DOI: 10.16725/j.1673-808X.202364

引用本文: 杨爽, 江泽涛. 一种低照度下的无监督域自适应目标检测方法[J]. 桂林电子科技大学学报, 2026, 46(3): 241-251.

Yang Shuang, Jiang Zetao. An unsupervised domain adaptive object detection method for low-light scenarios[J]. Journal of Guilin University of Electronic Technology, 2026, 46(3): 241-251.

一种低照度下的无监督域自适应目标检测方法

杨　爽[1,2]，江泽涛[1,2]

（1. 桂林电子科技大学 计算机与信息安全学院, 广西 桂林　541004;
2. 桂林电子科技大学 广西图像图形与智能处理重点实验室, 广西 桂林　541004）

摘　要: 针对低照度目标检测中数据集匮乏的问题, 提出一种低照度环境下的无监督域自适应目标检测方法。首先, 采用像素级对齐模块对源域图像进行图像翻译, 初步缩小源域和目标域之间的分布差异; 然后, 采用结合噪声注意力的多层特征对齐模块对低照度图像存在大量噪声的问题进行了特定处理, 并在多层网络中设置对抗分类器和梯度反转层, 促进了源域和目标域在全局特征上的进一步对齐; 最后, 提出融合全局特征的实例级特征对齐模块, 通过引入特征提取网络中的全局信息, 弥补实例级特征域泛化性较差的缺陷, 并对图像前景区域的特征进行对齐, 从而实现无监督目标检测。实验结果表明, 所提方法在不使用低照度数据集标注及低照度图片预处理的情况下, 取得了优异的低照度图像目标检测性能。

关键词: 低照度; 目标检测; 域自适应; 对抗训练; 无监督学习; 图像翻译

中图分类号: TP301.6　　**文献标志码:** A　　**文章编号:** 1673-808X(2026)03-0241-11

An unsupervised domain adaptive object detection method for low-light scenarios

YANG Shuang[1,2], JIANG Zetao[1,2]

（1. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China;
2. Guangxi Key Laboratory of Image Graphics and Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China）

Abstract: To address the scarcity of labeled data in low-light object detection, an unsupervised domain-adaptive object detection framework is proposed. First, a pixel-level alignment module is employed for image translation, thereby reducing the distribution discrepancy between the source and target domains. Then, a multi-layer feature alignment module with noise-aware attention is designed to mitigate severe noise in low-light images, where an adversarial classifier and gradient reversal layer(GRL) are incorporated to further align global features across domains. Finally, an instance-level feature alignment module with global context is introduced to improve generalization by enhancing feature consistency in foreground regions. Experimental results show that the proposed method significantly improves detection performance in low-light images without requiring additional annotations or preprocessing.

Key words: low-light; object detection; domain adaptation; adversarial training; unsupervised learning; image translation

目标检测是计算机视觉重要应用之一, 已广泛应用于安全、军事、交通、医疗等领域[1-3]。然而, 低照度环境下的目标检测仍存在诸多挑战。由于光照不足, 捕获的图像存在严重的退化现象, 如低可见度, 图像存在大量噪声以及色彩偏差等问题[4]。这些图像退化问题不仅会影响人眼观测, 还会严重影响机器在目

收稿日期: 2023-04-12

基金项目: 国家自然科学基金 (62172118); 广西自然学科基金 (2021GXNSFDA196002); 广西图像图形智能处理重点实验室基金 (GIIP2204); 广西研究生教育创新计划 (YCB2021070, YCBZ2018052, YCSW2022269, 2021YCXS071)

通信作者: 江泽涛 (1961—), 男, 教授, 博士, 研究方向为图像处理、计算机视觉。E-mail: zetaojiang@126.com

标检测任务中的表现，例如在监控视频分析、夜间自动驾驶等应用场景中会带来巨大的潜在风险。

低照度图像的形成主要源于夜间光照不足、相机曝光不足或过曝等，这导致低照度图像存在许多难以辨别的区域和观测障碍区域，因此难以对其进行准确标注，这对构建具有精准标注的高质量低照度目标检测数据集带来了巨大挑战，将会耗费大量的时间和人力。Loh 等[5] 收集了带有实例级注释的低照度数据集 ExDARK，但只包含有限类别，数据集的大小也不足以训练足够优越的检测模型。Sasagawa 等[6] 针对低照度下的 RAW 格式图片，提出了一种基于知识蒸馏的方法，将不同域中预训练得到的模型通过黏合层与生成模型合并在一起，实现了跨域的检测。然而，该方法提出的黏合层需要通过知识蒸馏进行额外训练，模型训练过程较为复杂，且该方法不能直接用于 RGB 格式的图片。Guo 等[7] 提出一种低照度图像增强方法 Zero-DCE，在实验结果中展示了其在低照度人脸检测任务中的有效性，然而这并非完整端到端方案，实际落地场景通用性较差。由此可见，低照度数据集的匮乏为低照度目标检测算法的研究带来了挑战。因此，在缺乏标注的情况下，通过无监督的方法训练低照度下的目标检测模型具有重要意义。

