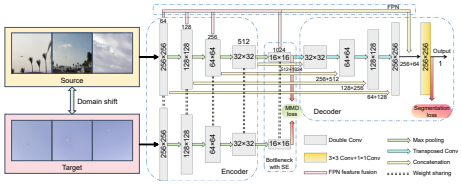
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 1 月, 第 53 卷, 第 1 期

基于领域自适应的轻量级改进 UNet 小目标飞机跨域分割模型



赵文峰[1,2,3,4], 夏运霞[1,2,3,4*], 贾昊龙[1,2,3,4], 包启亮[1,2,3,4]

[1] 中国科学院光场调控科学技术全国重点实验室，四川 成都 610209；
[2] 中国科学院光束控制重点实验室，四川 成都 610209；
[3] 中国科学院光电技术研究所，四川 成都 610209；
[4] 中国科学院大学，北京 100049

摘要：针对现有语义分割模型对小目标飞机分割效果不佳和跨域泛化能力弱的问题，本研究提出了一种基于领域自适应的语义分割模型 FPN-SE-DA-UNet。该模型引入 FPN (feature pyramid network) 模块，通过搭建特征金字塔，充分利用多尺度信息，增强模型对小目标飞机的分割能力；同时，引入 SE (squeeze-and-excitation) 注意力机制模块，增强模型对重要特征的捕捉能力；此外，引入领域自适应方法，将源域与目标域的特征分布进行对齐，显著增强了模型的跨域泛化能力；最后，以 FPN-SE-DA-UNet 作为教师模型，结合知识蒸馏策略，训练得到了轻量级的学生模型。实验结果表明，在跨域小目标飞机分割任务中，FPN-SE-DA-UNet 的 IoU (intersection over union) 与 DICE 指标分别达到了 0.5229 和 0.6866，显著优于现有主流语义分割模型。此外，通过知识蒸馏得到的学生模型，IoU、DICE 指标分别为 0.4557 和 0.6260，且在 NVIDIA GeForce RTX 3090 上平均运行速度达到 59.5 f/s，较 FPN-SE-DA-UNet 提升 49.1%，能够在实际应用中兼顾准确率与实时性。

关键词：语义分割；注意力机制；领域自适应；知识蒸馏

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250145 | CSTR: 32245.14.oee.2026.250145 | 中图分类号：TP391.4 | 文献标识码：A

赵文峰，夏运霞，贾昊龙，等. 基于领域自适应的轻量级改进 UNet 小目标飞机跨域分割模型 [J]. 光电工程，2026，**53**(1): 250145

Zhao W F, Xia Y X, Jia H L et al. A lightweight improved UNet model with domain adaptation for cross-domain segmentation of small aircraft targets[J]. *Opto-Electron Eng*, 2026, **53**(1): 250145

1 引　言

随着全球航空运输业和无人机技术的迅猛发展，航空器 (如直升机、客机、无人机等) 在民用安全领域的应用逐渐深化。空域监测与飞机异常行为识别在公共安全及民用航空管理等方面显得尤为重要，这使得如何对飞机进行高精度、高实时性检测成为了一个重要的课题[1]。

近年来，随着深度学习与计算机视觉技术兴起，基于卷积神经网络 (convolutional neural networks, CNN) 的预测模型在飞机检测任务中得到了广泛应用。在计算机视觉领域中，目标检测 (object detection) 与语义分割 (semantic segmentation) 是两个典型且重要的研究方向。目标检测指对图像中目标的类别和位置进行预测，通过边界框 (bounding box) 来表示目标的位置。然而，目标检测只能实现对目标所在区域的预测，以边界框的形式作为输出，不能做到像素级的分割，导致无法精确地提供目标的形状和细节信息；语义分割则能实现对目标的

收稿日期：2025-05-14；修回日期：2025-07-13；录用日期：2025-07-24；网络出版日期：2026-01-25
*通信作者：夏运霞，xyxhaha@163.com。
版权所有©2026 中国科学院光电技术研究所

