3 实验结果分析

### 3.1 统计性能分析

为了验证本文所提方法的有效性, 在 HSDD 和 SSDD 上进行了实验, 并将所提方法与基于 Resnet50<sup>[11]</sup> 的直推方法和现有的域适应方法作比较, 所对比的域适应方法包括基于 MK-MMD 的 DAN<sup>[12]</sup> 方法、基于领域鉴别器的 DANN<sup>[13]</sup> 方法、基于伪标签训练的 ADLA<sup>[14]</sup> 方法、基于双分类器的 BCDM<sup>[15]</sup> 方法和去噪分类差异 D-MCD<sup>[16]</sup> 方法。

在不同光照条件下 HSDD\_N1, HSDD\_N2, HSDD\_N3 数据集上进行的实验结果如表 1 所示。HSDD\_N1→HSDD\_N3 表示在 HSDD\_N1 上训练的模型在 HSDD\_N3 上测试。

表 1 本文方法与现有方法在 HSDD 上的统计性能对比

Tab. 1 Statistical performance comparison between our method and existing methods on HSDD

| 模型                       | HSDD_N1→ | HSDD_N1→ | HSDD_N2→ | HSDD_N2→ | HSDD_N3→ | HSDD_N3→ | AVG    |
|--------------------------|----------|----------|----------|----------|----------|----------|--------|
|                          | HSDD_N2  | HSDD_N3  | HSDD_N1  | HSDD_N3  | HSDD_N1  | HSDD_N2  |        |
| ResNet50 <sup>[11]</sup> | 47.533   | 19.682   | 51.100   | 83.170   | 41.610   | 86.320   | 54.902 |
| DAN <sup>[12]</sup>      | 88.685   | 63.210   | 87.510   | 90.963   | 80.246   | 97.223   | 84.639 |
| DANN <sup>[13]</sup>     | 92.690   | 92.233   | 90.352   | 95.167   | 94.216   | 91.562   | 92.703 |
| BCDM <sup>[15]</sup>     | 94.813   | 96.621   | 91.245   | 97.312   | 89.675   | 99.246   | 94.818 |
| ADLA <sup>[14]</sup>     | 93.566   | 97.524   | 93.628   | 97.719   | 92.527   | 98.172   | 95.523 |
| D-MCD <sup>[16]</sup>    | 97.888   | 95.669   | 94.937   | 96.382   | 95.776   | 97.781   | 96.315 |
| OURS                     | 99.410   | 99.240   | 98.120   | 99.470   | 99.220   | 99.680   | 99.190 |

由表 1 可以看出, 基于 Resnet50 的直推方法的平均精确率仅为 54.902%。DANN 方法把具有不同分布的源域和目标域中的数据映射到同一个特征空间, 进而对齐两域数据的特征分布, 在 HSDD\_N1→HSDD\_N3 上较 DAN 方法提高了 29.023%。BCDM 方法使用两个分类器来进行对抗学习, 在 HSDD\_N3→HSDD\_N2 上较 DANN 方法提高了 7.684%。ADLA 和 D-MCD 方法与本文类似结合伪

标签学习判别特征, 在不同数据集上获得的平均精确率达 95.523% 和 96.315%。本文提出的方法平均精确率达 99.190%, 验证了本文方法针对跨场景缺陷识别问题优于现有的域适应方法。

本文方法与现有域适应方法在光照分布更为复杂的 SSDD\_N1、SSDD\_N2 和 SSDD\_N3 数据集上的统计性能对比结果如表 2 所示。

表 2 本文方法与现有方法在 SSDD 上的统计性能对比

Tab. 2 Statistical performance comparison between our method and existing methods on SSDD

| 模型                       | SSDD_N1→ | SSDD_N1→ | SSDD_N2→ | SSDD_N2→ | SSDD_N3→ | SSDD_N3→ | ANG    |
|--------------------------|----------|----------|----------|----------|----------|----------|--------|
|                          | SSDD_N2  | SSDD_N3  | SSDD_N1  | SSDD_N3  | SSDD_N1  | SSDD_N2  |        |
| ResNet50 <sup>[11]</sup> | 76.556   | 63.143   | 81.286   | 80.256   | 64.938   | 79.716   | 74.315 |
| DAN <sup>[12]</sup>      | 84.026   | 64.920   | 83.664   | 85.168   | 70.623   | 83.562   | 78.660 |
| DANN <sup>[13]</sup>     | 81.937   | 73.624   | 89.568   | 83.296   | 74.668   | 84.725   | 81.303 |
| BCDM <sup>[15]</sup>     | 86.569   | 76.196   | 87.365   | 93.523   | 75.235   | 85.457   | 84.057 |
| ADLA <sup>[14]</sup>     | 87.023   | 81.598   | 88.563   | 94.248   | 78.556   | 87.963   | 86.325 |
| D-MCD <sup>[16]</sup>    | 86.854   | 80.764   | 90.286   | 94.481   | 77.938   | 87.561   | 86.481 |
| OURS                     | 92.850   | 88.290   | 92.680   | 94.510   | 86.970   | 92.370   | 91.280 |