在不使用低照度标注数据的情况下，使用大量有标注的正常照度数据对模型进行训练，在低照度图像上推理，往往会造成检测精度严重下降。这是由于训练集与测试集存在较大差异。这种训练集与测试集之间特征空间分布存在差异的情况通常称为域偏移，近年来迁移学习在解决域偏移问题上取得了一定进展[8]。无监督域自适应技术旨在解决源域和目标域之间的域偏移问题。其中，对抗学习在域自适应方法中扮演重要角色。随着域对抗神经网络（domain adversarial neural network，简称 DANN）[9] 的出现，相关学者开始通过对抗学习的方法对齐不同域间的分布，并在一些下游任务上取得了成功，如跨域的图像分类[10-11] 和语义分割[12-13]。近年来，无监督域自适应的方法已开始逐渐应用于目标检测任务。在无监督域自适应的目标检测算法中，Chen 等[14] 在图像级特征和实例级特征上添加了梯度反转层（gradient reverse layer，简称 GRL），以此来对齐源域和目标域特征。Zhu 等[15] 使用了 *k*-means 算法来获取图像中的关键区域，以进行域对齐。Saito 等[16] 在低级特征上对齐局部感受野，在全局特征上进行弱对齐。另外，Arruda 等[17] 通过图像翻译的方法生成了有标注的伪目标域数据集，以此来进行有监督训练。然而，伪目标域数据和真实目标域数据之间仍存在一定的域偏移。同时，以上这些方法所研究的跨域场景多为正常照度图像或不存在退化的图像，且域间差异较小，例如晴天与雨雾天气获取的图像[18]、不同相机捕获的风格迥异的图像[19]、真实世界图像与绘画图像[20] 等。在这些场景中，目标大多为人眼可见的实例对象，而低照度图像中图像严重退化，存在大部分目标实例不清晰的情况。当这些模型用于低照度图像的跨域检测时，效果往往不如域间差异较小的数据。

针对上述问题，提出了一种低照度场景下的无监督域自适应目标检测方法。与 YOLO 等单阶段目标检测方法[21-25] 相比，双阶段目标检测方法在第二阶段通过裁剪候选框区域，得到了实例级特征。而在域自适应目标检测方法中，为了更高效地对齐特征，需要获取图像中前景区域的特征，而双阶段目标检测方法提供了这一条件，因此，选取 Faster R-CNN[26] 目标检测网络作为基线网络（Baseline）。在本方法中，首先提出像素级对齐（pixel-level alignment，简称 PA）模块，然后提出结合噪声注意力的多层特征对齐（multi-layer feature alignment with noise attention，简称 MFA-NA）模块，最后提出融合全局特征的实例级对齐（instance-level feature alignment incorporating global features，简称 IFA-G）模块，三模块构建无监督域目标检测模型。像素级对齐模块通过图像翻译技术实现了源域到目标域图像风格的迁移，缩小了源域和目标域图像之间的像素级差异；结合噪声注意力的全局特征对齐模块，并借助噪声权重估计图以及多层特征的对抗学习，减弱了低照度图像中噪声对对抗训练的影响，增强了骨干网络所提取全局特征的域泛化能力；融合全局特征的实例级特征对齐模块将实例级特征与经过融合的多层全局特征合并，增加了全局特征信息的引导，通过对抗训练增强了前景区域特征的域泛化能力。

1 低照度下的无监督域自适应目标检测模型

1.1 总体设计思想

通过在 Faster R-CNN 目标检测网络中嵌入像素级对齐模块、结合噪声注意力的多层特征对齐模块、融合全局特征的实例级特征对齐模块，以一种渐进的方式在像素风格、全局特征、实例特征三个层面逐步缩小正常照度图像和低照度图像间的特征分布差异，得到具有域泛化能力的特征，在不使用低照度图像标注的情况下，提升了模型在低照度图像上的检测性能。

首先，正常照度图像通过 PA 模块的处理生成合成低照度数据集，与正常照度数据集相比，合成低照

度数据集的分布更接近真实低照度图像的数据分布, 在图像输入阶段减小了迁移难度。在模型训练过程中, MFA-NA 模块对主干网络中的多层特征进行对抗训练, 由于网络深层特征中不同尺度的特征图包含不同的语义信息, 多层特征对齐更全面地弥合了源域与目标域间的特征分布差异。其次, 在计算全局对齐损失的过程中, 通过估计图像的噪声分布生成噪声权重估计图, 为特征图中噪声较大的区域赋予较小的权重, 抑制低照度图像中噪声严重区域包含的不利信息对全局特征对齐过程的干扰。在模型经过 RoI Pooling[27] 生成实例级特征后, IFA-G 模块融合了实例级特征与主干网络提取的全局特征, 结合上下文语义信息对齐图像前景区域的特征分布, 弥补实例级特征域泛化性较差的缺陷。最后, 对源域中融合后的实例级特征计算检测损失, 使检测网络获得分类和定位的能力。

模型整体结构如图 1 所示, 可分为像素级对齐模块、结合噪声注意力的多层特征对齐模块、融合全局特征的实例级特征对齐模块以及 Faster R-CNN 检测网络四部分。在模型推理过程中, 只输入低照度图像进行检测, 将去除 PA 模块和 MFA-NA 模块, IFA-G 模块不参与实例级对齐损失的计算。

