SDDA:无监督的风格和分布域适应夜间语义分割方法

雷晓春[1,2],吴炜林[1*],江泽涛[1,2],朱文才[3],刘颖健[1],陈冬梅[1],吴思琦[1]
(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. 广西图像图形与智能处理重点实验室,广西桂林 541004;
3. 西北工业大学计算机学院,陕西西安 710072)

摘 要: 语义分割在自动驾驶、医工交叉和安防监控等多种实际应用中发挥着重要作用,但夜间语义分割仍然是未解决的一道难题。由于夜间光照不足,获取的图像细节模糊不清,导致数据集标注困难,因而人们首选探索无监督域适应夜间语义分割方法。虽然取得了一些进展,但仍然存在数据集跨域幅度太大难以直接进行域适应的问题,导致夜间场景的语义分割效果不理想。针对这个问题,本文提出了一种风格和分布域适应(Style and Distribution Domain Adaptation,SDDA)的无监督夜间语义分割方法,将夜间语义分割任务的域适应分为风格域适应和分布域适应,以此降低夜间分割任务的难度。将性能更优秀的Mamba架构模型引入无监督域适应夜间语义分割任务中,探索该架构模型在夜间语义分割任务的优势,以提升夜间分割任务的精度。提出了一个语义对齐图像翻译(Semantic Pairing GAN,SPG)模块,通过语义信息将非配对翻译和粗配对翻译相结合,以此将分割任务与SPG翻译模块进行语义关联,促进翻译内容更加适合分割任务,且不独立于分割任务。SPG模块先将源域白天图像翻译成夜间图像,然后分割模型用翻译后的图像进行训练,这样分割模型就能学习到风格域信息以减少风格域差异。提出了一种语义域混合(Semantic Domain Mixing,SDM)策略,利用语义信息将SPG翻译的动态物体提取并移动到目标域夜间静态物体图像的合理位置,重新组合成新的图像。分割模型利用这种风格域差异小的图像进行训练,可以更容易从分布域角度进行域适应,从而缩小分布域差距。通过风格域适应和分布域适应相结合,使模型从两种不同角度分别缩小域差异,整体上实现夜间分割任务的域适应,从而缓解现有数据集跨域幅度太大,难以直接域适应的问题。实验结果表明,本文的方法在Dark Zurich、ACDC Night和Nighttime Driving三个数据集上的mIoU指标分别取得60.0%、59.8%、59.1%,比现有最好的方法分别提高0.9%,0.4%和1.6%,对夜间复杂实际场景图像目标能进行精准的分割预测。

关键词: 无监督域适应;夜间语义分割;图像到图像翻译;深度学习;图像分割

基金项目: 国家自然科学基金(No.62473105);广西自然科学基金面上项目(No.2025JJA170088);广西图像图形智能处理重点实验项目(No. GIIP2305);桂林电子科技大学研究生教育创新计划项目(No. 2025YCXS051,No.2024YCXS035)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2026)01-0433-18

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20251221

SDDA: Unsupervised Style and Distribution Domain Adaptation Method for Nighttime Semantic Segmentation

LEI Xiaochun[1,2], WU Weilin[1*], JIANG Zetao[1,2], ZHU Wencai[3], LIU Yingjian[1], CHEN Dongmei[1], WU Siqi[1]
(1. *School of Computer and Information Security*, *Guilin University of Electronic Technology*, *Guilin*, *Guangxi* 541004, *China*;
2. *Guangxi Key Laboratory of Image and Graphics Intelligent Processing*, *Guilin*, *Guangxi* 541004, *China*;
3. *School of Computer Science*, *Northwestern Polytechnical University*, *Xi'an*, *Shaanxi* 710072, *China*)

Abstract: Semantic segmentation plays an important role in a variety of practical applications such as autonomous driving, doctor-worker intersection, and security monitoring. However, nighttime semantic segmentation is still an unsolved problem. Due to insufficient illumination at night, the details of the acquired image are unclear, which leads to the difficulty of dataset annotation. Therefore, unsupervised domain adaptation methods for nighttime semantic segmentation are preferred. As a result, the semantic segmentation effect of nighttime scenes is not ideal. To solve this problem, this paper proposes an unsupervised SDDA (Style and Distribution Domain Adaptation) method for nighttime semantic segmentation. The domain adaptation of nighttime semantic segmentation task is divided into style domain adaptation and distribution domain adaptation. In this way, the difficulty of the nighttime segmentation task is reduced. The Mamba architecture model

收稿日期:2026-01-12;录用日期:2026-01-19;责任编委:唐旭
*通信作者:吴炜林

with better performance is introduced into the unsupervised domain to adapt to the nighttime semantic segmentation task, and the advantages of this architecture model in the nighttime semantic segmentation task are explored to improve the accuracy of the nighttime segmentation task. This paper proposes a SPG (Semantic Pairing GAN) module, which combines the unpaired translation and rough paired translation through semantic information, so as to semantically associate the segmentation task with the SPG translation module, so as to promote the translation content to be more suitable for the segmentation task and not independent of the segmentation task. The SPG module translates the day images of the source domain into night images, and then the segmentation model is trained with the translated images, so that the segmentation model can learn the style domain information to reduce the style domain differences. This paper proposes a SDM (Semantic Domain Mixing) strategy, which uses semantic information to extract and move the dynamic objects translated by SPG to the reasonable position of the static object image at night in the target domain, and recombines them into a new image. The segmentation model is trained by using the images with small style domain differences, which makes it easier to perform domain adaptation from the perspective of distribution domain, so as to narrow the distribution domain gap. Through the combination of style domain adaptation and distribution domain adaptation, the model reduces the domain differences from two different perspectives, and realizes the domain adaptation of night segmentation tasks as a whole, so as to alleviate the problem that the existing data sets have too large cross-domain range and are difficult to directly adapt to the domain. The experimental results show that the mIoU index of the proposed method on Dark Zurich, ACDC Night and Nighttime Driving datasets achieves 60.0%, 59.8% and 59.1%, respectively, which is 0.9%, 0.4% and 1.6% higher than the best existing method. It can accurately segment and predict the image target of complex actual scene at night.

Keywords: unsupervised domain adaptation; nighttime semantic segmentation; image to image translation; deep learning; image segmentation

Foundation Item(s): National Natural Science Foundation of China (No.62473105); Natural Science Foundation of Guangxi (No.2025JJA170088); Guangxi Key Laboratory of Image and Graphic Intelligent Processing (No.GIIP2305); Innovation Project of GUET Graduate Education (No.2025YCXS051, No.2024YCXS035)

0 引言

语义分割是计算机视觉领域中的一个重要研究方向,旨在为图像的每个像素分配一个特定语义类别的标签,在自动驾驶、医工交叉和安防监控等领域有十分广阔的应用前景。近年来,随着深度学习的发展,语义分割技术取得了显著进展[1-5]。然而,由于人工难以对夜间数据集进行像素级标注,所以大量已有工作仅在正常白天环境下进行研究,而对于光照不充足的夜间环境下的语义分割方法缺乏探索。

最近,一些研究者提出从无监督的角度对夜间语义分割任务进行探索。这些无监督夜间语义分割方法大致分为两类。

(1)基于图像翻译的方法。这些方法[6-8]采用图像翻译模型生成夜间图像,并利用白天场景的标注进行训练,以解决夜间场景下像素级人工标注困难的问题。江泽涛等人[8-10]利用白天图像翻译成夜间图像训练分割网络,以学习夜间知识。

(2)基于Mixup的方法。现有的一些研究[11-13]采用Mix混合图像训练方法,通过将源域图像的物体和目标域图像的物体混合成新的图像进行训练,以缩小两个域之间的差异。DSRNSS[13]利用Mix方法将动态物体和小目标物体融合进目标域图像再进行特征原型对齐,缩小源域和目标域的差距。

然而,这两类方法均存在一定的缺点。具体来说,第一类方法常常直接采用翻译模型生成夜间图像,然后再进行分割任务。由于这种方法翻译模型和分割网络各自独立,所生成的夜间图像可能并不完全适用于分割任务,生成的图像与真实夜间场景下的图像之间有很大的域差距[14-15]。第二类Mix混合图像通常是来自两种不同的数据集,但这样的图像不仅存在风格域(白天和夜间环境)不一致的情况[13],还存在分布域(数据集来源分布差异)的不一致[16],这样的域差异过大不利于模型学习。

现有的无监督域适应夜间语义分割方法大多数还是以ResNet为骨干的网络[16-18],但其性能和以Transformers为骨干的网络存在很大的差距。直到DAFormer[19]首次将SegFormer[20](MiT)模型Transformers架构引入无监督域适应分割任务中,性能得到了提升,一些夜间任务的方法[21-22]也开始采用DAFormer模型来提升性能。但Transformers架构存在模型参数量大、计算资源高、推理速度慢等问题,难以运用于实际。

针对上述问题,本文提出一种风格和分布域适应(Style and Distribution Domain Adaptation,SDDA)的无监督夜间语义分割方法。主要贡献如下。

(1)创新性地将Mamba用于无监督域适应夜间语义分割任务,探索其在夜间分割的优势,提升夜间

分割的精度。

(2)提出了语义对齐图像翻译(Semantic Pairing GAN,SPG)模块,其非配对翻译结合语义指导的粗配对翻译可以减小风格域差异。SPG与分割任务建立语义联系,SPG翻译图像促进分割任务,分割任务指导SPG翻译图像,两者相互促进。

(3)提出了一种语义域混合(Semantic Domain Mixing,SDM)策略,利用SPG训练时生成的夜间图像和真实白天图像来生成较为合理的混合域图像和伪标签,促进模型分布域适应。

(4)在Dark Zurich、ACDC Night和Nighttime Driving三个数据集上的mIoU指标分别取得60.0%、59.8%、59.1%,优于一些现有的方法,验证了本文方法的有效性。

1 相关工作

1.1 分割模型

在语义分割任务中,卷积神经网络(Convolutional Neural Network,CNN)长期以来一直是主流的网络,通过卷积从浅层(纹理、边缘)到深层的高级语义信息逐层学习。FCN[23]开创了分割端到端训练的时代,随着U-Net[24]、RefineNet[25]、DeepLabv2[26]等模型的出现,进一步推动了语义分割的发展。RefineNet采用多路径细化网络,利用下采样过程中所有可用的信息,以使用残差连接实现高分辨率预测。但由于CNN在全局建模存在一定的局限性,容易受到局部特征干扰,性能因此受到限制。近年来,Transformer因其强大的全局上下文建模能力,开始在视觉任务中得到广泛应用。SETR[27]是首个将Transformer应用于语义分割任务的模型,解决了CNN无法有效全局建模的问题。SegFormer[20]、Swin Transformer[28]等模型进一步提升了语义分割性能。SegFormer提出一种分层结构Transformer编码器,使用简单的MLP解码器聚合来自不同层的信息,这些简单的设计实现了高效的分割。然而,Transformer的计算复杂度高、推理速度慢、硬件要求高,不适合应用部署。

近期提出的Mamba架构能够长距离依赖建模,其性能与Transformer及其变体相当,甚至在某些方面表现更优,其训练效率和推理速度也得到了极大地提升。同时研究表明[29],Mamba架构在语义分割任务中具有巨大的潜能。本文尝试将性能更优的架构应用于夜间无监督域适应任务,探索其在夜间的处理能力,以得到更好的分割性能。