由表 2 可以看出,本文提出的方法在跨场景下的平均精确率达 91.280%,验证了本文方法针对跨场景缺陷识别问题优于现有的域适应方法。基于 Resnet50 的直推方法的泛化性能较差,平均精确率仅为 74.315%。由于本文方法与伪标签方法相结合来学习判别特征,本文方法比基于对抗学习的方法(DAN、DANN、BCDM)取得了更好的性能。与本文方法类似,ADLA 和 D-MCD 方法也结合伪标签方法。在迁移任务 SSDD\_N1→SSDD\_N3、SSDD\_N2→

SSDD\_N1 和 SSDD\_N3→SSDD\_N2 上,本文方法相比 ALDA 和 D-MCD 方法取得了更优的性能。

### 3.2 特征可视化结果

为了进一步验证提出 PLCODA 方法的有效性,本文利用 T-SNE 方法将模型学习到的目标域高维特征向量映射为二维特征向量,并在二维嵌入空间中对 SSDD\_N1→SSDD\_N2 的特征可视化,结果如图 4 所示。

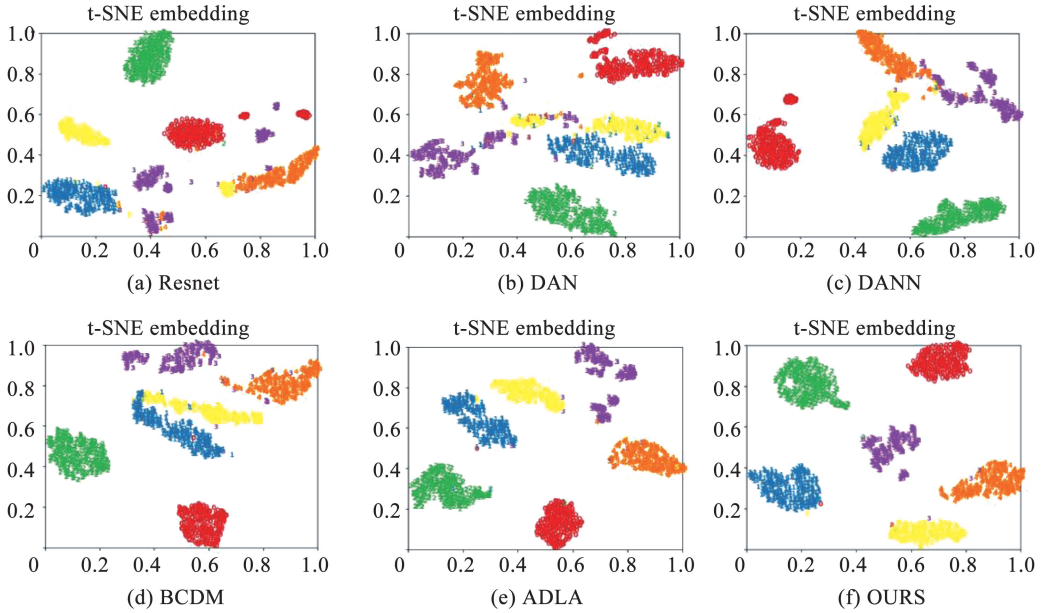


图 4 不同方法针对目标域提取的特征可视化结果

Fig. 4 Different methods visualize the results of the extracted features for the target domain

由图 4 可以看出,基于 Resnet50 的直推方法对各类缺陷提取的特征在嵌入空间中的类间距离较小。针对划痕、油点和油污缺陷,DANN 方法较 DAN 方法在嵌入空间中的类内距离更为紧凑,但边缘和网纹缺陷提取的特征相似度高。BCDM 方法对提取特征的类内间距较 DANN 方法和 DAN 方法更小,但划痕与油点缺陷在嵌入空间中存在远离聚类中心的样本。ADLA 方法较 BCDM 方法拉近了划痕缺陷特征的类内距离,但边缘缺陷提取的特征散度较大。本文提出的方法对目标域所有缺陷特征在嵌入空间中的类内距相比其他方法更为紧凑,类间距更大,验证了所提方法对目标域数据提取的特征具有类别区分性。

### 3.3 消融实验

为了进一步验证图像增强模块、伪标签校正策略和基于双模型的伪标签迭代优化策略的性能,以 SSDD\_N1→SSDD\_N2 为例设计了消融实验,结果如表 3 所示。

表 3 消融实验

Tab. 3 Ablation experiment

| 图像增强模块 | 伪标签校正策略 | 基于双模型的伪标签迭代优化策略 | 性能指标 AVG |
|--------|---------|-----------------|----------|
| —      | —       | —               | 81.937   |
| ✓      | —       | —               | 84.965   |
| ✓      | ✓       | —               | 87.683   |
| ✓      | —       | ✓               | 90.560   |
| ✓      | ✓       | ✓               | 92.850   |

由表 3 可以看出,在 DANN 方法下精确率仅为 81.937%。使用图像增强模块的方法相比没有使用的方法提升了 3.028%,表明图像增强模块能够提升模型泛化性能。使用伪标签校正策略和基于双模型的伪标签迭代优化策略,精确率分别提升了 2.718%和 5.595%。本文方法的精确率达 92.850%。综上,本文提出的图像增强模块和两个伪标签策略能够提升模型的泛化性能。