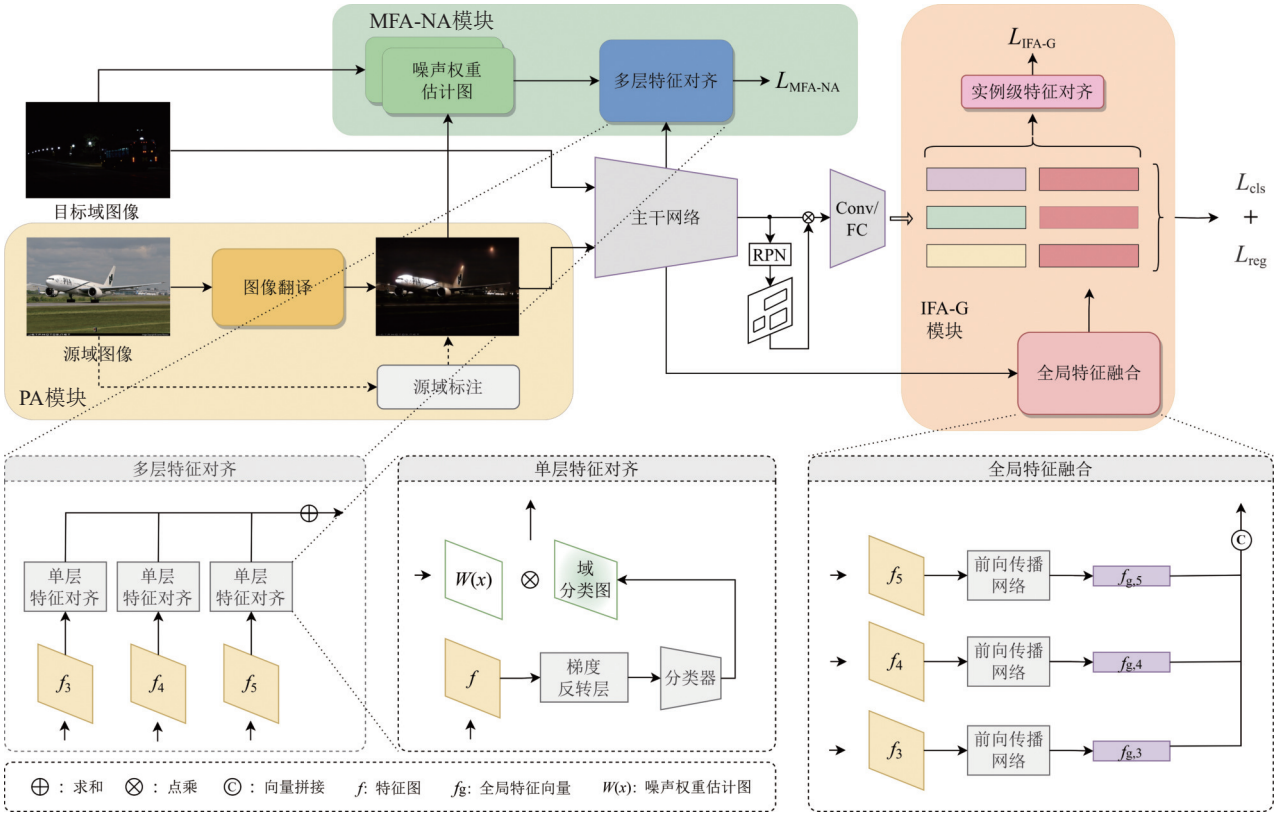


图 1　模型整体结构

1.2　像素级对齐模块

像素级对齐模块使用不成对的图像翻译技术来减小正常照度图像与低照度图像之间的像素级分布差异。由于对抗生成网络(generative adversarial network, 简称 GAN)[28] 在生成数据时, 无需数据的标注信息, 可实现有标注的源域数据与无标注的目标域数据之间的任意转换。像素级对齐模块结构如图 2 所示, 通过训练好的 CycleGAN[29] 模型将正常照度图像翻译为合成低照度图像, 正常照度图像中的地面真实标注信息也同时适用于合成低照度图像, 这样便获得了大量具有精准标注信息的合成低照度数据。

以假乱真的合成低照度图像大量地扩增了低照度图像训练数据, 这使得模型可以学到更多与目标域相近的表征和知识。与源域的正常照度数据集相比, 合成低照度图像在特征空间分布上更接近目标域, 因此减少了源域到目标域之间的迁移难度。为了直观地观察图像翻译前后分布的差异, 使用 t-SNE 算法[30] 将图像映射到二维的低维空间中, 模拟图像在高维空间中的分布。图 3(a) 为 PASCAL VOC[31] 和 ExDARK 数据的特征空间可视化图。将经过 CycleGAN 图像翻译后得到的合成低照度图像数据集表示为 Syn_D, 图 3(b) 为 Syn_D 与 ExDARK 数据集的特征空间可视化图。对比图 3(a) 和图 3(b) 可看出, 与正常照度图

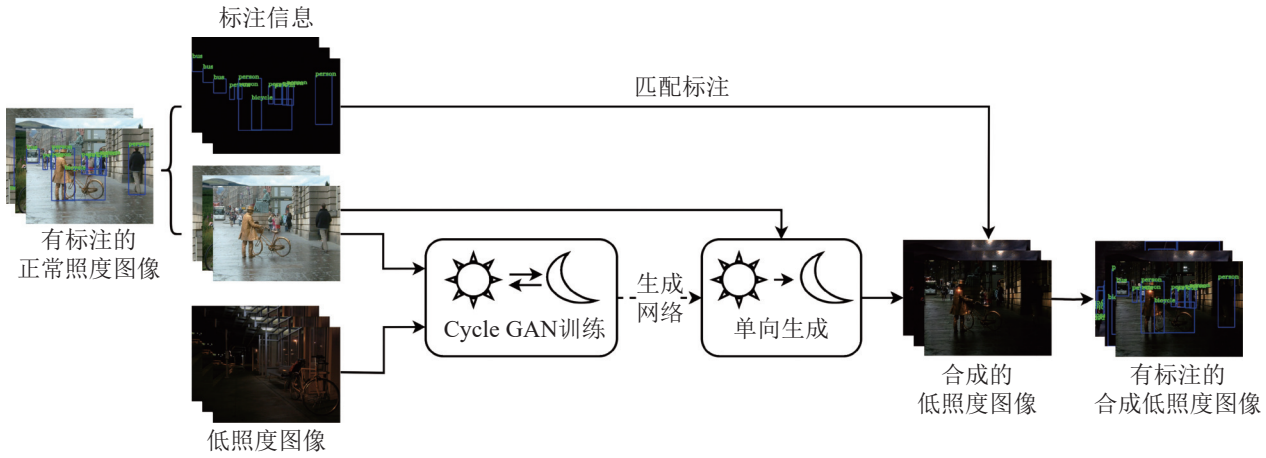
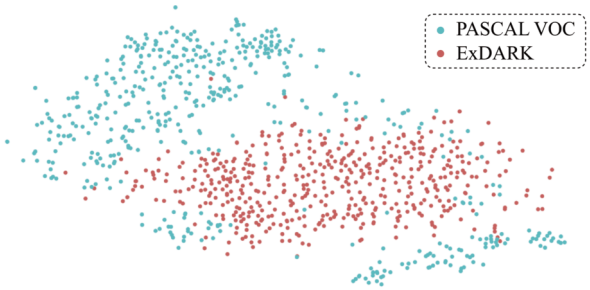