像素级检测[2]，将图像中的每个像素进行分类，从而获得更精细的目标表示，便于后续对目标的轮廓、姿态等进行更深入的分析。为获取更精确的目标信息，一些研究人员采取了语义分割的方式实现飞机检测：Shaar 等[3]为了实现对遥感图像中飞机的分割，提出了一种基于 UNet 的二值图像分割模型，在该任务中取得了较好效果；Wang 等[4]构建了一种具有注意力机制的多尺度残差 UNet (MSRAUNet) 模型，设计并引入了一个多尺度卷积模块、两个修改过的 Respath 和两种注意力模块到 UNet 架构中，以提取多尺度特征，增强模型对不同尺度飞机的分割能力；刘思婷等[5]利用 Mask R-CNN 网络对飞机进行分割，随后引入条件随机场 CRF (conditional random fields) 算法对分割结果进行精化，为后续的任务提供了更丰富、更精确的边界轮廓信息。

然而，上述工作考虑的大部分是针对大或中目标飞机的分割任务，针对小目标飞机的相关研究仍显不足，且面临诸多挑战[6]。小目标飞机在图像中占据的像素区域有限且分辨率较低，传统网络往往难以捕捉到目标的关键细节，无法实现高精度分割。此外，在大多数深度学习模型中，通常默认“训练集与测试集来自相同分布”，但在实际应用中，图像数据往往采集自不同的传感器设备、机场环境和气象条件，因此训练集和测试集会存在特征分布差异，导致模型面临“域偏移 (domain shift)”问题[7]，在新场景中模型的预测性能显著下降。同时，为了追求更高的分割精度，过去的研究大多采用了更为复杂的网络结构，以高计算复杂度实现分割精度的提升，这会导致推理时间延长，难以满足实际应用中对实时性的严格要求。这些挑战的存在，使得小目标飞机的实时精确分割仍然是一个亟待解决的重要研究课题。针对这些问题，本研究提出了一种基于领域自适应与知识蒸馏的高泛用性和实时性语义分割方法，主要工作如下：

1) 引入 FPN (feature pyramid network) 模块与注意力机制 (attention mechanism) 模块：通过搭建特征金字塔，将不同尺度的特征进行有效融合，增强模型对小目标飞机的分割能力；通过引入注意力机制模块，使模型能够聚焦于对任务更关键的信息，降低对其他无关信息的关注度，增强对小目标飞机特征的捕捉能力。

2) 引入领域自适应 (domain adaptation) 方法：通过将源领域与目标领域的特征分布进行对齐，使模型能够在不同域之间学习到共性特征，增强模型在不同场景下对目标飞机的分割能力，提升模型的跨域泛用性。

3) 引入知识蒸馏 (knowledge distillation) 技术：通过构建教师-学生网络框架，利用教师模型 (teacher model) 产生的“软标签 (soft labels)”指导轻量级的学生模型 (student model) 进行训练，从而在显著缩短推理时长的同时，维持其在小目标飞机分割任务上的高精度表现。

2 相关工作

2.1 UNet 网络

UNet 是一种常用的语义分割网络，最初由 Ronneberger 等[8]提出。如图 1 所示，该网络采用 U 形架构，由编码器 (encoder)、解码器 (decoder) 和跳跃连接 (skip connections) 构成。其中，编码器用于提取特征信息，解码器用于恢复空间分辨率，跳跃连接则用于融合高分辨率细节与深层语义信息，从而实现目标的精细分割。