## 4 结 论

1)为了解决深度学习模型在跨场景的带钢表面缺陷识别中泛化性能差的问题,提出了一种伪标签校正和优化的带钢缺陷识别域适应模型(PLCODA)。

2)设计了基于最大熵和亮度约束的 Retinex 图像增强模块,该模块生成一个数据分布不同于源域和目标域且保留源域标签信息的中间域,该域作为源域和目标域之间的桥接,有效缓解两域数据分布差异。

3)设计了一个双预测对抗耦合模型,中间域和源域分别与目标域进行对抗学习,以获得域不变表示,并针对目标域样本生成初始的伪标签。

4)设计了伪标签校正策略和基于双模型的伪标签迭代提纯策略来优化跨场景的目标域样本伪标签,并通过设计的标签差异度量函数提升伪标签置信度。

5)在邯郸钢铁集团带钢缺陷数据集以及谢维尔钢铁集团带钢缺陷数据集上分别进行了验证,结果表明,本文方法对跨场景的带钢缺陷识别具有更好的泛化性。

## 参考文献

- [1]易梦云. 带钢表面质量缺陷检测与分类方法研究[D]. 武汉: 湖北工业大学, 2019  
YI Mengyun. Research on surface quality defect detection and classification methods for strip steel[D]. Wuhan: Hubei University of Technology, 2019
- [2]ZHANG Xu, YU Xinnan, WANG Shengjin, et al. Deep transfer network: unsupervised domain adaptation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE Computer Society, 2015: 11. DOI:10.48550/arXiv.1503.00591
- [3]GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation [C]// International Conference on Machine Learnings. Stroudsburg, PA, USA: International Machine Learning Society, 2015, 37: 1180. DOI:10.5555/3045118.3045244
- [4]GHIFARY M, KLEIJN W B, ZHANG M, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[C]// European Conference on Computer Vision, Cham. Switzerland: Lecture Notes in Computer Science (LNCS), 2016, 28: 597. DOI:10.1007/978-3-319-46493-0\_36
- [5]LIU Huijun, YANG Chuya, LI Aocheng, et al. Deep domain adaptation for pavement crack detection[J]. IEEE Transactions on

- Intelligent Transportation Systems, 2021, 24: 1669. DOI:10.1109/TITS.2022.3225212
- [6]GYEONGDO H, YUCHEOL C, JAE H L, et al. P-pseudolabel: enhanced pseudo labeling framework with network pruning in semi-supervised learning[J]. IEEE Access, 2022, 10: 115652. DOI: 10.1109/Access.2022.3218161
- [7]JIANG X, LAO Qicheng, MATWIN S, et al. Implicit class-conditioned domain alignment for unsupervised domain adaptation [C]//International Conference on Machine Learning. Vienna, Austria: Proceedings of Machine Learning Research, 2020, 119: 4816
- [8]LI Shuang, XIE Mixue, LV F, et al. Semantic concentration for domain adaptation [C]//IEEE/CVF International Conference on Computer Vision and Pattern Recognition. Stroudsburg: IEEE Computer Society, 2021: 9082. DOI: 10.1109/ICCV48922.2021.00897
- [9]BEN-DAVID S, BLITZER J, CRAMMER K, et al. A theory of learning from different domains[J]. Machine Learning, 2010, 79 (1/2): 151. DOI:10.1007/s10994-009-5152-4
- [10]WEI C, WANG W, YANG W, et al. Deep retinex decomposition for low-light enhancement [C]//British Machine Vision Conference. Durham: BMVA Press, 2018, 55(4): 124. DOI:10.48550/arXiv.1808.04560
- [11]HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE Computer Society, 2016: 770. DOI:10.1109/CVPR.2016.90
- [12]LONG Mingsheng, ZHU Hanhua, WANG Jianmin, et al. Deep transfer learning with joint adaptation networks [C]//International Conference on Machine Learning. Sydney, Australia: Proceedings of Machine Learning Research (PMLR), 2017, 70: 2208. DOI: 10.48550/arXiv.1605.06636
- [13]CHEN Minghao, ZHAO Shuai, LIU Haifeng, et al. Adversarial-learned loss for domain adaptation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Stroudsburg: AAAI Press, 2020, 34(4): 3521. DOI:10.1609/aaai.v34i04.5757
- [14]LI Shuang, LÜ Fangrui, XIE Binhui, et al. Bi-classifier determinacy maximization for unsupervised domain adaptation [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, California: AAAI Press, 2021, 35(10): 1854. DOI:10.1609/aaai.v35i10.17027
- [15]CHU Tong, LIU Yahao, DENG Jinhong, et al. Denoised maximum classifier discrepancy for source free unsupervised domain adaptation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington DC: AAAI Press, 2022: 472. DOI:10.1609/aaai.v36i1.19925
- [16]LAURENS V D M, HINTON E G. Visualizing data using t-SNE [J]. Journal of Machine Learning Research, 2008, 9(2605): 2579

(编辑 刘 彤)