图 2　像素级对齐模块结构

像 PASCAL VOC 相比，合成低照度图像 Syn_D 在特征空间中与真实低照度数据 ExDARK 更接近。由此可知，图像翻译模块有效地减小了域偏移。但从图 3(b) 可看出，经过迁移的源域数据与目标域数据仍存在一定的分布差异，因此需要进一步对齐两者之间的特征分布。



(a) PASCAL VOC和ExDARK



(b) Syn_D和ExDARK

图 3　t-SNE 算法下数据集的特征空间可视化分布图

1.3　结合噪声注意力的多层特征对齐模块

特征提取网络的中层特征和高层特征包含大量图像中的关键特征信息，将这部分特征进行对齐，可显著减小源域和目标域之间的特征偏移。因此，本方法采用骨干网络中层和高层部分的多层特征来进行对齐，在特征提取网络的多层设置域分类器进行对抗训练，从而在多个尺度的特征上弥合源域与目标域之间的偏移，以达到域间差异最小化的目的。假设输入的源域图像、目标域图像分别为 x_s、x_t，骨干网络部分为 B，以 ResNet101 为例，骨干网络由 5 个卷积块组成，即 $B=\{B_1,B_2,B_3,B_4,B_5\}$，其中 B_m 表示第 m 个卷积块。每个卷积块都会输出一个特征图，假设特征图表示为 $F=\{f_1,f_2,f_3,f_4,f_5\}$，其中 f_m 表示网络中第 m 个卷积块输出的特征图，其中 $f_1=B_1(x)$，在 $m\in[2,5]$ 时 $f_m=B_m(f_{m-1})$。结合噪声注意力的多层特征对齐模块结构如图 4 所示，在 f_3、f_4、f_5 三特征图上分别设置对抗域分类器，分类器分别表示为 D_3、D_4、D_5，用来辨别当前特征图中每个元素属于源域还是目标域，预测得到域标签预测图 p_m，其中 $p_m=D_m(B_m(x))$。在实验中，设置源域标签为 1，目标域标签为 0。分类器由全卷积网络构成，卷积核大小为 1，通过计算逐点的域判别损失，得到与当前特征图大小相等的域标签预测图 p_m。假设当前图像域标签表示为 d_i，在第 m 个卷积块输出的特征图 f_m 中，位置在 $(w,\ h)$ 处的元素，其判别器损失函数定义为

$$L_{\text{glb}}=d_i\log D_m(B_m(f_{m-1}))_{w,h}+(1-d_i)\times\log(1-D_m(B_m(f_{m-1}))_{w,h})。$$

通过对抗训练的方式来减小源域与目标域之间的特征分布差异。对抗训练主要通过梯度反转层和域分类器 2 个组件实现，其中梯度反转层在提取域不变特征的过程中起关键作用。域分类器的目的是对当前特征进行分类，分辨当前特征属于源域还是目标域，这一步通过最小化分类损失来实现。梯度反转层在网络中介于分类器和骨干网络之间，只在反向传播时影响梯度计算。在反向传播过程中，梯度值经过梯度反转层后会对梯度取反，梯度逆转后，骨干网络的参数会向相反的方向更新。最终，骨干网络输出的特征企图骗过分类器，使分类器损失最大化。在这样最大化、最小化的博弈中，骨干网络获得了混淆源域和

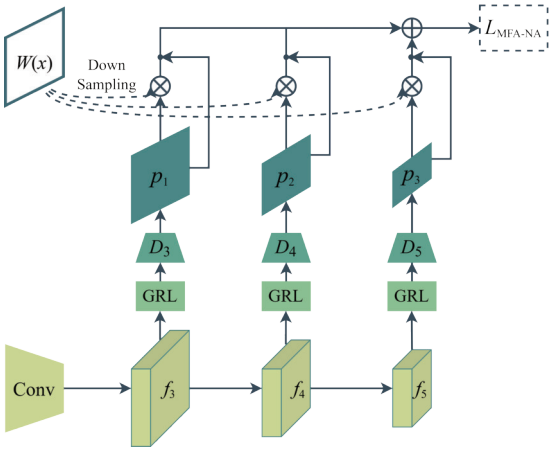


图 4　结合噪声注意力的全局特征对齐模块

目标域特征的能力, 输出的源域和目标域特征获得了一定的域泛化性。假设骨干网络中权重参数为 θ_B, 分类器中权重参数为 θ_D, 网络的优化目标可表示为 $\min\limits_{\theta_D}\max\limits_{\theta_B} L_{glb}$。

低照度图像往往含有大量噪声, 这会对全局特征的对齐产生较大干扰。例如, 源域中的前景区域特征与一个含有大量噪声的目标域图像同时进行对抗训练时, 噪声信息会对包含关键信息的前景特征提取产生干扰。因此, 为了减小噪声严重区域带来的影响, 将估计输入图像中各区域的噪声, 通过噪声估计图计算得到不同区域的权重值, 为噪声较低的区域赋予较高的权重; 反之, 将赋予较低的权重值。噪声的估计采用分块法[32], 将大小为 (M, N) 的图像划分成若干子块, 子块大小为 4×4, 划分后整体子块尺寸为 $(M/4)\times(N/4)$。为每个子块计算噪声值, 将子块灰度值的标准差作为该子块的噪声值, 由此得到噪声估计图 $N(x)$, 噪声越大的区域将被赋予越小的权重, 因此噪声值与权重值应呈反比例关系。为了防止特征过度表达, 该关系通过凹函数计算, 设权重 $W(x)=(3-2\times N(x))/(2\times N(x)+1)$。得到的噪声权重图与域标签预测图尺寸不一致, 因此需要对噪声权重图进行下采样计算, 池化后的噪声权重图与域标签预测图尺寸一致。将权重估计图与域标签估计图进行点乘计算, 并将逐层损失相加, 得出最终对齐损失。由此, 网络将针对具有更多有效信息的区域进行对齐。