1.2 夜间语义分割

现有大多数工作都是在光线良好的白天场景下进行的,但也有一些研究专注于夜间这种更具挑战性的环境。一些研究者通过引入一个中间域的方法[30-32],渐进地将模型从源域过渡到中间域最后迁移到目标域。MGCDA[32]先从白天数据训练,再进行黄昏数据训练,最后到夜间数据训练,从相对简单的白天条件训练逐步到比较困难的夜间条件训练。这类方法使训练流程变得复杂,需要分阶段训练,效率低下且繁琐。另一些研究从夜间光照影响的角度出发,得到了先增强夜间图像[33-36]再进行分割的思路[37-40]。DTP[37]将夜间图像分解为反射分量和光照分量,使分割网络在不同的照明条件下提取一致的特征,减弱光照对模型的影响。然而这些方法需要标签进行有监督训练,但夜间数据集的标签较少,所以导致模型表现不佳。在其他领域的研究者成功将无监督域适应方法应用于合成数据集到真实数据集的分割任务,同样的思路也可以被用于夜间分割。针对夜间分割,可以把白天场景和夜间场景看成两个不同的域,这样就可以运用域适应的方法[17-18,41]来解决夜间分割。DANNet[17]采用对抗训练,使用白天图像上静态类别的预测作为伪标签来分割其对应的夜间图像。还有一些与域适应结合的方法,比如结合原型的域适应[13],利用特征原型缩小域间差距。还有通过传统的图像增强方法降低图像的亮度生成低照度图像,再用原图像的标签作为低照度图像标签[42-43]进行监督训练。虽然这些方法尝试结合多种方法来应对夜间问题,但它们往往过于复杂,解决方案层层叠加,导致模型难以收敛,而且传统的图像增强可能导致过曝或欠曝的情况,这样的图像反而会增加模型的训练难度。

1.3 Mix

Mix是一种数据增强的方法,其中具有代表性的方法有CutMix[44]和ClassMix[45]。CutMix是从两张图像按一定比例随机裁剪互补合成一张新图像。Attentive Cutmix[46]在CutMix基础上加上注意力机制,使用注意力减少随机性使图像融合更加符合实际。ClassMix利用分割标签的掩码可以精确地裁剪出对应类别的像素区域,再粘贴到另一张图像上,可以生成更加合理的图像。这些Mix方法都专注于精准分割出一张图片的对象,而忽略了粘贴到另一张图像的整体合理性。DACS[47]利用ClassMix思想,将源域图像剪切到目标域图像,这样的混合域图像可以有利于缩小域差异。Zhou等人[48]则提出了一种新的Mix——上下文掩码生成策略,通过挖掘源域的先验空间分布和目标域的上下文关系,用上下文掩码生成混合图像。Chen等人[49]提出一种双教师模型,一个教师从局部尺度的混合数据训练,另一个从整体尺度的混合数据训练,最后两者共同训练学生模型。这些方法通过充

分挖掘合成图像的信息，来促进模型学习到更多知识，但没有考虑到两张图片的来源不同，这样合成的图像既有风格域差异又有分布域差异，导致域差异过大等问题。

和现有方法不同的是，本文提出的SDM策略通过重新随机组合来自不同数据源的图像中的静态物体和动态物体，生成新的图像和标签。在这一过程中，存储的动态物体图像由SPG生成，原图像属于白天域，但逐渐过渡到夜间域。同时，将动态物体移至特定的语义位置。这种方法能够生成合理的混合图像，并通过使用中间域图像来合成混合域图像训练，从而解决传统Mix方法中混合域图像存在的过大域差异问题。

2 方法

2.1 设计思想

本方法的设计思想是将无监督域适应夜间语义分割任务细化为风格域适应和分布域适应两个子任务，这一细化将降低任务难度，让模型更有效地学习。首先，针对风格域适应，提出了SPG语义对齐图像翻译模块，它将与分割网络一同训练。SPG训练过程将非配对和粗配对翻译相结合。图像整体采用非配对翻译，翻译网络生成的夜间图像进行有监督的分割训练，进而提升分割网络对夜间图像的分割能力。粗配对翻译则利用分割网络得到的目标域静态类部分进行配对翻译。两者结合训练提升翻译网络的性能，这样可以将翻译网络和分割网络建立起语义关联，两者相互促进使得风格域差异减小。其次，针对分布域适应，提出了一种SDM语义域混合策略，它先将模型输出的目标域白天静态部分预测、目标域夜间静态部分预测、高置信度的预测和源域动态物体标签制作成伪标签，然后将目标域夜间图像和动态物体对应的SPG生成的夜间图像组合成新图像。其中，会将动态类物体移动到特定语义位置，生成更加符合现实的、合理的图像。动态物体图像是由SPG生成的属于白天域但会逐渐接近夜间域的图像，这样生成的图像在模型训练初期减少了风格域的影响，以便促进分布域的适应。而在后期，分布域保留的同时，风格域的影响逐渐增加，这种由简单到困难的逐渐性学习可以更好地促进分布域适应。

2.2 整体架构

本文的整体架构如图1所示。本架构需要一个有标签的源域$\boldsymbol{S}_d$和两个无标签粗对齐的目标域$\boldsymbol{T}_d$和$\boldsymbol{T}_n$。首先，源域数据$\boldsymbol{X}_{S_d}\in\boldsymbol{S}_d$具有像素级注释标签$\boldsymbol{Y}_S$，这样$\boldsymbol{X}_{S_d}$经过分割网络$f_\theta$得到的预测结果$\bar{\boldsymbol{Y}}_{S_d}$与标签$\boldsymbol{Y}_S$用交叉熵损失做监督训练$\mathcal{L}_{\text{source}}=\mathcal{L}_{\text{ce}}\left(f_\theta(\boldsymbol{X}_{S_d}),\boldsymbol{Y}_S\right)$。

$$\mathcal{L}_{\text{ce}}\left(\bar{\boldsymbol{Y}},\boldsymbol{Y}\right)=-\sum_{c=0}^{C}\sum_{h=0}^{H}\sum_{w=0}^{W}\boldsymbol{Y}^{c,h,w}\log\left(\bar{\boldsymbol{Y}}^{c,h,w}\right)\tag{1}$$

其中，$\mathcal{L}_{\text{ce}}$为式(1)的交叉熵损失函数；C、H、W分别为预测图像的通道数、高和宽。

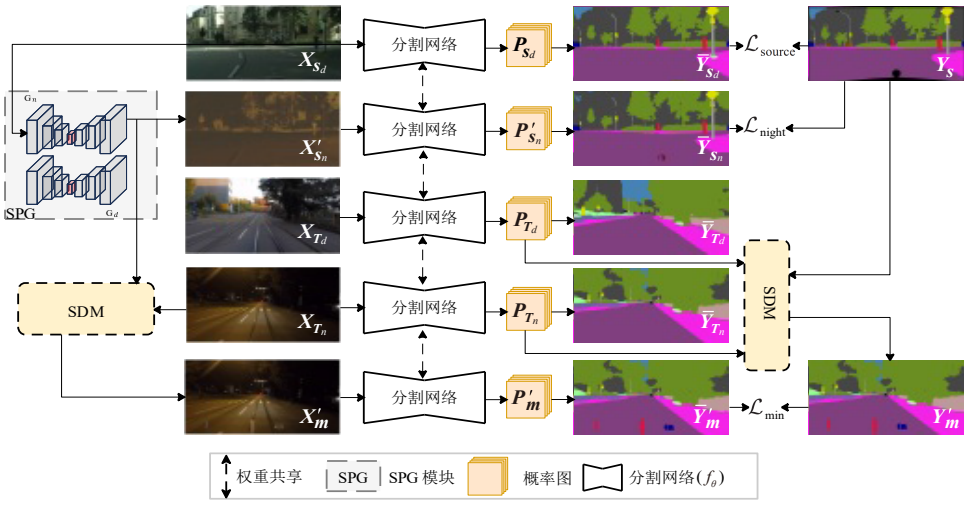


图1 SDDA方法的整体结构

Figure 1 Overall structure of the SDDA method

其次，$\boldsymbol{X}_{S_d}$经过SPG模块可以生成夜间图像$\boldsymbol{X}'_{S_n}$，$\boldsymbol{X}'_{S_n}$和$\boldsymbol{X}_{S_d}$除了白天和夜间域不一样，图像结构是一致的，所以$\boldsymbol{Y}_S$也是$\boldsymbol{X}'_{S_n}$的标签，可以与模型预测的$\bar{\boldsymbol{Y}}_{S_n}$做交叉熵损失进行监督训练$\mathcal{L}_{\text{night}}=\mathcal{L}_{\text{ce}}\left(f_\theta(\boldsymbol{X}'_{S_n}),\boldsymbol{Y}_S\right)$。

然后，目标域的白天图像$\boldsymbol{X}_{T_d}\in\boldsymbol{T}_d$和夜间图像$\boldsymbol{X}_{T_n}\in\boldsymbol{T}_n$是没有标签的。通过使用分割网络分别获得

它们的分割预测概率 $\boldsymbol{P}_{T_d}$ 和 $\boldsymbol{P}_{T_n}$,将两者和 $\boldsymbol{Y}_S$ 一起输入到 SDM,同样地,$\boldsymbol{X}'_{S_n}$ 和 $\boldsymbol{X}_{T_n}$ 也输入到 SDM。最后,SDM 会随机提供混合图像 $\boldsymbol{X}'_m$ 和伪标签 $\boldsymbol{Y}'_m$,混合图像 $\boldsymbol{X}'_m$ 通过分割网络得到预测 $\bar{\boldsymbol{Y}}'_m$ 与伪标签 $\boldsymbol{Y}'_m$ 做交叉熵损失。

2.3　SDDA 分割网络

之前的无监督域适应夜间语义分割方法[16,18,21-22]大多数使用 RefineNet[25]、DeepLabv2[26]、MiT[20]网络架构,但这些网络的性能已经比不过最新的网络。因此,本文尝试将性能更优的 VMamba[50]引入本文方法,以提高模型的性能和域适应能力。

本文将这些常见的模型[20,25-26,50]进行有监督的夜间语义分割训练。具体实验细节将在 3.3.1 节详细介绍。其训练的可视化部分结果如图 2 所示。

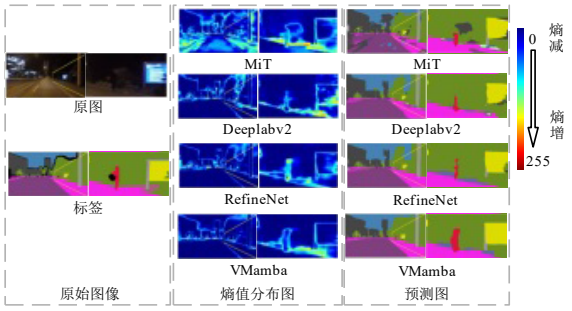


图 2　夜间分割细节可视化结果

Figure 2　Visualization results for nighttime segmentation details

从图 2 中可以看到,在夜间低照度环境下,人眼很难分辨出图中放大的部分存在物体,通过标签进行对应可以发现,确实存在一个行人。本文分别用这些模型[20,25-26,50]进行预测输出得到熵值分布图和预测图。由香农公式(2)得出,熵值越大不确定性越大,即模型预测正确的可能性很低(图中蓝色部分是低响应的熵值,熵值越小说明模型预测得越好)。

$$H=-\sum_{i=1}^{N}\left(\boldsymbol{p}_i\cdot\log(\boldsymbol{p}_i)\right) \tag{2}$$

观察图 2 中预测图可以发现,四个模型都可以预测到行人,但 MiT、Deeplabv2、RefineNet 效果并不好。对应到熵值分布图,MiT 和 RefineNet 虽然能预测到行人,但它们的熵值很高,也就是说模型的不确定性很高,模型性能还是比较差。从图片整体来看,MiT 预测和熵值分布图都表现很差,Deeplabv2、RefineNet 相当,VMamba 预测图效果好且熵值分布整体较低,所以整体来看 VMamba 性能最好。从 3.3.1 节的实验结果可以看到,VMamba 不仅在正常场景下的性能很好,而且在夜间低照度场景下的分割性能同样很好。

经过实验验证,VMamba 模型的 mIoU 性能和可视化结果都表现出色,在夜间低照度场景下的分割性能优于其他模型,证明其非常适合夜间任务,所以本文尝试将其运用于无监督夜间语义分割任务中,以提升无监督夜间分割任务的性能。