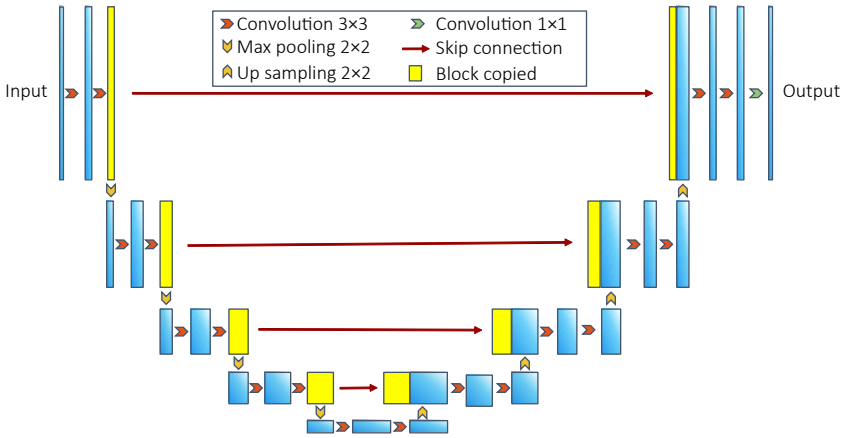


图 1 UNet 网络结构示意图

Fig. 1 Schematic diagram of the UNet architecture

在经典 UNet 的基础上，为适应不同下游任务的需求，提升网络在复杂场景下的分割精度，研究者提出多种改进策略：Zhou 等[9] 在编码器和解码器之间采用嵌套和密集的跳跃连接，并结合深度监督机制，提出了 UNet++，以满足医学图像中更准确分割的需求；Oktay 等[10] 在 UNet 中引入注意力机制，提出了 Attention UNet，该网络能够抑制输入图像中的无关区域并增强局部关键特征，在 CT 胰腺分割任务中取得了较好的效果；此外，更多基于 UNet 的改进方法在医学影像分割[11]、遥感影像分割[12] 以及工业检测[13] 等诸多领域均表现出优异的性能。

2.2 领域自适应

领域自适应是迁移学习中的一种方法，旨在应对“域偏移”问题，以此增强模型的泛化能力。如图 2 所示，域偏移指源领域与目标领域的特征和类别信息相同，但特征分布不同。为应对此问题，通常采取领域对齐的方法，使源领域和目标领域的特征分布尽可能相似。然而在实际应用中，可能存在标注数据昂贵或难以获取的情况，无法获得足够多样化的标注样本，此时目标领域和源领域可能存在域偏移，这种特征分布的差异会显著影响模型在目标领域上的表现。为了有效应对此问题，研究者们提出了多种策略：Tzeng 等[14] 将统计学中的最大均值差异 (maximum mean discrepancy, MMD) 指标[15] 引入深度学习模型中，用于度量源领域和目标领域在再生希尔伯特空间中的距离，并通过最小化该距离，对齐源领域和目标领域的特征分布，以此减少域偏移带来的影响；Sun 等[16] 提出了 CORAL 方法，通过对源领域和目标领域特征的第二阶统计量 (即协方差矩阵) 进行对齐，优化模型在目标领域上的表现。

2.3 知识蒸馏

考虑到当前许多性能优异的深度学习模型存在参数规模庞大、计算复杂度高等问题，不利于在资源受限的实际场景中部署以及实时检测，因此亟须探索更加高效且实用的模型压缩方法。其中，知识蒸馏作为一种有效的模型压缩技术，最早由 Hinton 等[17] 提出，旨在通过将大型的、复杂的教师模型训练得到的知识传递给一个较小的学生模型，提升学生模型的性能。如图 3 所示，其核心思想是在用硬标签 (hard labels) 对学生模型训练的同时，引入教师模型生成的软标签，用于指导学生模型训练。软标签提教师模型对训练数据的“软”概率分布，使学生模型不仅能学习目标标签信息，还能捕捉更多潜在、丰富的知识信息，从而改善其在任务中的表现。

3 模型设计

本研究在经典 U-Net 架构的基础上，从四个方面进行了扩展与改进，以强化对小目标飞机的分割能力、提升跨域泛化性能并兼顾模型的轻量化。如图 4 所示，首先，引入 FPN 模块，构建多尺度特征金字塔，增强模型多尺度特征提取能力；其次，在瓶颈层引入 SE 注意力机制模块，通过对通道进行自适应加权，突出关键通道并抑制冗余特征；再次，引入领域自适应方法，利用最大均值差异度量并对齐源领域与目标领域的特征分布，从而有效提升网络在跨域场景下的鲁棒性；最后，引入知识蒸馏技术，利用教师模型生成的软标签指导学生模型训练，得到轻量级的学生模型，在降低参数量的同时，维持其在小目标飞机分割任务上的高精度表现。