但由于该种方式计算产生的噪声估计图并不完全准确, 一些图像的边缘细节部分会被误认为是噪声较大区域。因此, 借鉴残差网络的思想, 通过在网络中增加跳跃连接来增强误判区域的特征表达。基于以上思路, 结合噪声注意力的多层特征对齐模块的损失函数可定义为

$$L_{\text{MFA-NA}}=\sum_{m,h,w}(1+S_m(W(x)_{w,h}))L_{glb},$$

其中, S_m 表示将权重估计图调整为第 m 层特征图尺寸的下采样操作。

1.4　融合全局特征的实例级特征对齐模块

实例级特征只保留了局部特征信息, 缺乏对上下文信息的关注, 实例级对齐过程中通常更多地关注域内信息, 而域泛化性相对较弱; 经过全局对齐的特征具有较强的域泛化性, 而域内特征表达较弱。因此本方法通过融合两者的特征, 弥补各自的不足。图 5 为融合全局特征的实例级特征对齐模块, 特征图 f_3、f_4、f_5 经过前向传播网络得到了该层特征图的特征向量, 将各层的特征向量拼接后, 得到融合后的全局向量 $\boldsymbol{f}_g$, $\boldsymbol{f}_g$ 包含了图像中的全局特征信息。为了使实例级特征得到全局特征的引导, 增强实例级特征的上下文信息, 将得到的全局特征向量 $\boldsymbol{f}_g$ 与实例特征向量 $\boldsymbol{f}_i$ 拼接, 得到融合特征 $\boldsymbol{f}_{fus}$。$\boldsymbol{f}_{fus}$ 结合了前景区域的特征信息和图片的语义信息。特征 $\boldsymbol{f}_{fus}$ 经过梯度反转层和实例级域分类器, 通过对抗训练实现了源域与目标域之间实例级特征对齐。其中, 为了防止前向传播网络在反向传播过程中影响特征提取网络的参数更新, 导致特征提取网络提取到被实例级特征影响的全局特征, 本方法阻断了前向传播网络到骨干网络的反向传播。前向传播网络只需从骨干网络中单向地获取各层全局特征信息, 通过检测损失和实例级对齐损失更新自身参数。阻断梯度的传播会保证得到的 $\boldsymbol{f}_g$ 不受实例级特征干扰, 保留完整的全局特征信息, 保证全局特征的域泛化性。各层的前向传播网络结构如图 6 所示。

为了增强模型对于分类困难样本的关注, 忽略易于分类的样本, 以达到更好的分类性能。IFA-G 模块的损失函数通过 Focal Loss[33] 计算。损失函数由源域实例级对齐损失和目标域实例级对齐损失两部分组成, 源域实例级特征表示为 $\boldsymbol{f}_{fus}$, 目标域实例级特征表示为 $\boldsymbol{f}_{fus}$, 假设第 i 张图像中共检测到 j 个实例, 则 IFA-G 模块的损失函数为

$$L_{\text{IFA-G}}=\sum_{i,j}(1-D_{ins}(f_{fus}^{i,j}))^{\gamma}\log(D_{ins}(f_{fus}^{i,j}))+\sum_{i,j}D_{ins}(f_{fus}^{i,j})^{\gamma}\log(1-D_{ins}(f_{fus}^{i,j}))。$$

1.5　总损失函数

总损失函数分为检测损失和特征对齐损失两部分。检测损失包括分类损失 L_{cls} 和回归损失 L_{reg}, 分别用来评估分类的精确度和预测边界框与标注框之间的重合度, 这两项损失只对源域数据进行计算。特征

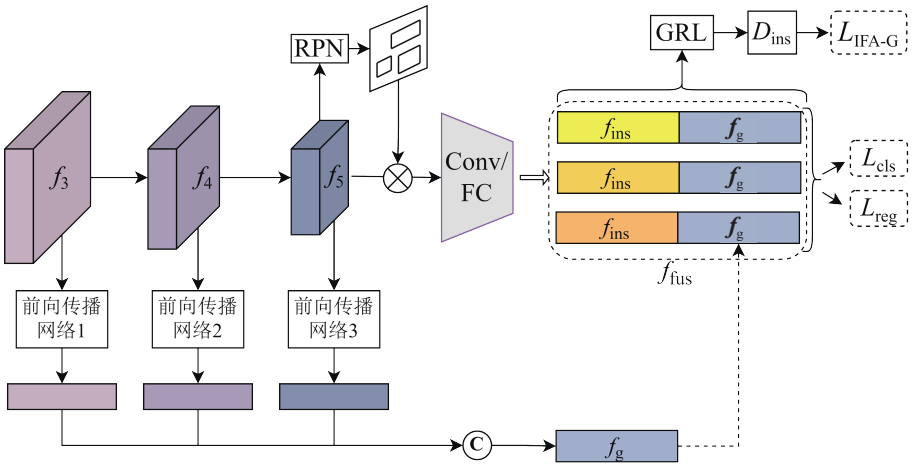


图 5　融合全局特征的实例级特征对齐模块

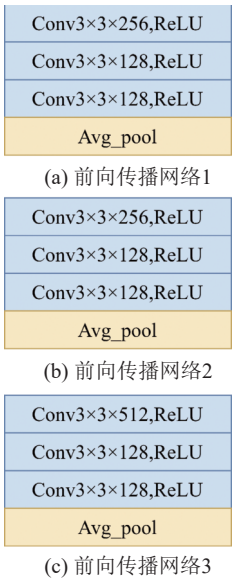


图 6　前向传播网络结构

对齐损失由全局特征对齐损失 $L_{\text{MFA-NA}}$ 和实例级特征对齐损失 $L_{\text{IFA-G}}$ 组成，源域和目标域数据均需计算这两项损失。最终，对以上损失函数求和，得到模型的整体目标函数：

$$L = L_{\text{cls}} + L_{\text{reg}} - \lambda(L_{\text{MFA-NA}} + L_{\text{IFA-G}})。$$