2.4　语义对齐图像翻译模块 SPG

2.4.1　非配对图像翻译

为了使分割模型缩小风格域差距,可以采用图像翻译网络将白天图像翻译成夜间图像,利用白天的标签让模型学习到夜间知识。然而,直接采用已有的翻译网络会与分割任务分离,导致翻译的内容并不适用分割任务。所以本文需要让图像翻译网络和分割网络建立语义关系训练,以适配分割任务。现实世界图像很难找到白天场景和夜间场景同一个位置且物体位置不变的匹配图像,因此很难做到有监督的训练翻译网络。CycleGAN[51]等非配对翻译模型的出现,解决了需要配对图像才能翻译的问题,近年来的一些扩散模型[52]也可以实现非配对翻译。但由于大多数扩散模型在生成图像时需要多个推理步骤,生成速度相对较慢,并且对显卡硬件资源的要求更高[53],不利于与分割网络一同训练。由于生成的图像结构化信息要与原图像保持一致,才能进行后续的分割,因此,相对于扩散模型而言,循环生成对抗模型更加适合非配对的夜间语义分割任务。所以本文利用这种循环生成对抗思想结合粗配对翻译,设计了一个 SPG 模块,如图 3 所示。本文将分别从图像层面和特征层面进行图像翻译约束,以更加全面地将白天域图像迁移到夜间域,让模型学习到更多夜间域知识。

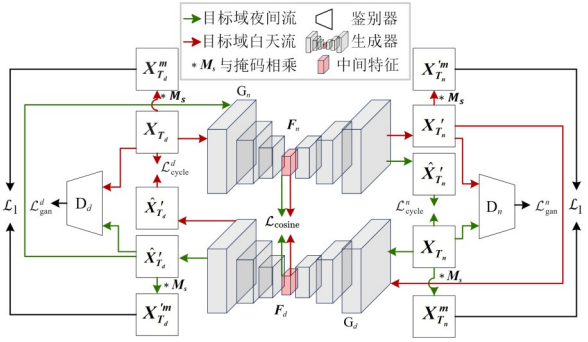


图 3　SPG 模块

Figure 3　SPG module

图像层面翻译,白天域图像 $\boldsymbol{X}_{T_d}$ 通过生成器 G_n 生成夜间域图像 $\boldsymbol{X}'_{T_n}$,$\boldsymbol{X}'_{T_n}$ 和真实夜间图像 $\boldsymbol{X}_{T_n}$ 通过判别器 D_n 进行 $\mathcal{L}^n_{\text{gan}}$ 损失计算 $\mathcal{L}^n_{\text{gan}}=\mathcal{L}_{\text{GAN}}(D_n,G_n,\boldsymbol{X}_{T_d},\boldsymbol{X}_{T_n})$,以此训练生成器和判别器,$D_n$ 促使 $\boldsymbol{X}'_{T_n}$ 像素更加接近真实图像 $\boldsymbol{X}_{T_n}$ 像素。

$$\mathcal{L}_{\mathrm{GAN}}(\mathrm{D},\mathrm{G},\boldsymbol{X},\boldsymbol{Y})=\mathrm{E}_{\boldsymbol{Y}\sim p_{\mathrm{data}}(\boldsymbol{Y})}\left[\log\mathrm{D}(\boldsymbol{Y})\right]+\mathrm{E}_{\boldsymbol{X}\sim p_{\mathrm{data}}(\boldsymbol{X})}\left[\log\left(1-\mathrm{D}(\mathrm{G}(\boldsymbol{X}))\right)\right] \quad (3)$$

其中，$\mathcal{L}_{\mathrm{GAN}}$ 为式(3)。接着，$\boldsymbol{X}'_{T_n}$ 经过生成器 G_d 生成白天域图像 $\hat{\boldsymbol{X}}'_{T_d}$，$\hat{\boldsymbol{X}}'_{T_d}$ 和 $\boldsymbol{X}_{T_d}$ 进行循环一致性损失 $\mathcal{L}^d_{\mathrm{cycle}}$ 计算 $\mathcal{L}^d_{\mathrm{cycle}}=\mathcal{L}_{\mathrm{CYC}}(\mathrm{G}_d,\mathrm{G}_n,\boldsymbol{X}_{T_d})$。

$$\mathcal{L}_{\mathrm{CYC}}(\mathrm{H},\mathrm{G},\boldsymbol{X})=\mathrm{E}_{\boldsymbol{X}\sim p_{\mathrm{data}}(\boldsymbol{X})}\left[\left\|\mathrm{H}(\mathrm{G}(\boldsymbol{X}))-\boldsymbol{X}\right\|_1\right] \quad (4)$$

其中，$\|\ \|_1$ 为 $\mathcal{L}_1$ 损失。因为白天域图像 $\boldsymbol{X}_{T_d}$ 先经过夜间生成器 G_n，再经过白天生成器 G_d 生成的白天图像 $\hat{\boldsymbol{X}}'_{T_d}$ 应该与原图像 $\boldsymbol{X}_{T_d}$ 是一样的，这样能保证翻译后的图像整体结构保持不变，以防图像翻译朝着不可预知的方向进行。同理，夜间域图像 $\boldsymbol{X}_{T_n}$ 也是先后经过 G_d 和 G_n，生成 $\boldsymbol{X}'_{T_d}$ 和 $\hat{\boldsymbol{X}}'_{T_n}$ 分别进行 $\mathcal{L}^d_{\mathrm{gan}}=\mathcal{L}_{\mathrm{GAN}}(\mathrm{D}_d,\mathrm{G}_d,\boldsymbol{X}_{T_n},\boldsymbol{X}_{T_d})$ 和 $\mathcal{L}^n_{\mathrm{cycle}}=\mathcal{L}_{\mathrm{CYC}}(\mathrm{G}_n,\mathrm{G}_d,\boldsymbol{X}_{T_n})$ 损失计算。

特征层面翻译，因为生成器 G_d 和 G_n 都是使图像从一个域到另一个域的转变，所以可以假设它们都是先剥离一个域的特征然后再赋予另一个域的特征，或者说模型中间的特征应该远离源域而靠近目标域。例如，$\boldsymbol{X}_{T_d}$ 经过 G_n 的中间特征 $\boldsymbol{F}_n$ 应该与 $\boldsymbol{X}'_{T_n}$ 经过 G_d 的中间特征 $\boldsymbol{F}_d$ 是相似的。所以，可以将 G_d 和 G_n 生成器的中间特征提取出来进行余弦相似度的计算，以此促进特征层的翻译，使生成器更平滑地进行翻译。

$$\mathcal{L}_{\mathrm{COSINE}}(\boldsymbol{x},\boldsymbol{y})=\frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{\boldsymbol{x}_i\cdot\boldsymbol{y}_i}{\|\boldsymbol{x}_i\|\|\boldsymbol{y}_i\|}\right) \quad (5)$$

其中，式(5)为余弦相似度公式，假设中间特征 $\boldsymbol{F}_n$ 和 $\boldsymbol{F}_d$ 维度为 C、H、W，C 为特征维度，H 和 W 分别为特征图高和宽，则 N 为 $C\times H\times W$。中间特征的相似度计算为

$$\mathcal{L}_{\mathrm{cosine}}=\frac{1}{2}\left[\mathcal{L}_{\mathrm{COSINE}}\left(\boldsymbol{F}_n^{X_{T_d}},\boldsymbol{F}_d^{X_{T_d}}\right)+\mathcal{L}_{\mathrm{COSINE}}\left(\boldsymbol{F}_d^{X_{T_n}},\boldsymbol{F}_n^{X_{T_n}}\right)\right] \quad (6)$$

其中，$\boldsymbol{F}_n^{X_{T_d}}$ 为 $\boldsymbol{X}_{T_d}$ 的 G_n 的中间特征，$\boldsymbol{F}_d^{X_{T_d}}$ 为 $\boldsymbol{X}_{T_d}$ 的 G_d 的中间特征，$\boldsymbol{F}_d^{X_{T_n}}$、$\boldsymbol{F}_n^{X_{T_n}}$ 同理。

2.4.2 粗配对图像翻译

粗配对图像翻译指的是在图像部分配对的像素块上进行配对监督翻译。在夜间语义分割任务中，研究人员制作了几个夜间数据集[30-31,54]。其中ACDC[54]和Dark Zurich[31]数据集都有在白天和夜间场景的相同位置上获取图像，所以这些图像的一些物体可以粗略地对齐。这些数据集都是车辆行驶在道路上获取的，物体类别有19种，根据规律可以将这些需要分割物体分为两类，即动态类物体和静态类物体。由于获取的图像从白天到夜间跨越了一段时间，所以同一个地方的动态物体大概率不会继续在原位置，但静态物体大概率会在。本文的主要任务是语义分割，就是将物体分割出来，理论上可以使用分割算法将静态物体都获取到。利用这个规律和方法，可以只对静态物体进行一个粗配对的图像翻译，如图3所示。

白天域图像 $\boldsymbol{X}_{T_d}$ 和生成的夜间域图像 $\boldsymbol{X}'_{T_n}$，都乘以静态掩码 $\boldsymbol{M}_s$（由SDM提供）分别得到只有静态物体的 $\boldsymbol{X}^{\boldsymbol{m}}_{T_d}$ 和 $\boldsymbol{X}'^{\boldsymbol{m}}_{T_n}$ 图像。同理，夜间域图像 $\boldsymbol{X}_{T_n}$ 和生成的白天域图像 $\boldsymbol{X}'_{T_d}$ 也乘以静态掩码 $\boldsymbol{M}_s$ 分别得到 $\boldsymbol{X}^{\boldsymbol{m}}_{T_n}$ 和 $\boldsymbol{X}'^{\boldsymbol{m}}_{T_d}$，$\boldsymbol{X}^{\boldsymbol{m}}_{T_n}$、$\boldsymbol{X}'^{\boldsymbol{m}}_{T_n}$、$\boldsymbol{X}^{\boldsymbol{m}}_{T_d}$ 和 $\boldsymbol{X}'^{\boldsymbol{m}}_{T_d}$ 分别进行 $\mathcal{L}_1$ 损失计算 $\mathcal{L}_{\mathrm{pair}}=\left\|\boldsymbol{X}'^{\boldsymbol{m}}_{T_n}-\boldsymbol{X}^{\boldsymbol{m}}_{T_n}\right\|_1+\left\|\boldsymbol{X}'^{\boldsymbol{m}}_{T_d}-\boldsymbol{X}^{\boldsymbol{m}}_{T_d}\right\|_1$。这样就可以将生成的图像与真实图像进行配对监督训练，使生成图像的领域更接近真实的领域，增强图像层面的翻译。

$$\mathcal{L}_{\mathrm{SPG}}=\lambda_1\mathcal{L}_{\mathrm{cosine}}+\lambda_2\mathcal{L}_{\mathrm{pair}}+\mathcal{L}^d_{\mathrm{gan}}+\mathcal{L}^n_{\mathrm{gan}}+\mathcal{L}^d_{\mathrm{cycle}}+\mathcal{L}^n_{\mathrm{cycle}} \quad (7)$$

其中，λ_1 和 λ_2 是超参数，取值为 $\lambda_1=0.1$，$\lambda_2=0.1$。

2.5 语义域混合SDM

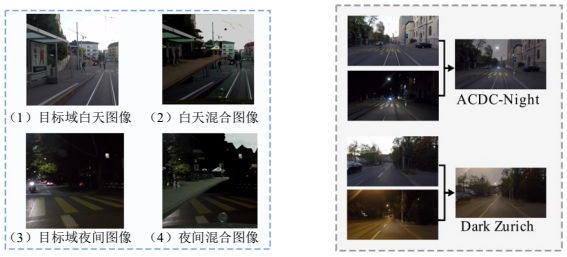
语义域混合SDM与现有的Mix方法不同。如图4所示，Mixup采用的是两张图像随机像素插值得到的混合图像，Cutmix采用源域图像随机截取一部分连续的像素合并到目标域中，Classmix则利用源域的语义标签，精确裁剪某些类别物体粘贴到目标域图像中。Mixup方法像素不连续导致语义信息不连贯；Cutmix方法的随机图像块无法保证截取的物体整体性，会导致一个完整物体被切分；Classmix方法虽然能完整截取物体，但没有考虑两张图像的位置和大小不一样，导致在源域物体粘贴到目标域图像的位置不合理。上述方法都没有考虑两种域图像来源不同，风格域也会不一样。SDM通过动静态物体分类，将SPG生成夜间动态物体移动到道路上，使白天图像和夜间图像合成，促进风格域缩小差距。如图4的SDM所示，其中SPG生成的夜间域汽车被移动到了道路上（红色外框只起到强调提示作用，正常的合成图像不存在）。