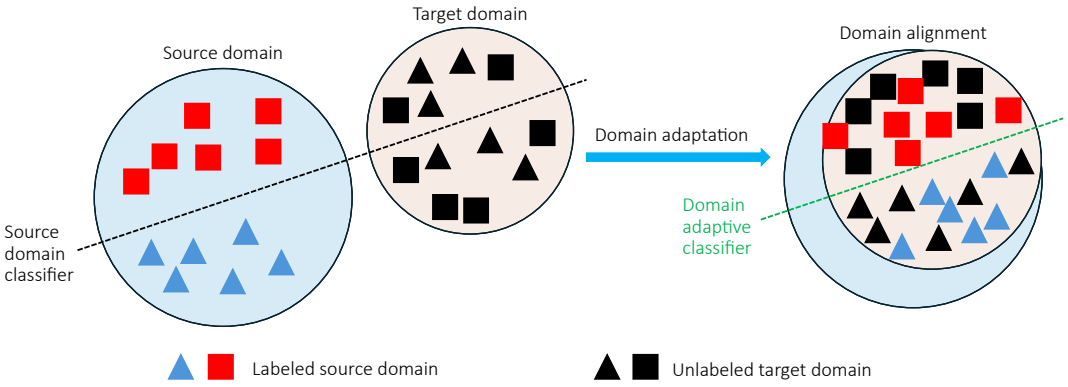


图 2 域偏移示意图

Fig. 2 Illustration of domain shift

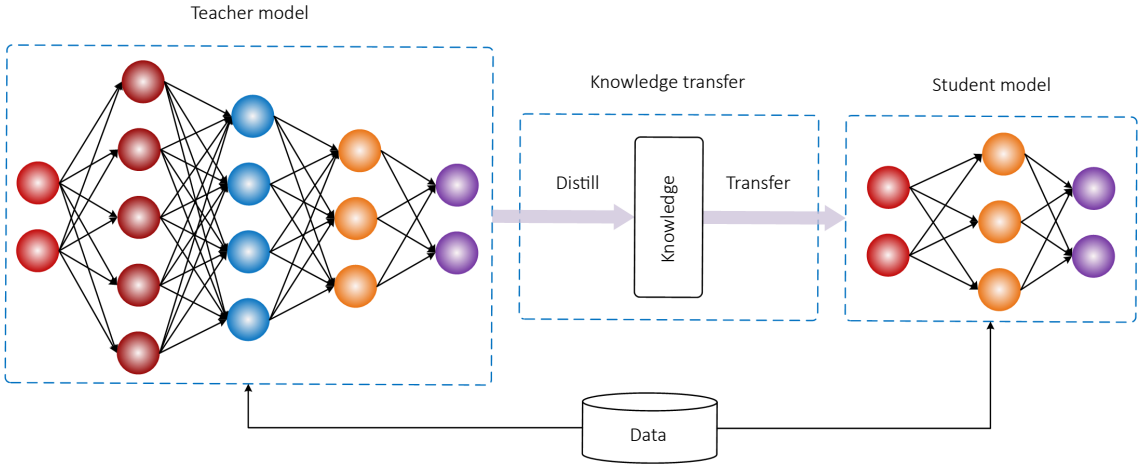


图 3 知识蒸馏架构图

Fig. 3 Architecture of knowledge distillation

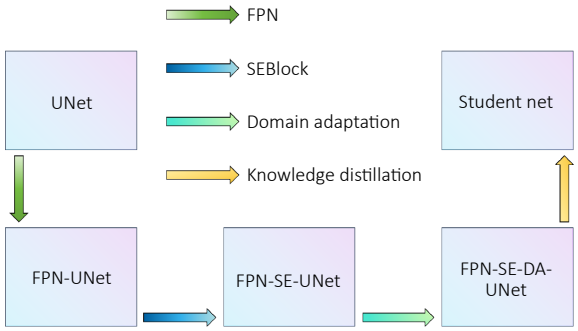


图 4 本研究模型设计流程图

Fig. 4 Design flowchart of the proposed model

3.1 Backbone 架构——UNet

本文选用经典的 UNet 模型[8]作为分割网络的 Backbone，其整体框架以及参数设计如图 5 所示。该架构采用对称式的编码器和解码器结构，其中编码器部分通过连续的卷积与池化操作，实现从浅层到深层的多尺度特征提取，并在瓶颈层充分融合更高层次的全局语义特征。随后，解码器将得到的特征图进行逐级上采样，并通过跳跃连接接收来自同层编码器的细节特征，在恢复空间分辨率的同时保持对目标边缘细节的精准感知。在网络末端，通过 1×1 卷积将多通道特征映射为单通道，再经由 Sigmoid 函数输出目标与背景的二值概率图。