其中，超参数 λ 用来平衡 Faster R-CNN 网络的检测损失与域自适应过程中特征对齐损失之间的权重。

2　实验

为了验证本方法的有效性，将 PASCAL VOC 数据集作为源域数据集，将 ExDARK 数据集作为目标域数据集，将正常照度图像的知识迁移到低照度图像，以在低照度图像中实现较好的检测效果。

2.1　实验设置

实验在配置了 Intel Xeon CPU E5-2650 v4 @ 2.20 GHz 处理器、128 GiB 内存、Tesla P40 图形处理器、安装了 Ubuntu 20.04.3 系统的服务器上完成。训练和测试在 PyTorch 深度学习平台完成。使用 Image Net 数据集[34] 预训练 Resnet101 模型初始化参数，之后使用 PASCAL VOC 和 ExDARK 数据集对模型进行训练，初始学习率设为 0.001，batch_size 设为 2，每次迭代时输入一张源域图片和一张目标域图片，输入图像的短边被调整为 600 像素，超参数 λ 设为 1.0，γ 设为 5.0，权重衰减设为 0.000 5，共训练 20 个 epoch。在测试阶段，将交并比阈值为 0.5 时得到的平均精度（average precision，简称 AP）与平均精度均值（mean average precision，简称 mAP）作为检测性能的评估指标，分别评估各类别检测精度与整体检测精度。

2.2　数据集与评估准则

作为源域的 PASCAL VOC 数据集共 20 个目标类别，图片主要拍摄于光照良好的环境下，图像中目标轮廓清晰，所有图像都具有图像级和实例级标注。在训练阶段，通过源域数据集的标注信息计算检测损失，使目标检测网络获得分类和定位的能力。

作为目标域的 ExDARK 数据集是为了解决低照度目标检测问题而制作，完全由低照度环境下捕获的图像组成，具有低照度情况下不同场景类型的图片，概括了现实中部分低照度场景，由 12 个目标类别组成，具有一定的普适性，且所有图像均具有图像级和实例级标注。本研究用于测试阶段对目标域检测结果的评估。

在 PASCAL VOC 数据集和 ExDARK 数据集中共有 10 个相同的目标种类，分别是自行车、船舶、瓶子、公交车、轿车、猫、椅子、狗、摩托车、行人。在本方法的训练过程中，使用带有标注的 PASCAL VOC 训练集图片 11 540 张以及无标注的 Ex-

DARK 训练集图片 5 890 张; 在测试阶段使用了带有标注的 ExDARK 测试集图片 1 409 张。使用 mAP 评估模型在低照度图像上的检测效果。

2.3　消融实验

为验证网络中 PA、MFA-NA、IFA-G 的有效性, 对各模块进行消融实验。

为了减小源域和目标域之间的特征空间分布差异, 像素级对齐的方法还可以通过对低照度图像进行增强处理的手段, 使低照度图像向正常照度图像的分布靠近。实验使用近期达到 SOTA 表现的 SCI[35] 图像增强方法对低照度图像进行增强并训练, 将增强低照度数据之后进行迁移的方法称为 Enh, 表 1 为使用不同像素级对齐方法的检测结果。从表 1 可看出, 增强的 mAP 比增强前降低了 0.68 个百分点, 与添加 PA 模块相比, 精度下降了 1.53 个百分点, 说明对低照度图像进行增强并不能达到缩小源域和目标域特征差异的目的。因此, 采取图像翻译方法来实现图像像素级对齐。

表 1　使用不同像素级对齐方法的检测精度　%

方法	AP										mAP
	自行车	船舶	瓶子	公交车	轿车	猫	椅子	狗	摩托车	行人	
Baseline	59.54	30.73	54.06	65.59	52.73	54.01	42.75	60.88	52.15	55.21	52.76
Baseline + Enh	60.83	34.79	50.22	63.13	52.80	52.59	40.59	60.20	52.67	53.01	52.08
Baseline + PA	62.35	34.96	51.00	67.08	53.10	56.97	42.61	62.14	51.03	54.84	53.61

在 Baseline 中加入不同模块后的检测精度如表 2 所示, 可看出, 在增加像素级特征对齐模块的处理后, 与 Baseline 相比, 检测精度提升了 1.78 个百分点, 这一步显著提升了模型在低照度图像中的检测精度。由此可得到结论: 将正常照度数据进行风格迁移处理后, 图像的数据分布更接近真实的低照度图像, 经过合成低照度图像数据的训练后, 检测网络在真实低照度图像上的检测能力得到提升。

表 2　Baseline 中加入不同模块后的 mAP　%

方法	mAP
Baseline	52.76
Baseline + PA	54.54
Baseline + MFA-NA	54.20
Baseline + MFA-NA + IFA-G	55.47
Baseline + PA + MFA-NA	55.82
Baseline + PA + MFA-NA + IFA-G	56.72

在 Baseline 中添加 MFA-NA 模块后, 检测精度提升了 1.44 个百分点。在增加 PA 模块的基础上添加 MFA-NA 模块, 检测精度提升了 1.28 个百分点。由此可知, 经 PA 模块处理后, 仍存在一定的域偏移, MFA-NA 模块在像素级对齐的基础上促进了源域与目标域全局特征的对齐。

仅采用 PA+MFA-NA 组合时 mAP 为 55.82, 在此基础接入 IFA-G 全局实例对齐模块后, mAP 提升了 0.90 个百分点。在增加 MFA-NA 模块的基础上添加 IFA-G 模块后, 检测精度提升了 1.27 个百分点。在增加了 MFA-NA 模块后, 模型经过初步的全局特征对齐获得了具有域泛化性的语义信息。在此基础上, 对于融合了全局特征的实例级特征再进行对齐, 进一步提升了模型在目标域中的检测性能。