图4 常见的Mix方法示意图

Figure 4 Illustration of the common Mix method

SPG在一定程度上缩小了白天和夜间域的风格域差异，然而本文的任务是在源域（如Cityscapes数据集）训练，然后迁移到目标域（如Dark Zurich数据集）。

它们的差异不只是白天域和夜间域,其数据分布也不一致[16]。有些研究[47,49]采用混合域图像促进模型学习两个域之间的差异,但这些方法都没有考虑混合合成图像的合理性和域差距过大的问题。合理性指的是物体位置不符合常理,如图5(a)所示,图5(a)(1)原本是公交站的图像,图5(a)(2)混合变成一条路在公交站中间,图5(a)(3)和图5(a)(4)中的混合同样出现了问题。这样的图像可能会导致全局语义信息的错乱,使分割性能变差,因为本文的任务是语义分割,需要的是语义信息进行的分割,如果缺乏语义信息,实际上就退化为传统的阈值分割。域差距过大是指源域图像和目标域图像直接Mix混合的新图像跨了两个域——风格域(白天和夜间)和数据分布域,难度太大不利于模型学习。



(a) DTBS混合图像　(b) 白天夜间重叠图像

(a) DTBS mix images　(b) Day and night overlapping images

图5　混合图像和图像重叠示意图

Figure 5　Sketch of the mixture image and image overlap

为了解决这些问题,本文将ClassMix和课程学习思想相结合,并充分利用SPG的合成图像,提出了一种渐进性语义Mix混合方法。根据课程学习的思想,可以先学习简单的知识再学习困难的,因此可以用ClassMix方法将SPG合成的夜间域图像$\boldsymbol{X}'_{S_n}$和目标域夜间图像$\boldsymbol{X}_{T_n}$进行混合形成新的图像。SPG是和分割模型一起训练的,$\boldsymbol{X}'_{S_n}$会逐渐靠近夜间真实的图像,所以它就会既有白天域特征又有夜间域特征,天然地拥有混合域的特性。这样$\boldsymbol{X}'_{S_n}$和$\boldsymbol{X}_{T_n}$合成的图像,模型前期可以先学习简单的$\boldsymbol{X}'_{S_n}$白天特征,再利用$\boldsymbol{X}'_{S_n}$的夜间域特征往$\boldsymbol{X}_{T_n}$真实夜间域靠近。尽管$\boldsymbol{X}'_{S_n}$后期夜间域特征占主导,但模型后期性能有了提升,可以学习比较困难的内容,$\boldsymbol{X}'_{S_n}$夜间域特征会继续促进模型往真实夜间域靠近,最终实现从白天域到夜间域的适应。又因为引入了$\boldsymbol{X}'_{S_n}$,使模型在风格域影响较小,模型在混合图像上就能逐渐地学习到数据分布域的差异,从而缓解传统ClassMix方法导致的域差距过大问题。

对于合成图像合理性问题,本文将道路场景的图像物体分为动态物体和静态物体,动态物体包括行人、骑手、汽车、卡车、公交车、火车、摩托车和自行车。可以发现,大部分类别都是可以在道路上出现的,而且模型训练采用的是驾驶数据集,道路类别作为数据中出现频率最高、像素总量最大的类别,其纹理也相对简单,这使得模型容易学习并实现较好的分割性能。因此,道路类别的重要程度是最低的,当动态物体移动到道路上,道路被覆盖后受到的影响也小。本文将问题简化,直接定义这些类别在合成图像中都出现在道路上。ClassMix方法混合图像时,需要有分割标签指定哪些像素是什么类别物体,同样的动态类别和公路类别也从标签里面筛选,然后计算位置偏移,将动态类别移动到公路上,形成比较合理的混合图像。考虑到让模型学习到更多知识并减小任务难度,本文在$\boldsymbol{X}'_{S_n}$只取动态类别不取静态类别,避免把静态物体也移动到公路上。然后互补地在$\boldsymbol{X}_{T_n}$取静态类别,将两者合成为有动态和静态类别的新图像,保证动静类别都存在,让模型充分学习到各种知识。

由于目标域是没有标签的,$\boldsymbol{X}_{T_n}$的标签只能通过伪标签获取。通过观察发现,Dark Zurich和ACDC数据集的$\boldsymbol{X}_{T_d}$和$\boldsymbol{X}_{T_n}$是同一地点不同时间拍摄的,所以它们大部分的静态部分是重合的,如图5(b)所示。借助此规律,可以认为模型对不同时间同一地点的图像预测得到伪标签$\bar{\boldsymbol{Y}}_{T_n}$和$\bar{\boldsymbol{Y}}_{T_d}$,它们重叠的静态部分可以认为是正确的标签,所以直接采用这种正确的静态伪标签作为$\boldsymbol{X}_{T_n}$的标签。

SDM详细过程如算法1所示。其中,$\boldsymbol{M}_s$由目标域白天预测$\bar{\boldsymbol{Y}}_{T_d}$和目标域夜间预测$\bar{\boldsymbol{Y}}_{T_n}$的one-hot编码取相同部分得到;staticbuff和dynbuff是大小50的数组。本文还需要设置一个阈值τ生成伪标签掩码$\boldsymbol{M}_c$,τ取0.2。SDM的具体流程如图6所示。大致分为三个步骤:首先,分割得到道路和动态物体的掩码或图像;接着,计算底部和中点坐标;最后,计算偏移量将SPG的动态物体移动到对应位置,即可得到移动后的图像。其中除了正常驾驶数据集的角度,我们还使用了其他角度的摄像头数据集[55]进行了可视化展示,可以看到SDM策略满足不同的摄像头角度。

混合后的图片进行分割和交叉熵损失计算,如图1所示,得到mix损失$\mathcal{L}_{\text{mix}}=\mathcal{L}_{\text{ce}}\left(f_\theta\left(\boldsymbol{X}'_{\boldsymbol{m}}\right),\boldsymbol{Y}'_{\boldsymbol{m}}\right)$。

由于SPG和SDM的相互依赖关系,它们之间也会受到对方的影响,即风格域适应和分布域适应也会相互影响。对于SDM而言,SPG模块翻译的夜间域图像将提供给SDM用于合成合理的图像进行训练,因此,SPG生成的夜间域图像越真实、信息越丰富,就越

算法 1 SDM

输入：$\boldsymbol{P}_{T_d}, \boldsymbol{P}_{T_n}, \boldsymbol{X}_{T_n}, \boldsymbol{X}'_{S_n}, \boldsymbol{Y}_S$

输出：$\boldsymbol{Y}'_m$

1. $\bar{\boldsymbol{Y}}_{T_d}=1(\boldsymbol{P}_{T_d})$; $\bar{\boldsymbol{Y}}_{T_n}=1(\boldsymbol{P}_{T_n})$ 1操作是将概率图变成one-hot标签

2. $\boldsymbol{M}_s=\left(\bar{\boldsymbol{Y}}_{T_d}=\bar{\boldsymbol{Y}}_{T_n}\right)$ 重叠的静态部分掩码 $\boldsymbol{M}_s$，SPG模块也需要使用

3. $\boldsymbol{M}_c=\dfrac{-\text{softmax}\left(\boldsymbol{P}_{T_n}\right)\log\left(\text{softmax}(\boldsymbol{P}_{T_n})\right)}{\log(C)}\leqslant\tau$ 归一化熵，C 是类别数，τ 是超参数

4. $\boldsymbol{M}_{s,c}=\boldsymbol{M}_s\mid\boldsymbol{M}_c$; $\boldsymbol{Y}'_{s,c}=\boldsymbol{M}_{s,c}\cdot\bar{\boldsymbol{Y}}_{T_n}$

5. $\boldsymbol{M}_d=\boldsymbol{Y}_S\geqslant 11\ \&\ \boldsymbol{Y}_S\leqslant 18$; $\boldsymbol{Y}'_d=\boldsymbol{M}_d\cdot\boldsymbol{Y}_S$; $\boldsymbol{X}'_d=\boldsymbol{M}_d\cdot\boldsymbol{X}'_{S_n}$; 11和18是动态类下标范围

6. $\hat{\boldsymbol{X}}_{s,c}, \hat{\boldsymbol{Y}}'_{s,c}=\text{staticbuff}(\boldsymbol{X}_{T_n}, \boldsymbol{Y}'_{s,c})$ 存入静态缓存器并随机取出一对图像和标签

7. $\hat{\boldsymbol{X}}'_d, \hat{\boldsymbol{Y}}'_d=\text{dynbuff}(\boldsymbol{X}'_d, \boldsymbol{Y}'_d)$ # 存入动态缓存器并随机取出一对图像和标签

8. dyn_xy = find_bottom_center $\left(\hat{\boldsymbol{Y}}'_d, \text{range}(11,19)\right)$ find_bottom_center根据类别计算整体的中间坐标和底部坐标

9. static_xy = find_bottom_center $(\hat{\boldsymbol{Y}}'_{s,c}, 0)$ 计算0类即“道路”这个类别的中间坐标和底部坐标

10. x, y = shift (dyn_xy, static_xy) 计算两者坐标的偏移量

11. $\tilde{\boldsymbol{X}}'_d, \tilde{\boldsymbol{Y}}'_d, \tilde{\boldsymbol{M}}_d=\text{roll}(\hat{\boldsymbol{X}}'_d, \hat{\boldsymbol{Y}}'_d, x, y)$ 根据偏移量将动态类别移动到“道路”上

12. $\boldsymbol{X}'_m=\text{Cover}(\boldsymbol{X}_{T_n}\cdot\sim\tilde{\boldsymbol{M}}_d+\tilde{\boldsymbol{X}}'_d)$ Cover操作为随机覆盖一部分像素

13. $\boldsymbol{Y}'_m=\hat{\boldsymbol{Y}}'_{s,c}\cdot\sim\tilde{\boldsymbol{M}}_d+\tilde{\boldsymbol{Y}}'_d$

能促进缩小风格域的差异，从而有利于SDM的分布域适应。反之，翻译的图像不正确，将会使得SDM的训练变得更加困难，效果也会随之下降。同样，对于SPG而言，SDM提供的掩码 $\boldsymbol{M}_s$ 可用信息很少或者不正确信息太多，就会导致SPG粗配对翻译训练有误，进而影响SPG的翻译过程。反之，SDM提供的信息越准确，会促进SPG翻译得越好。

2.6 目标函数

本文使用 $\mathcal{L}_{\text{SPG}}$ 约束生成的图像，使网络从白天域转移到夜间域，使用 $\mathcal{L}_{\text{mix}}$ 从分布层面进行域适应，使用 $\mathcal{L}_{\text{source}}$ 在源域有监督的训练，$\mathcal{L}_{\text{night}}$ 则是使用生成图像有监督的训练。综上所述，整个网络的损失表示为

$$\mathcal{L}=\lambda_3\mathcal{L}_{\text{mix}}+\mathcal{L}_{\text{SPG}}+\mathcal{L}_{\text{source}}+\mathcal{L}_{\text{night}} \tag{8}$$