3.2 FPN 模块

为了充分利用多尺度信息，提升模型对小目标的分割能力，本文引入 FPN[18]模块，构建了 FPN-UNet。该模块结构主要包括自底向上的特征提取、自顶向下的特征融合以及横向连接三部分。随着网络深度的加深，特征图的空间分辨率逐渐降低，语义信息不断增强。本研究提取了编码器以及瓶颈层输出的多分辨率特征图，即图 5 与图 6 中的 E1~E5。由于不同尺度的特征图重要程度不同且过多特征图的融合会显著增加模型的参数量，经验证，舍弃 E4 部分特征图不会影响模型的预测性能，故只取 E1、E2、E3、E5。通过 1×1 卷积将每层特征图的通道数统一为 256，得到 C1、C2、C3、C5，实现 FPN 结构中自底向上的特征提取。后续网络对高层特征图进行上采样，以提高其分辨率，并通过横向连接与对应的低层特征图进行逐元素相加，实现自顶向下的特征融合，在每次融合后进行一次 3×3 卷积，用于进一步平滑融合后的特征，最终得到包含多尺度语义信息的特征金字塔 P5、P3、P2、P1。在该特征金字塔的每一层上，都可以独立进行预测任务。考虑到 FPN 最高分辨率的特征图 (P1) 包含了最丰富的语义信息，本研究特别将这层与 UNet 最后一层 (D1) 的解码信息进行拼接，并通过一次 3×3 卷积将二者有效融合后，再接 1×1 卷积输出最终的分割结果。通过这种多尺度特征的补充与融合，模型能够在保持高分辨率细节的同时，充分利用深层语义信息，从而增强对不同尺度目标的分割能力。

3.3 SE 注意力机制模块

SE[19]模块是通道维度上的注意力机制，能够对特征图的通道维度进行自适应加权，从而突出网络中更加重要的通道特征，抑制冗余信息。SEBlock 的结构如图 7 所示，包含四个操作步骤。