2.4　检测性能

2.4.1　ExDARK 数据集中的检测结果

在目标域数据集 ExDARK 上综合评估不同检测方法的性能, Baseline 为在正常照度图像中训练得到的 Faster R-CNN 检测模型, Oracle 为无监督模型检测结果的上限值, 由 Faster R-CNN 对有标注的低照度图像进行有监督训练得到。表 3 为本方法与近年提出的无监督域自适应目标检测方法在 ExDARK 数据集上的检测结果。由表 3 可知, 本方法的 mAP 为 56.72%, 在 Baseline 基础上提升了 3.96 个百分点, 与表现最好的 SCL[36] 相比, mAP 提升了 1.12 个百分点。表 4 为不同方法在 ExDARK 数据集上路面常见目标类别的检测结果。从表 4 可看出, 本方法在低照度路面常见目标类别上的 mAP 为 62.07%, 在 Baseline 的基础上提升了 5.03 个百分点, 与 SCL 相比, 提升了 1.35 个百分点。

但类别“瓶子”和“椅子”在经过迁移学习之后, 检测精度均有下降。这可能与这两类图片在黑暗环境中轮廓不清晰有关, 如玻璃杯透明, 轮廓不清晰; 椅子部分轮廓纤细。

图 7 为本方法与其他方法的部分可视化数据, 可

表 3 ExDARK 数据集上不同域自适应目标检测方法的检测精度 %

方法	AP										mAP
	自行车	船舶	瓶子	公交车	轿车	猫	椅子	狗	摩托车	行人	
Baseline	59.54	30.73	54.06	65.59	52.73	54.01	42.75	60.88	52.15	55.21	52.76
SWDA	62.28	38.28	43.75	74.54	54.96	61.02	42.06	64.22	52.19	56.73	55.00
DAfrcnn	62.31	36.69	46.47	72.87	54.99	61.49	42.27	65.02	51.27	57.44	55.08
ATF[37]	51.17	24.17	41.68	60.31	44.27	44.79	32.06	47.50	40.64	46.80	43.34
SCL	64.26	38.63	43.21	72.84	55.54	62.15	43.31	65.13	53.27	57.67	55.60
I^3Net[38]	57.79	30.14	38.40	60.23	46.88	47.11	42.78	54.06	48.44	44.35	47.02
本方法	64.76	40.45	45.31	75.13	55.62	61.79	42.48	66.73	55.94	58.91	56.72
Oracle	77.62	63.70	65.29	78.97	72.59	70.91	60.40	75.43	68.96	66.88	70.08

表 4 不同方法在 ExDARK 数据集路面常见目标中的检测精度 %

方法	AP					mAP
	自行车	公交车	轿车	摩托车	行人	
Baseline	59.54	65.59	52.73	52.15	55.21	57.04
SWDA	62.28	74.54	54.96	52.19	56.73	60.14
DAfrcnn	62.31	72.87	54.99	51.27	57.44	59.78
SCL	64.26	72.84	55.54	53.27	57.67	60.72
I^3Net	63.02	72.95	52.43	49.93	47.64	57.19
本方法	64.76	75.13	55.62	55.94	58.91	62.07
Oracle	77.62	78.97	72.59	68.96	66.88	73.00

看出，本方法更容易检测到隐藏在黑暗环境中的目标，误检率和漏检率均下降，更加近地面真实标注。

2.4.2 极低照度环境下的检测结果

在极低照度图像上进行实验。通过伽马变换模拟极低照度的环境，对 ExDARK 图像进行 γ 随机阈值在 [2.5,3.5] 的调暗处理，得到了模拟的极低照度数据集 Ex^2DARK。图 8 为 Ex^2DARK 数据集与 ExDARK 数据集的对照，以及本方法在 Ex^2DARK 数据集上的检测结果可视化。表 5 为不同方法在源域为 PASCAL VOC、目标域为 Ex^2DARK 条件下训练得到的模型在 Ex^2DARK 测试集上的检测结果。从表 5 可看出，在极低照度的环境下，PA 模块的效果与普通低照度相比更加显著，由此可知，在域偏移极大的情况下，像素级对齐对于域间特征空间差异的缩小起关键作用。

3 结束语

针对无监督低照度目标检测问题，提出了一种低照度下的域自适应目标检测方法。为了能更精准地对齐正常照度和低照度图像之间的特征，设计了像素级对齐模块，通过生成低照度合成图像来拉近源域和目标域之间的距离。另外，设计了结合噪声注意力的多层特征对齐模块，减弱了噪声对对抗训练的影响。在实例级特征对齐阶段，增加了全局特征的引导，增强了实例级特征的域泛化性，提高了无监督域自适应模型的检测能力。与其他无监督域目标检测方法相比，本方法在低照度环境下具有更高的检测精度。然而，本方法的训练步骤仍有待简化，基于 Faster R-CNN 的模型检测速度也有待提高。在后续工作中，将针对无监督域目标检测模型的简化与轻量化开展进一步研究。

参考文献：

[1] Zaidi S S A, Ansari M S, Aslam A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.

[2] 吴相远, 申诺, 蒙玉洪, 等. 基于 U-Net++的人体脊椎 MRI 图像识别[J]. 桂林电子科技大学学报, 2022, 42(1): 36-42.

[3] 吕方方, 陈光喜, 刘家畅, 等. 基于卷积神经网络的小目标检测改进算法[J]. 桂林电子科技大学学报, 2021, 41(5): 368-374.

[4] Li Chongyi, Guo Chunle, Han Linghao, et al. Lighting the darkness in the deep learning era[PP/OL]. arXiv (2021-04-21)[2023-10-13]. https://arxiv.org/pdf/2104.10729v1.pdf.