其中，λ_3 是超参数，取值为 $\lambda_3=0.8$。

3 实验

对于所有实验，本文采用所有类别的交并比（IoU）的平均值（mIoU）作为评估指标，并将3.1节的数据集用于模型训练和性能评估。

3.1 数据集

Cityscapes[56]是城市道路场景的语义分割数据集，其中包含2 975张分辨率为2 048 × 1 024的训练图像。该数据集只包含白天场景，并且每张图像都有对应的标签，这些数据作为本文方法中源域数据部分，用于有监督的训练。

Dark Zurich[31]是一个由苏黎世城市道路场景图像组成的数据集，其中包含3 041张白天图像和2 416张夜间图像，分辨率为1 920 × 1 080，这些图像是在相同位置下于不同光照条件下拍摄的。本文去除了重复图像，仅选取1 426对未标注的日夜图像作为本文方法的目标域数据部分进行无监督训练，在验证和测试阶段则使用数据集提供的50张有标注的验证图像进行验证，151张未标注的测试图像用于在线测试。

ACDC[54]数据集包含4 006张图像，包括了四种不良环境条件（雾、雨、雪和夜间）。其中，夜间图像

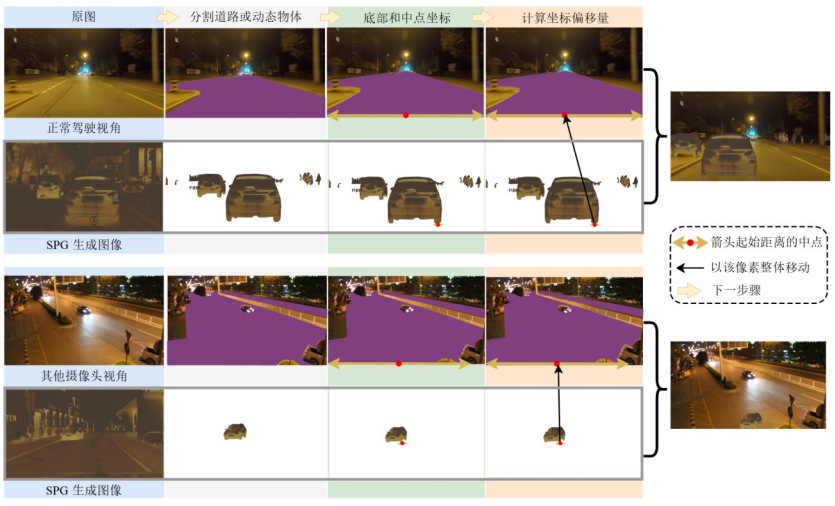


图6 SDM具体流程

Figure 6 SDM specific process

有 1 006 张,分辨率为 1 920 × 1 080,标记为 ACDC Night,且在相同位置下的白天图像也有 1 006 张。在使用 ACDC 数据集进行训练时,从训练集中选取 400 对未标注的日夜图像作为本文方法中的目标域数据部分进行无监督训练。在验证和测试阶段则使用数据集提供的 106 张有标注的验证图像进行验证,并用 500 张未标注的测试图像进行在线测试。除了在模型对比实验 3.3.1 节中额外使用了 400 张图像和标签监督训练,其余实验均不使用标签训练。

Nighttime Driving[30]夜间驾驶数据集包含 50 张有标注的 1 920 × 1 080 分辨率的夜间驾驶场景图像,在本文中仅用于测试评估本文网络的域适应能力。

3.2　实验细节

本文基于 mmsegmentation 框架[57]实现本文的方法,选择 VMamba[50]作为网络的骨干网络,在 ImageNet 数据集上进行了预训练。模型采用 AdamW[58]优化器进行训练,学习率为 6×10^{-5},权重衰减设置为 1×10^{-2},并采用线性学习率预热,预热比例为 1×10^{-6},批量大小设置为 2。其中,SPG 模块训练时采用 AdamW 优化器,学习率为 1×10^{-4}。在训练过程中,使用大小为 256 × 512 的随机裁剪,并应用随机水平翻转,总训练迭代次数设置为 80 k。

3.3　实验结果分析

3.3.1　夜间分割模型比较

首先将 RefineNet[25]、DeepLabv2[26]、MiT[20]和 VMamba[50]模型在 Cityscapes 数据集上进行监督训练,在 Cityscapes 验证集上进行评估,从表 1 可以发现,VMamba 模型性能断层领先,在正常照度下的分割性能非常优秀。接着,将这些模型在 ACDC Night 数据集上进行有标签的监督训练,在 Dark Zurich 和 Nighttime Driving 数据集上进行泛化性评估,结果如表 1 所示。因为在有监督的训练实验中,可以排除无监督难收敛的影响,直接得到模型对夜间分割能力的表现。从表 1 中可以得到,VMamba 模型整体性能效果最好,分别在 ACDC Night 和 Nighttime Driving 上取得最优,Dark Zurich 次优,证明其在夜间分割任务中是最有潜力的。正常照度数据训练的模型性能排行和夜间监督训练的模型性能排行大体一致,可以得出在正常照度下模型的性能越好,夜间的分割性能也会越好。

3.3.2　Dark Zurich

在表 2 中比较了一些现有最先进的方法,包括 GCMA[31]、DAFormer[19]、DTBS[21]、HRDA[59]、PIG[60]等方法,以及其他方法[8,17,32,61]。本文的方法在 Dark Zurich Test 的 mIoU 指标上达到了 60.0%,此外,本文的策略在一些类别得到很大地改善,特别是在行人、交通信号标、交通信号灯和公共汽车等类别。其中,公共汽车这个类别的性能提升最为明显,比次优值还要高 18.1 个百分点,验证了 SDM 动态和静态混合图像的有效性,通过动态和静态结合使模型不偏向某一类别,让模型学习到更多类别。也因为 SDM 的移动策略,导致植物和围墙等物体的性能不好,相较于最高值分别低了 4.5 个百分点和 12.6 个百分点。因为移动策略无法检测到动态物体的大小,虽然能保证动态物体移动到道路上,但动态物体太大,可能会遮住植物和围墙等物体,导致这类物体学习到的内容较少,性能较弱。如图 7 所示,图 7(a)中大物体会覆盖到其他类别;图 7(b)中小物体只在道路内部不会覆盖到其他类别;图 7(c)中差异比较大的物体存在时,大物体出现覆盖情况,小物体虽然也有些覆盖但覆盖范围小;图 7(a)和(c)的可视化也体现了存在植物和围墙这些类别被覆盖的情况。

表 1　各种模型在验证集上分割结果对比　　单位:%

Table 1　Comparison of segmentation results of various models on the validation set　　unit: %

方法	mIoU ↑			
	Cityscapes	ACDC Night	Dark Zurich	Nighttime Driving
MiT-B5[20]	68.87*	34.83	32.94	34.88
DeepLabv2[26]	74.90*	46.26	41.74	44.49
RefineNet[25]	<u>76.17*</u>	<u>50.50</u>	**51.83**	<u>46.44</u>
VMamba-s[50]	**81.57***	**54.63**	<u>49.27</u>	**53.69**

注:加粗表示最优值,下划线表示次优值。*由于所有方法的参数需要统一,因此本文对该方法进行了重新训练得到的结果。

为了证明本文模型的泛化性,在表 3 中,本文在 Cityscapes→Dark Zurich 训练的模型直接进行 ACDC Night Test 的评估,本文的方法在 mIoU 指标上达到了 59.5%,并且很多类别精度取得了最优或次优。在表 4[17,19,22,59,61-62]进行了 Nighttime Driving Test 的评估,本文的方法在 mIoU 指标上达到了 59.1%。表 3 和表 4 说明,本文的方法在没有训练过的数据集上具有一定的泛化能力。

本文在 Dark Zurich Val 进行了分割结果可视化,如图 8 所示。可以看到,本文的方法在这些图片中都能够识别到行人,而其他方法几乎没有识别到。其中第三行图片中的建筑有个发光的牌子,标签里面也将其标记为建筑的一部分,只有本文的方法能将其识别为建筑的一部分,其他方法都将这个发光的牌子错误地识别为了交通信号灯或交通信号标。

3.3.3　ACDC

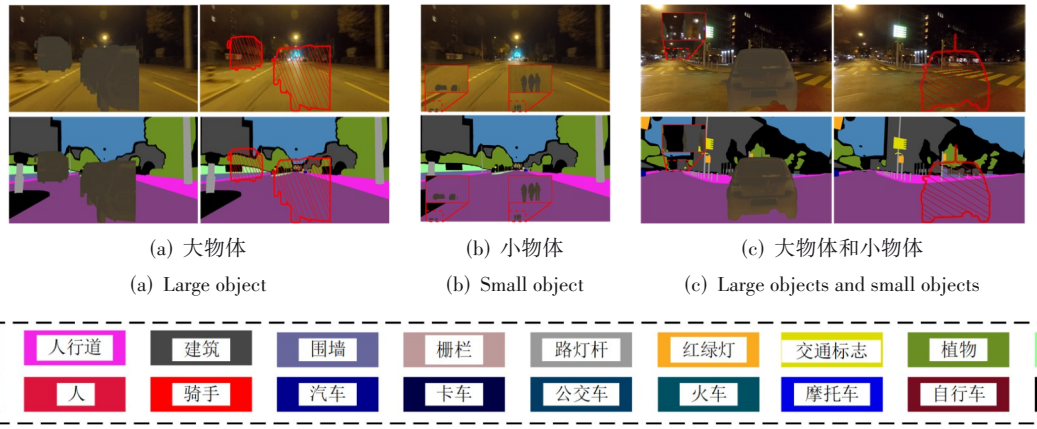
在表 5 中,通过与其他域适应方法进行比较,包括 DAFormer[19]、DTBS[21]、HRDA[59]、CISS[63]等[8,61,60]方法,在 ACDC Night Test 数据集上进行测试。本文的方法 mIoU 达到了 59.8%,优于先进方法 CISS[63]。从表 5

表 2 与 Cityscapes→Dark Zurich 域适应任务的最先进方法在 Dark Zurich Test 上的性能比较 单位:%

Table 2 Performance comparison with the state of the art method on the Cityscapes→Dark Zurich domain adaptation task on the Dark Zurich Test unit: %

方法	道路	人行道	建筑	围墙	栅栏	路灯杆	红绿灯	交通标志	植物	绿化带	天空	人	骑手	汽车	卡车	公交车	火车	摩托车	自行车	mIoU ↑
GCMA[31]	81.7	46.9	58.8	22.0	20.0	41.2	40.5	41.6	64.8	31.0	32.1	53.5	47.5	75.5	39.2	0.0	49.6	30.7	21.0	42.0
MGCDA[32]	80.3	49.3	66.2	7.8	11.0	41.4	38.9	39.0	64.1	18.0	55.8	52.1	53.5	74.7	66.0	0.0	37.5	29.1	22.7	42.5
DANNet[17]	90.0	54.0	74.8	41.0	21.1	25.0	26.8	30.2	72.0	26.2	84.0	47.0	33.9	68.2	19.0	0.3	66.4	38.3	23.6	44.3
DAFormer[19]	**93.5**	65.5	73.3	39.4	19.2	53.3	44.1	44.0	59.5	34.5	66.6	53.4	52.7	82.1	52.7	9.5	89.3	50.5	38.5	53.8
SePiCo[61]	93.2	68.1	73.7	32.8	16.3	54.6	49.5	48.1	**74.2**	31.0	**86.3**	57.9	50.9	82.4	52.2	1.3	83.8	43.9	29.8	54.2
DTBS[21]	93.1	62.8	76.9	39.5	17.9	51.9	29.0	40.0	63.9	27.4	77.7	56.0	53.6	78.0	**81.3**	13.9	90.1	44.8	40.2	54.6
HRDA[59]	90.4	56.3	72.0	39.5	19.5	**57.8**	52.7	43.1	59.3	29.1	70.5	60.0	58.6	84.0	75.5	11.2	90.5	51.6	**40.9**	55.9
VSA-DANet[8]	91.3	58.6	76.6	**46.0**	**24.6**	52.7	47.0	49.4	64.5	**39.4**	78.7	55.6	52.6	79.9	75.6	0.0	**90.6**	**55.4**	38.3	56.7
PIG[60]	91.8	**73.3**	73.4	43.6	20.8	57.7	49.4	54.3	71.7	38.1	80.5	58.7	56.5	82.4	80.7	17.3	89.9	41.8	40.4	59.1
SDDA(ours)	93.3	68.1	**78.6**	33.4	21.5	57.2	**56.3**	**60.4**	69.7	27.2	80.2	**65.5**	**61.7**	**86.1**	76.4	**35.4**	88.3	42.0	38.2	**60.0**