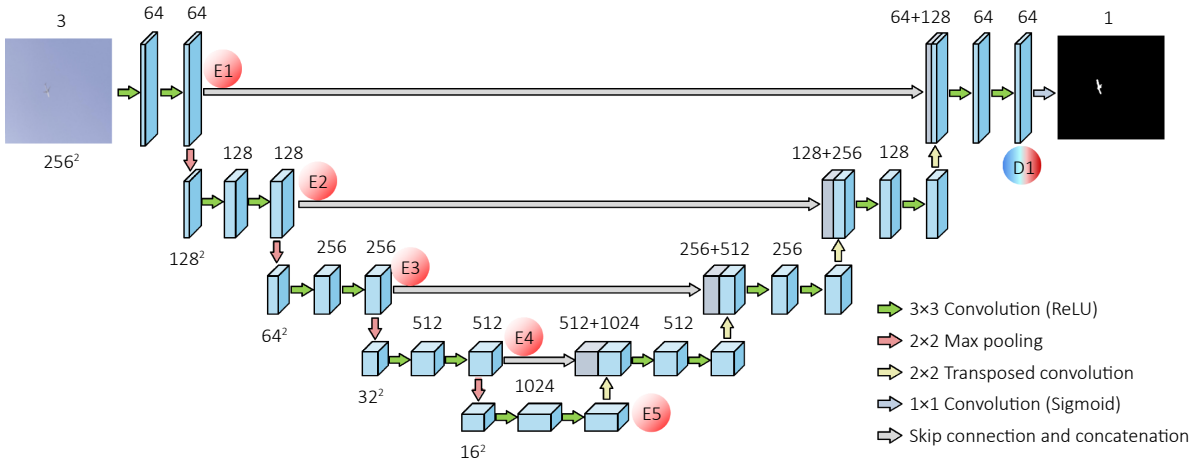


图 5 本研究采用的 UNet 架构图

Fig. 5 UNet architecture used in this research

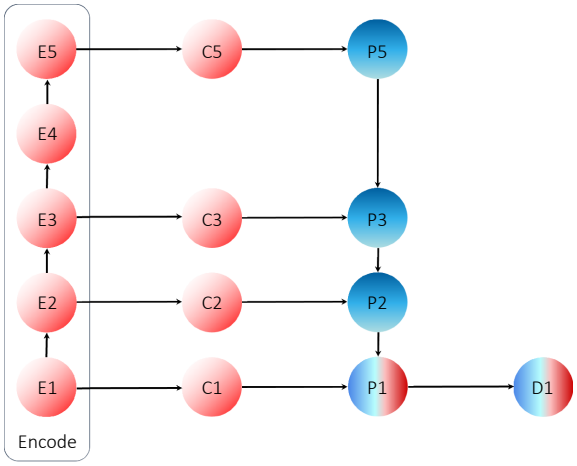


图 6 FPN 模块架构

Fig. 6 FPN module architecture

1) 初始阶段。输入特征图 X 需要进行一个标准的卷积操作作为转换操作，即 F_1，得到维度为 (W, H, C) 的特征图 U。其数学表达式为

$$F_1: X \to U,\ X \in \mathbf{R}^{W' \times H' \times C'},\ U \in \mathbf{R}^{W \times H \times C}, \tag{1}$$

式中：(W', H', C') 表示 X 的维度；(W, H, C) 表示 U 的维度。

2) 对特征图 U 做全局平均池化 (global average pooling)，输出 $1\times1\times C$ 的向量，将每个通道上对应的 $H\times W$ 维度的空间信息压缩到对应通道中，用一个数字 z 表示，计算式为

$$z = F_2(u) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u(i,j), \tag{2}$$

式中：z 为某一通道的平均值；u 为特征图 U 中某一个通道的信息。

3) 之后是通道压缩 (squeeze) 操作与通道激励

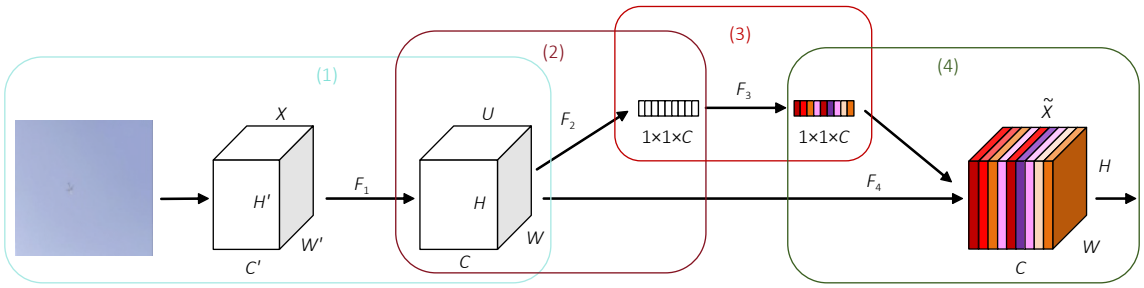


图 7 SEBlock 模块

Fig. 7 Squeeze-and-excitation block

(excitation) 操作，即 F_3。通道向量 z 依次通过降维与升维的全连接操作，以此融合各通道的特征信息。本研究用 s 表示各通道的权重，其数学表达式为

$$s = Sigmoid(W_2 \times ReLU(W_1 z)), \tag{3}$$

式中：W_1为全连接降维层；W_2为全连接增维层。

4) 最后是F_4操作，即对特征图U的每个通道乘以各自的权重，实现对 U 的通道的重要程度重新分配。

由于瓶颈层位于网络的最深处，处理的是最抽象和最具代表性的特征，为了帮助模型更有效地捕捉和利用这些深层特征，本研究在该层引入该通道注意力机制，构建了 FPN-SE-UNet 结构。

3.4 领域自适应方法

由于飞机目标图像在不同环境、摄像条件下的特征分布存在明显差异，仅依赖有标签的源领域进行训练，往往无法在无标签的目标领域中取得理想性能。为此，本文在网络中引入了领域自适应方法，构建了 FPN-SE-DA-UNet，其网络架构如图 8 所示。