[5] Loh Y P, Chan C S. Getting to know low-light images with the Exclusively Dark dataset[J]. Computer Vision and Image Understanding, 2019, 178: 30-42.

[6] Sasagawa Y, Nagahara H. Yolo in the dark-domain adaptation method for merging multiple models[C]//Computer Vision–ECCV 2020. Berlin: Springer, 2020: 345-359.

[7] Guo Chunle, Li Chongyi, Guo Jichang, et al. Zero-reference deepcurve estimation for low-light image enhance-



图 7　部分模型在 ExDARK 数据集中检测结果可视化



图 8　模型在 Ex^{2}DARK 中检测结果可视化

表 5　Ex^{2}DARK 数据集上不同无监督域自适应目标检测方法的检测精度　　%

方法	不同类别 AP										mAP
	自行车	船舶	瓶子	公交车	轿车	猫	椅子	狗	摩托车	行人	
Baseline	21.33	16.17	17.16	36.39	18.88	12.38	12.60	22.99	20.17	19.14	19.72
SWDA	20.08	9.57	9.76	29.98	17.18	12.68	8.01	10.16	19.13	18.87	15.54
DAfrcnn	22.32	13.48	21.79	33.69	20.44	15.31	13.03	19.71	21.97	21.16	20.29
SCL	26.03	12.47	19.25	37.38	21.19	12.75	12.47	16.33	24.24	21.66	20.38
Baseline + PA	36.23	18.79	23.39	44.12	22.14	23.33	15.27	29.31	28.35	28.86	26.98
本方法	35.23	17.30	22.38	45.84	26.40	29.07	14.44	29.12	32.09	29.23	28.12
Oracle	61.76	42.59	46.92	72.48	51.90	49.92	36.11	55.98	56.00	49.68	52.33

ment[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 1777-1786.

[8] Wang Mei, Deng Weihong. Deep visual domain adaptation: a survey[J]. Neurocomputing, 2018, 312: 135-153.

[9] 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.

[10] Shrivastava A, Shekhar S, Patel V M. Unsupervised domain adaptation using parallel transport on grassmann manifold[C]// IEEE Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE Press, 2014: 277-284.

[11] Sun B C, Feng J S, Saenko K. Return of frustratingly easy domain adaptation[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2016: 2058-2065.

[12] Hoffman J, Wang Dequan, Yu F, et al.Fcns in the wild:pixel-level adversarial and constraint-based adaptation[PP/OL]. arXiv (2016-12-08)[2023-10-13]. https://arxiv.org/pdf/1612.02649.pdf.

[13] Tsai Y H, Hung W C, Schulter S, et al. Learning to adapt structured output space for semantic segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 7472-7481.

[14] Chen Yuhua, Li Wen, Sakaridis C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 3339-3348.

[15] Zhu Xinge, Pang Jiangmiao, Yang Ceyuan, et al. Adapting object detectors via selective cross-domain alignment[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 687-696.

[16] Saito K, Ushiku Y, Harada T, et al. Strong-weak distribution alignment for adaptive object detection[C]//International Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press, 2019: 1-8.

[17] Arruda V F, Paixao T M, Berriel R F, et al. Cross-domain car detection using unsupervised image-to-image translation: from day to night[C]// International Joint Conference on Neural Networks. Piscataway, NJ: IEEE Press, 2019: 1-8.

[18] Sakaridis C, Dai Dengxing, Van-Gool L. Semantic foggy scene understanding with synthetic data[J]. International Journal of Computer Vision, 2018, 126(9): 973-992.

[19] Johnson-Roberson M, Barto C, Mehta R, et al. Driving in the matrix: can virtual worlds replace human-generated annotations for real world tasks?[PP/OL]. arXiv (2016-10-06)[2023-10-13]. https://arxiv.org/pdf/1610.01983.pdf.

[20] Inoue N, Furuta R, Yamasaki T, et al. Cross-domain weakly-supervised object detection through progressive domain adaptation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 5001-5009.

[21] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 779-788.

[22] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 6517-6525.

[23] Redmon J, Farhadi A. Yolov3: an incremental improvement[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1-6.

[24] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: optimal speed and accuracy of object detection[PP/OL]. arXiv (2020-04-23)[2023-10-13]. https://arxiv.org/pdf/2004.10934.pdf.

[25] Ge Z, Liu S T, Wang F, et al. Yolox: exceeding yolo series in 2021[PP/OL]. arXiv (2021-07-18)[2023-10-13]. https://arxiv.org/pdf/2107.08430.pdf.

[26] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[27] Girshick R. Fast R-CNN[C]//IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 1440-1448.

[28] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: an overview[J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.

[29] Zhu Junyan, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 2242-2251.

[30] Van Der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11): 2579-2605.

[31] Everingham M, Gool L, Williams C K I, et al. The PASCAL visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[32] Olsen S I. Estimation of noise in images: an evaluation [J]. CVGIP: Graphical Models and Image Processing,

1993, 55(4): 319-323.

[33] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 2999-3007.

[34] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[35] Ma Long, Ma Tengyu, Liu Risheng, et al. Toward fast, flexible, and robust low-light image enhancement[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2022: 5637-5646.

[36] Shen Z Q, Maheshwari H, Yao W C, et al. SCL: towards accurate domain adaptive object detection via gradient detach based stacked complementary losses[PP/OL]. arXiv (2019-11-06)[2023-10-13]. https://arxiv.org/pdf/1911.02559.pdf.

[37] He Zhiwei, Zhang Lei. Domain adaptive object detection via asymmetric tri-way Faster-RCNN[C]//Computer Vision-ECCV 2020. Cham: Springer, 2020: 309-324.

[38] Chen Chaoqi, Zheng Zebiao, Huang Yue, et al. I3Net: implicit instance-invariant network for adapting one-stage object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2021: 12571-12580.

编辑：张所滨