注:加粗表示最优值,下划线表示次优值。



(a) 大物体 (b) 小物体 (c) 大物体和小物体

(a) Large object (b) Small object (c) Large objects and small objects

(d) 图例

(d) legend

图 7 不同尺寸物体覆盖情况

Figure 7 Coverage of objects of different sizes

表 3 与其他域适应方法在零样本泛化到 ACDC Night Test 的性能比较 单位:%

Table 3 Performance comparison with other domain adaptation methods on zero-shot generalization to ACDC Night Test unit: %

方法	道路	人行道	建筑	围墙	栅栏	路灯杆	红绿灯	交通标志	植物	绿化带	天空	人	骑手	汽车	卡车	公交车	火车	摩托车	自行车	mIoU ↑
GCMA[31]	78.6	45.9	58.5	17.7	18.6	37.5	43.6	43.5	58.7	39.2	22.4	57.9	29.9	72.1	21.5	56.2	41.8	35.7	35.4	42.9
MGCDA[32]	74.5	52.5	69.4	7.7	10.8	38.4	40.2	43.3	61.5	36.3	37.6	55.3	25.6	71.2	10.9	46.4	32.6	27.3	33.8	40.8
DANNet[17]	90.7	61.1	75.5	35.9	**28.8**	26.6	31.4	30.6	**70.8**	39.4	**78.7**	49.9	28.8	65.9	24.7	44.1	61.1	25.9	34.5	47.6
DAFormer[19]	91.5	61.9	67.7	30.9	15.0	44.6	43.3	40.0	55.2	41.4	44.6	54.1	31.9	74.7	9.1	44.8	83.3	38.1	45.0	48.3
HRDA[59]	87.5	48.1	77.6	**43.2**	23.2	51.1	53.2	50.2	54.1	35.8	55.6	63.2	40.4	80.7	**63.5**	81.8	80.6	46.0	49.5	57.1
SDDA(ours)	**92.8**	**68.4**	**80.2**	28.9	21.0	**51.8**	**56.6**	**60.5**	66.6	**45.2**	74.8	**68.6**	**44.5**	**83.2**	18.9	**84.0**	**85.0**	**46.3**	**53.2**	**59.5**

注:加粗表示最优值,下划线表示次优值。

中观察到,本文的方法在某些类别表现较好,如天空、骑手、建筑和公共汽车。实验结果证明了本文方法的优越性,在不同的数据集训练时,一样能达到很好的效果。

本文在 ACDC Night Val 进行了分割结果的可视化,如图 9 所示。可以看到,本文的方法在 ACDC

表4　与Cityscapes→Dark Zurich域适应任务的方法在Nighttime Driving Test上的性能比较　　单位:%

Table 4　Performance comparison with the Cityscapes→Dark Zurich domain adaptation task method on Nighttime Driving Test　unit: %

方法	mIoU ↑
DANNet[17]	47.7
DAFormer[19]	51.8
SePiCo[61]	56.9
HRDA[59]	57.7
MIC[22]	58.7
DPC[62]	58.1
SDDA(ours)	**59.1**

注:加粗表示最优值,下划线表示次优值。

Night数据集中一样可以识别到行人,但其他方法几乎不能识别到,并且可以看到DAFormer和HRDA这些方法几乎将天空都错误地识别为绿植,而本文的方法却能将天空准确识别。

3.4　消融实验

3.4.1　网络组件消融

为了验证不同组件的有效性,本文在Cityscapes→Dark Zurich任务上进行消融实验,所有实验采用相同的设置和超参数。本节实验的基线为VMamba[50],结果如表6所示,与基线模型相比,当采用SPG模块时,Dark Zurich、ACDC Night和Nighttime Driving数据集上的性能分别提升了6.85、6.56和9.66个百分点。说明SPG生成的图像有利于模型从源域适应到目标域。最后加上SDM策略时,模型整体较基线提升了17.88、19.54和18.43个百分点。

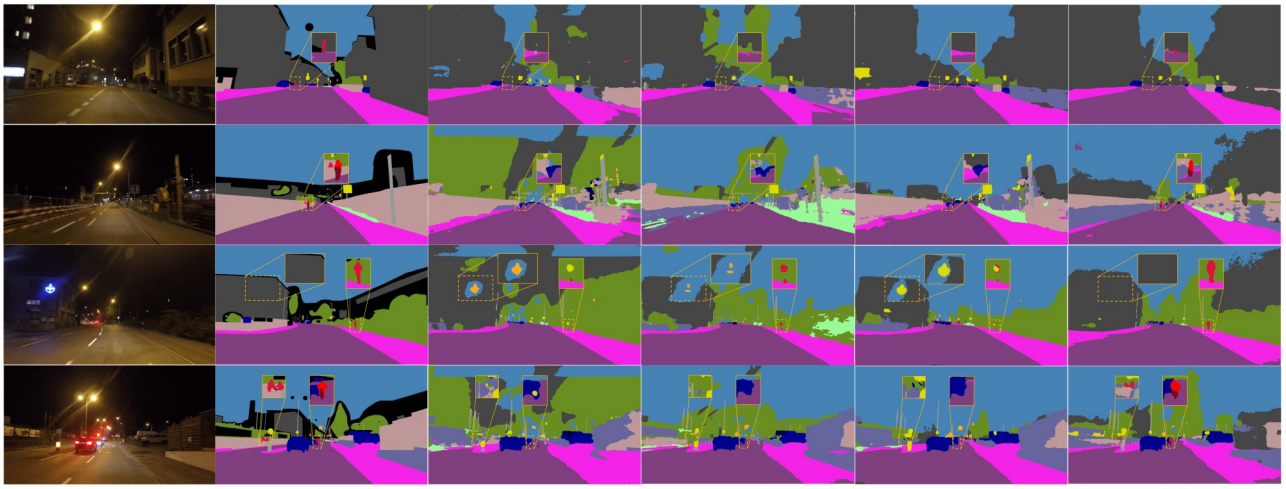


图8　不同方法在Dark Zurich Val上的分割结果的可视化比较

Figure 8　Visual comparison of segmentation results of different methods on Dark Zurich Val

表5　与Cityscapes→ACDC Night域适应任务的方法在ACDC Night Test上的性能比较　　单位:%

Table 5　Performance comparison with the Cityscapes→ACDC Night domain adaptation task method on ACDC Night Test　unit: %

方法	道路	人行道	建筑	围墙	栅栏	路灯杆	红绿灯	交通标志	植物	绿化带	天空	人	骑手	汽车	卡车	公交车	火车	摩托车	自行车	mIoU ↑
DAFormer[19]	92.3	64.6	70.1	28.7	18.5	45.8	11.3	41.5	42.7	41.9	0.0	55.4	29.8	74.3	40.3	45.8	81.3	39.4	47.0	45.8
HRDA[59]	87.2	46.9	79.1	46.2	18.0	51.4	41.0	48.5	41.8	46.7	0.0	63.2	36.9	81.0	**65.2**	77.7	83.6	46.0	49.0	53.1
SePiCo[61]	89.9	56.8	75.6	35.3	28.4	49.5	24.7	50.1	43.4	44.5	4.8	61.1	34.1	77.3	62.0	52.9	79.5	41.2	48.3	50.5
DTBS[21]	92.8	66.1	75.6	37.5	27.4	46.4	17.6	44.3	60.0	40.8	60.7	58.3	31.4	77.5	38.6	62.6	83.9	44.6	46.2	53.3
PIG[60]	91.9	70.8	81.3	44.7	13.9	50.5	44.6	51.8	68.8	45.7	78.6	62.1	42.2	76.6	41.7	63.6	78.7	25.2	48.4	56.9
VSA-DANet[8]	89.2	55.5	78.7	44.1	**30.0**	46.8	52.0	49.3	66.1	38.3	75.8	58.5	37.8	78.8	57.1	71.7	72.6	50.4	46.4	57.8
CISS[63]	**94.7**	**74.5**	81.2	**48.2**	28.4	**52.2**	50.1	**58.6**	43.2	**53.4**	2.6	**65.7**	39.0	**83.8**	63.2	74.7	**86.6**	**52.9**	**53.5**	58.2
SDDA(ours)	93.8	70.6	**83.6**	33.9	18.0	50.3	**54.1**	58.4	**69.7**	42.5	**80.3**	64.7	**44.8**	82.6	24.6	**79.4**	84.6	49.2	50.5	**59.8**

注:加粗表示最优值,下划线表示次优值。

本文还在图10展示了加入SPG模块时,生成的源域夜间图像与其他真实数据集的T-SNE分布情况。其中CityDay是Cityscapes,CityNight是Cityscapes经过SPG生成的夜间图像,DarkZurichDay是DarkZurich白

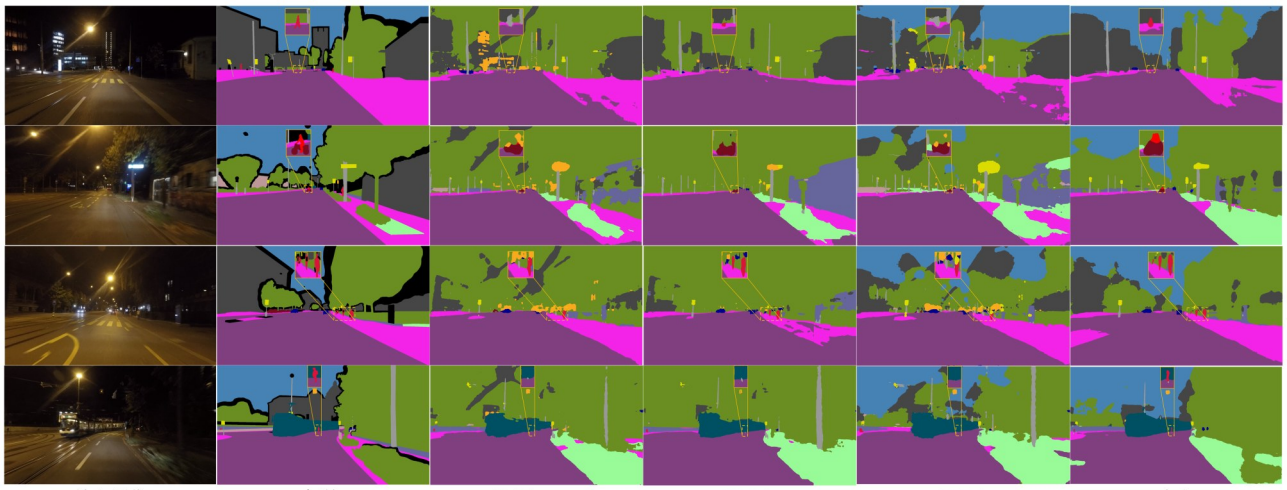


图9 不同方法在ACDC Night Val上的分割结果的可视化比较

Figure 9 Visual comparison of segmentation results of different methods on ACDC Night Val

表6 消融实验在Dark Zuirch Val、ACDC Night Val和Nighttime Driving Test数据集上的分割结果对比 单位:%

Table 6 Comparison of segmentation results of ablation experiments on Dark Zuirch Val, ACDC Night Val, and Nighttime Driving Test datasets unit: %

Baseline	SPG	SDM	mIoU ↑		
			Dark Zurich	ACDC Night	Nighttime Driving
√			26.03	29.96	40.73
√	√		<u>32.88</u>	<u>36.52</u>	<u>50.39</u>
√	√	√	**43.91**	**49.50**	**59.16**

注:加粗表示最优值,下划线表示次优值。

天图像,DarkZurichNight是DarkZurich夜间图像。观察到CityNight是远离了CityDay,往DarkZurichNight靠近,说明SPG生成的夜间图像接近于DarkZurichNight,证明了SPG的有效性,而且也证明本文在SDM策略中使用的CityNight包含白天和夜间域的信息,可以促进模型学习到这种混合域知识。

3.4.2 SPG的组件消融

本文对SPG组件进行消融实验,所有实验在Cityscapes→Dark Zurich训练。如表7所示,非配对翻译:加入生成对抗GAN进行风格学习时性能提升明显,证明GAN非配对翻译学习到了风格域信息,并且由于其与语义结合,也促进了分割网络学习到了风格域内容。粗配对翻译:当只加入cosine损失时,Dark Zurich和ACDC Night的精度都有上升,Nighttime Driving性能相当,整体性能上升。证明了本文假设模型中间的特征应该越远离原来的域而靠近目标域的有效性,有助于模型特征层面的翻译。单独加入pair损失之后,在ACDC Night数据集上略有提升,说明粗配对翻译的有效性,两者都使用时,模型性能整体最优。

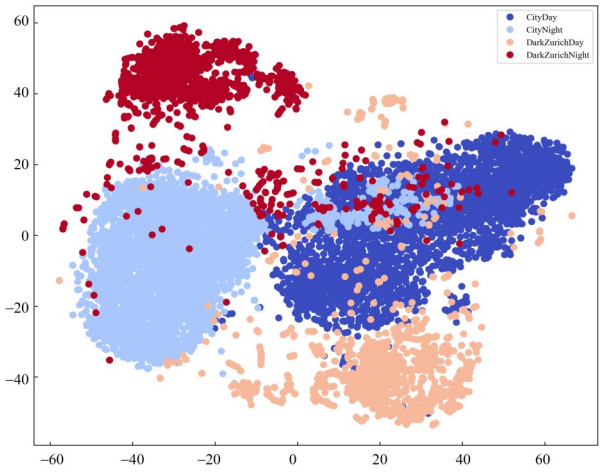


图10 T-SNE可视化各种数据集分布结果

Figure 10 T-SNE visualizes various dataset distribution results

表7 对SPG的组件进行有效性验证 单位:%

Table 7 The components of the SPG are verified for validity unit: %

Baseline	GAN	cosine	pair	mIoU ↑		
				Dark Zurich	ACDC Night	Nighttime Driving
√				26.03	29.96	40.73
√	√			30.18	33.51	**51.29**
√	√	√		<u>32.05</u>	35.64	49.53
√	√		√	30.17	<u>35.67</u>	<u>51.00</u>
√	√	√	√	**32.88**	**36.52**	50.39

注:加粗表示最优值,下划线表示次优值。

3.4.3 SDM的组件消融

本文对SDM进行消融实验,所有实验在Cityscapes→Dark Zurich训练。如表8所示,当生成的图像不合理和存在域差距过大时,就是普通的Class-

Mix,其性能是下降的。当将动态物体移动到道路上或者采用SPG生成的图像时,可以使模型性能提升,两者都采用时,模型整体性能最优。

表8　SDM策略消融　　单位:%

Table 8　SDM policy ablation　　unit: %

SDM	没有SPG图像	没有在道路上	mIoU ↑		
			Dark Zurich	ACDC Night	Nighttime Driving
√	√	√	42.23	45.22	53.99
√	√		42.47	48.21	53.99
√		√	42.48	48.09	53.73
√			**43.91**	**49.50**	**59.16**

注:加粗表示最优值,下划线表示次优值。

3.4.4　SDM与其他Mix方法对比

在表9中,SDM方法与其他Mix方法进行比较,所有实验在Cityscapes→Dark Zurich训练。本文的SDM方法在三个数据集的验证集上性能都是最优,证明SDM的设计能有效解决其他Mix方法的合成图像不合理以及合成图像跨域幅度过大的问题。

表9　SDM与其他Mix方法对比　　单位:%

Table 9　Comparison of SDM with other Mix methods　　unit: %

方法	mIoU ↑		
	Dark Zurich	ACDC Night	Nighttime Driving
Mixup*	36.13	36.22	48.02
Mixcut*	22.46	23.20	38.00
Classmix*	26.23	30.33	46.74
SDM	**43.91**	**49.50**	**59.16**

注:加粗表示最优值,下划线表示次优值。*由于所有方法需要适配SDDA的架构,因此本文对该方法进行了重新训练得到的结果。

3.4.5　超参数$\lambda_1,\lambda_2,\lambda_3,\tau$消融

对本文方法中采用的损失的超参数$\lambda_1,\lambda_2,\lambda_3$进行分析,分析它们对模型性能的影响。如表10所示,所有实验在Cityscapes→Dark Zurich训练。通过控制一个参数的改变,然后进行性能测试。由于三者相互作用和影响,每个参数改变时,不同数据集的性能表现都不太一样,结果呈现一定的随机性。只从Dark Zurich数据集上看,大致可以看到,λ_3、λ_2任意取值都表现得平滑,λ_1= 0.5时性能最优;当三者都取0.5时,模型性能次优;最终结果表明,当λ_3= 0.8,λ_2= 0.1,λ_1= 0.1时,性能最好。

超参数τ是伪标签质量的一个阈值,它越小证明伪标签质量越高,但质量高的伪标签就越少,模型可能已经学习到对这部分高质量的预测了,对模型的作用也就越小。它越大错误的信息也会越多,导致模型学习到错误知识影响模型性能。本文对其取值进行了分析,如表11所示,所有实验在Cityscapes→Dark Zurich训练。当τ=0.1时,模型可以学习到正确的知识;当τ=0.2时,正确知识和错误知识得到权衡,此时模型整体最佳;当τ从0.3到0.5时,模型性能逐渐下降。

表10　$\lambda_1,\lambda_2,\lambda_3$定量分析实验结果

Table 10　$\lambda_1,\lambda_2,\lambda_3$ quantitative analysis of experimental results

λ_3	λ_1	λ_2	mIoU ↑ /%		
			Dark Zurich	ACDC Night	Nighttime Driving
1.0	1.0	1.0	42.6	49.6	56.5
0.5	1.0	1.0	42.2	45.3	54.2
0.8	1.0	1.0	42.1	47.9	56.1
1.0	0.5	1.0	43.3	46.8	54.4
1.0	0.1	1.0	42.3	44.9	56.5
1.0	1.0	0.5	42.1	47.2	57.1
1.0	1.0	0.1	42.4	45.3	56.1
0.5	0.5	0.5	41.3	49.6	59.0
0.8	0.5	0.1	42.3	**50.2**	56.5
0.8	0.1	0.1	**43.9**	49.5	**59.1**

注:加粗表示最优值,下划线表示次优值。

表11　τ定量分析实验结果

Table 11　τ quantitative analysis of experimental results

τ	mIoU ↑ /%		
	Dark Zurich	ACDC Night	Nighttime Driving
0.5	34.88	41.15	53.16
0.4	38.50	44.85	54.15
0.3	42.18	**49.68**	54.26
0.2	**43.91**	49.50	**59.16**
0.1	43.09	49.20	56.12

注:加粗表示最优值,下划线表示次优值。

3.4.6　staticbuff和dynbuff存储大小消融

对本方法中采用的缓冲区(staticbuff和dynbuff)的存储大小进行了分析,两者大小一致。缓冲区将一段时间内模型输出的内容进行存储,相较于直接使用模型即时预测作为伪标签,这种方式更加稳定。因为在训练过程中,模型输出的内容可能会受到错误的影响,而通过存储历史输出信息,随机取出,可以确保模型学习到之前状态的重要信息而非当前预测的内容,从而防止性能的过度恶化。如表12所示,所有实验在Cityscapes→Dark Zurich训练,缓冲区的存储容量越大,有助于模型的稳定运行,能够学习到更多的正确信息。达到70左右效果不会继续增加,因为存储的是模型的之前状态,存储越多只会缓解模型学习错误信息,不会一直增加。但越小的话,之前状态的信息就会变少,无法有效减少模型错误学习。

表 12 缓冲区大小定量分析实验结果

Table 12 Buffer size quantitative analysis of experimental results

存储大小	mIoU ↑ /%		
	Dark Zurich	ACDC Night	Nighttime Driving
12	41.61	47.41	54.11
25	41.79	<u>48.52</u>	54.80
50	**43.91**	**49.50**	**59.16**
75	<u>42.25</u>	47.18	<u>56.10</u>

注:加粗表示最优值,下划线表示次优值。

3.4.7 分割网络消融

为了验证本文方法在不同的分割网络都有效,本文选取了一些具有代表性的语义分割网络[20,25-26],它们分别采用的分割网络架构为 ResNet 和 Transformer。这些模型在 Cityscapes 数据集上训练,然后对比使用了本文的方法进行训练的模型性能。如表 13 所示,在其他分割网络上对本文方法进行有效性验证,所有非域适应方法都是在 Cityscapes 上进行训练,域适应方法在 Cityscapes→Dark Zurich 训练,所有实验参数设置都一样,在 Dark Zuirch Val、ACDC Night Val 和 Nighttime Driving Test 数据集上的分割结果对比。在使用了本文方法之后,不同的分割网络性能都有一定的提升,证明本文方法让模型学习到的知识是通用的,不依赖于分割网络,但由于分割网络本身的能力上限,最终不如 VMamba 的网络。同样证明了 VMamba 在语义分割潜能和模型的能力决定了域适应方法的能力。本文方法在推理阶段是没有使用 SPG、SDM 这些模块和策略的,与 Baseline 的结构是一样的,也就是说本文方法提升的是 Baseline 性能,而不是依赖这些模块才有的性能提升,因此推理时不会增加额外参数和时间。

表 13 与其他分割网络上对本文的方法进行有效性验证 单位:%

Table 13 The effectiveness of the method in this paper is verified with other segmentation networks unit: %

方法	UDA	Backbone	mIoU ↑					
			Dark Zurich	ΔmIoU	ACDC Night	ΔmIoU	Nighttime Driving	ΔmIoU
MiT-B5[20]		Transformer	8.74	—	10.81	—	19.07	—
Ours(MiT)	√		22.43	+13.69	23.42	+12.61	30.08	+11.01
DeepLabv2[26]		ResNet	8.47	—	9.99	—	16.03	—
Ours(DeepLabv2)	√		23.41	+14.94	25.08	+15.09	35.98	+19.95
RefineNet[25]		ResNet	14.66	—	15.11	—	23.3	—
Ours(RefineNet)	√		27.05	+12.39	27.65	+12.54	32.94	+9.64
VMamba-s[50]		VMamba	26.03	—	29.96	—	40.73	—
SDDA	√		43.91	+17.88	49.5	+19.54	59.16	+18.43

4 结束语

本文提出了一种基于风格和分布域适应的无监督夜间语义分割方法,首先将更适用于夜间分割任务的 VMamba 模型引入,然后采用 SPG 模块生成夜间图像训练网络,以此缩小模型的风格域差异,最后采用 SDM 策略生成更加合理的混合域图像进行训练,将模型的分布域差异缩小。风格和分布域适应分别从不同角度进行缩小差距,最终将模型从源域迁移到目标域,实验结果也证明了本文方法的有效性。然而,在 SDM 策略中,源域的动态物体的尺寸可能存在较大差异。较大的动态物体移动到特定位置可能会覆盖其他静态类别,导致模型无法学习这部分内容;较小的动态物体因像素数量太少,很难为模型提供有效信息。在未来的研究中,我们将进一步完善 SDM 策略,尝试将尺寸差异大的物体分开处理。对于造成遮挡的较大的动态物体,可以适当缩小分辨率以避免覆盖其他静态物体;对于像素太少的较小的物体,可以适当放大分辨率并结合多尺度方法,为模型提供更多可学习信息。

参考文献

[1] Fu Y X, Lou M, Yu Y Z. SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2025: 19077-19087.

[2] Chen L W, Fu Y, Gu L, et al. Spatial frequency modulation for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(11): 9767-9784.

[3] Xu Z Z, Wu D Y, Yu C Q, et al. SCTNet: Single-branch CNN with transformer semantic information for real-time segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(6): 6378-6386.

[4] Zhao Z, Long S F, Pi J M, et al. Instance-specific and mod-

el-adaptive supervision for semi-supervised semantic segmentation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 23705-23714.

[5] 杨维静, 徐瑞, 顾浩文, 等. 基于伪标签去噪和SAM优化的大规模无监督语义分割[J]. 电子学报, 2025, 53(3): 716-727.
Yang Weijing, Xu Rui, Gu Haowen, et al. Pseudo-label denoising and SAM optimization for large-scale unsupervised semantic segmentation[J]. Acta Electronica Sinica, 2025, 53(3): 716-727. (in Chinese)

[6] Sun L, Wang K W, Yang K L, et al. See clearer at night: Towards robust nighttime semantic segmentation through day-night image conversion[C]//Artificial Intelligence and Machine Learning in Defense Applications. SPIE, 2019: 2532477.

[7] Shen Z Q, Huang M Y, Shi J P, et al. Towards instance-level image-to-image translation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 3678-3687.

[8] 江泽涛, 廖培期, 黄钦阳, 等. 基于向量符号架构-域适应网络的低照度图像语义分割方法[J]. 计算机辅助设计与图形学学报, 2025, 37(8): 1371-1382.
Jiang Z T, Liao P Q, Huang Q Y, et al. Low-light image semantic segmentation method based on vector symbolic architecture-domain adaptation network[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(8): 1371-1382. (in Chinese)

[9] 江泽涛, 朱文才, 金鑫, 等. 一种基于双重语义协作网络的图像描述方法[J]. 计算机研究与发展, 2024, 61(11): 2897-2908.
Jiang Zetao, Zhu Wencai, Jin Xin, et al. An image captioning method based on DSC-net[J]. Journal of Computer Research and Development, 2024, 61(11): 2897-2908. (in Chinese)

[10] 江泽涛, 翟丰硕, 钱艺, 等. 结合特征增强和多尺度感受野的低照度目标检测[J]. 计算机研究与发展, 2023, 60(4): 903-915.
Jiang Zetao, Zhai Fengshuo, Qian Yi, et al. Low illumination object detection combined with feature enhancement and MultiScale receptive field[J]. Journal of Computer Research and Development, 2023, 60(4): 903-915. (in Chinese)

[11] Yang G L, Zhong Z, Tang H, et al. Bi-mix: Bidirectional mixing for domain adaptive nighttime semantic segmentation[PP/OL]. V1.arXiv (2021-11-19)[2026-01-25]. https://doi.org/10.48550/arXiv.2111.10339.

[12] Luo X Y, Zhang J M, Yang K L, et al. Towards robust semantic segmentation of accident scenes via multi-source mixed sampling and meta-learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2022: 4428-4438.

[13] Pan J Y, Li S H, Chen Y C, et al. Towards dynamic and small objects refinement for unsupervised domain adaptative nighttime semantic segmentation[C]//2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE, 2024: 2720-2727.

[14] Sankaranarayanan S, Balaji Y, Jain A, et al. Learning from synthetic data: Addressing domain shift for semantic segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3752-3761.

[15] Zhang Y H, Qiu Z F, Yao T, et al. Fully convolutional adaptation networks for semantic segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6810-6818.

[16] Gao H, Guo J C, Wang G L, et al. Cross-domain correlation distillation for unsupervised domain adaptation in nighttime semantic segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9903-9913.

[17] Wu X Y, Wu Z Y, Guo H, et al. DANNet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 15764-15773.

[18] Liu W Y, Li W T, Zhu J K, et al. Improving nighttime driving-scene segmentation via dual image-adaptive learnable filters[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(10): 5855-5867.

[19] Hoyer L, Dai D X, Van Gool L. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9914-9925.

[20] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[C/OL]//The 35th Conference on Neural Information Processing Systems (NeurIPS 2021). 2021: 12077-12090. https://proceedings. neurips. cc/paper_files/paper/2021/file/64f1f27bf1b4ec22924fd0acb550c235-Paper.pdf.

[21] Huang F D, Yao Z H, Zhou W H. DTBS: Dual-teacher bi-di-

rectional self-training for domain adaptation in nighttime semantic segmentation[PP/OL]. V1. arXiv (2024-01-02) [2026-01-25]. https://doi.org/10.48550/arXiv.2401.01066.

[22] Hoyer L, Dai D X, Wang H R, et al. MIC: Masked image consistency for context-enhanced domain adaptation[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 11721-11732.

[23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3431-3440.

[24] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[M]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.

[25] Lin G S, Milan A, Shen C H, et al. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5168-5177.

[26] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[27] Zheng S X, Lu J C, Zhao H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 6877-6886.

[28] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9992-10002.

[29] Yu W H, Wang X C. MambaOut: Do we really need mamba for vision [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2025: 4484-4496.

[30] Dai D X, Van Gool L. Dark model adaptation: Semantic image segmentation from daytime to nighttime[C]//2018 21st International Conference on Intelligent Transportation Systems. Piscataway: IEEE, 2018: 3819-3824.

[31] Sakaridis C, Dai D X, Van Gool L. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 7373-7382.

[32] Sakaridis C, Dai D X, Van Gool L. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3139-3153.

[33] Wu X, Lai Z H, Yu S Q, et al. Coarse-to-fine low-light image enhancement with light restoration and color refinement[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(1): 591-603.

[34] Wu X, Lai Z H, Zhou J, et al. Light-aware contrastive learning for low-light image enhancement[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, 20(9): 1-24.

[35] Wu X, Hou X X, Lai Z H, et al. A codebook-driven approach for low-light image enhancement[J]. Engineering Applications of Artificial Intelligence, 2025, 156: 111115.

[36] 秦嘉奇, 江泽涛, 雷晓春. 基于ICFIE-YOLO的低照度图像目标检测方法[J]. 电子学报, 2025, 53(2): 514-526.
Qin Jiaqi, Jiang Zetao, Lei Xiaochun. Low illumination image object detection method based on ICFIE-YOLO[J]. Acta Electronica Sinica, 2025, 53(2): 514-526. (in Chinese)

[37] Wei Z X, Chen L, Tu T, et al. Disentangle then parse: Night-time semantic segmentation with illumination disentanglement[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 21536-21546.

[38] Tan X, Xu K, Cao Y, et al. Night-time scene parsing with a large real dataset[J]. IEEE Transactions on Image Processing, 2021, 30: 9085-9098.

[39] Deng X Q, Wang P, Lian X C, et al. NightLab: A dual-level architecture with hardness detection for segmentation at night[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 16917-16927.

[40] Xie Z F, Wang S, Xu K, et al. Boosting night-time scene parsing with learnable frequency[J]. IEEE Transactions on Image Processing, 2023, 32: 2386-2398.

[41] Nag S, Adak S, Das S. What's there in the dark[C]//2019 IEEE International Conference on Image Processing. Piscataway: IEEE, 2019: 2996-3000.

[42] Luo R D, Wang W J, Yang W H, et al. Similarity Min-max: Zero-shot day-night domain adaptation[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 8070-8080.

[43] Liang D, Li L, Wei M Q, et al. Semantically contrastive learning for low-light image enhancement[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 1555-1563.

[44] Yun S, Han D, Chun S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 6022-6031.

[45] Olsson V, Tranheden W, Pinto J, et al. ClassMix: Segmentation-based data augmentation for semi-supervised learning[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2021: 1368-1377.

[46] Walawalkar D, Shen Z Q, Liu Z C, et al. Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2020: 3642-3646.

[47] Tranheden W, Olsson V, Pinto J, et al. DACS: Domain adaptation via cross-domain mixed sampling[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2021: 1378-1388.

[48] Zhou Q Y, Feng Z Y, Gu Q Q, et al. Context-aware mixup for domain adaptive semantic segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(2): 804-817.

[49] Chen S J, Jia X, He J Z, et al. Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 11013-11022.

[50] Jiao J B, Liu Y, Liu Y F, et al. VMamba: Visual state space model[C]//Advances in Neural Information Processing Systems 37. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 103031-103063.

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

[52] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C/OL]//The 34th International Conference on Neural Information Processing Systems. 2020: 6840-6851. https://proceedings. neurips. cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[53] Shen Z Y, Mao M Y, Fan P F. A primary comparison of diffusion models and generative adversarial networks for image synthesis[C]//Proceedings of the 2024 7th International Conference on Machine Learning and Machine Intelligence (MLMI). New York: ACM, 2024: 225-234.

[54] Sakaridis C, Dai D X, Van Gool L. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 10745-10755.

[55] Lyu S W, Chang M C, Du D W, et al. UA-DETRAC 2017: Report of AVSS2017 & IWT4S challenge on advanced traffic monitoring[C]//2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE, 2017: 8078560.

[56] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 3213-3223.

[57] MMSEGMENTATION C. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark[EB/OL]. (2020-07-09)[2026-01-25]. https://github.com/open-mmlab/mmsegmentation.

[58] Loshchilov I, Hutter F. Decoupled weight decay regularization[PP/OL]. V3. arXiv (2019-01-04) [2026-01-25]. https://doi.org/10.48550/arXiv.1711.05101.

[59] Hoyer L, Dai D X, Van Gool L. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation[M]//Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 372-391.

[60] Xie Z F, Qiu R, Wang S, et al. PIG: Prompt images guidance for night-time scene parsing[J]. IEEE Transactions on Image Processing, 2024, 33: 3921-3934.

[61] Xie B H, Li S, Li M J, et al. SePiCo: Semantic-guided pixel contrast for domain adaptive semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7): 9004-9021.

[62] Lu Y W, Lang J C, Ding M R. Dual-path consistency unsupervised domain adaptation for nighttime semantic segmentation[C]//ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2025: 10887918.

[63] Sakaridis C, Bruggemann D, Yu F, et al. Condition-invariant semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 3111-3125.

作者简介

雷晓春 女，1981 年 2 月出生于广西壮族自治区南宁市。现为桂林电子科技大学计算机与信息安全学院博士、高级实验师。主要研究方向为图像处理、计算机视觉、人工智能。

E-mail: glleixiaochun@qq.com

吴炜林 男，2000 年 9 月出生于广西壮族自治区北海市。现为桂林电子科技大学计算机与信息安全学院硕士研究生。主要研究方向为图像处理、计算机视觉。

E-mail: 23032303013@mails.guet.edu.cn

江泽涛 男，1961 年 3 月出生于江西省九江市。现为桂林电子科技大学计算机与信息安全学院博士、教授。主要研究方向为图像处理、计算机视觉、人工智能。

E-mail: zetaojiang@126.com

朱文才 男，1999 年 11 月出生于河北省沧州市。现为西北工业大学计算机学院博士研究生。主要研究方向为图像处理、计算机视觉。

E-mail: zwc33@mail.nwpu.edu.cn

刘颖健 男，2000 年 3 月出生于江西省宜春市。现为桂林电子科技大学计算机与信息安全学院硕士研究生。主要研究方向为图像处理、计算机视觉。

E-mail: 2547051960@qq.com

陈冬梅 女，2000 年 5 月出生于广西壮族自治区钦州市。现为桂林电子科技大学计算机与信息安全学院硕士研究生。主要研究方向为图像处理、计算机视觉。

E-mail: 23032303002@mails.guet.edu.cn

吴思琦 男，2004 年 3 月出生于湖北省孝感市。现为桂林电子科技大学计算机与信息安全学院本科生。主要研究方向为计算机视觉与计算生物学。

E-mail: 2201610118@mails.guet.edu.cn