由于编码器部分包含更关键的特征信息，本研究在编码器部分采取“双域输入”，即在一次 forward 中，同时输入带标签的源域和无标签的目标域图像，经过瓶颈层后，提取源领域和目标领域的特征信息，采用 MMD 度量二者在再生希尔伯特空间 (reproducing kernel Hilbert space, RKHS) 中的距离，其数学表达式为

$$MMD(X_S, X_T) = \left\| \frac{1}{|X_S|}\sum \phi(x_s) - \frac{1}{|X_T|}\sum \phi(x_t) \right\|, \tag{4}$$

式中：ϕ表示映射，即把原变量映射到再生希尔伯特空间中，作用于源数据点$x_s \in X_S$和目标数据点$x_t \in X_T$；$MMD(X_S, X_T)$表示源数据X_S和目标数据X_T之间的距离。

通过最小化该距离，在特征空间中实现源领域和目标领域特征分布对齐，以此有效提升模型在目标领域的预测性能。同时，为了合理控制源域和目标域特征分布对齐的程度，避免由于特征分布的过度对齐导致模型分割性能下降，本文引入超参数λ，通过在损失函数中引入$\lambda MMD^2(X_S, X_T)$项，实现对特征分布对齐程度的动态调节，引入领域自适应后，模型总损失函数为

$$L = L_s + \lambda MMD^2(X_S, X_T), \tag{5}$$

式中：L 为引入领域自适应方法后的总损失；L_s为分割损失；λ为领域自适应的权重系数。

3.5 知识蒸馏架构

由于引入领域自适应方法使模型需要同时处理源域和目标域的信息，提升了模型的复杂度，为兼顾推理效率与分割精度，本文引入知识蒸馏策略，以训练好的 FPN-SE-DA-UNet 为教师模型，设计了教师-学生 (teacher-student) 网络架构，在不显著降低性能的情况下，训练出了一个轻量级的学生模型，网络架构如图 9 所示。

在本研究中，学生网络的设计基本沿袭教师网络的构架，并完整继承了 FPN、SE 注意力机制以及领域自适应模块，以此确保在多尺度特征融合、通道注意力强化和跨域特征对齐等方面与教师模型保持一致。具体而言，除瓶颈层外，学生网络将初始的 UNet 架构中编码

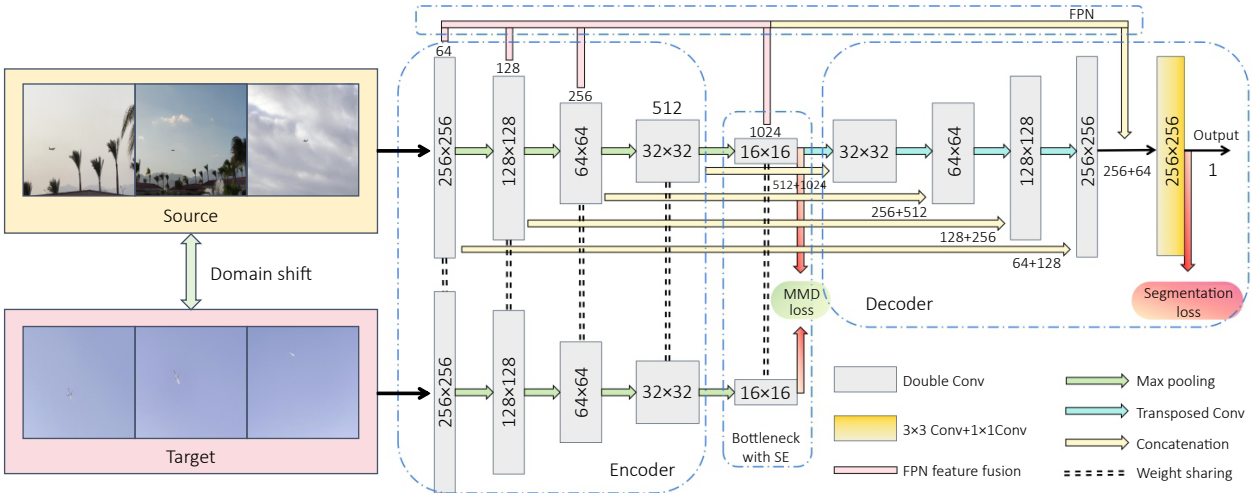


图 8 FPN-SE-DA-UNet 网络架构图

Fig. 8 FPN-SE-DA-UNet architecture

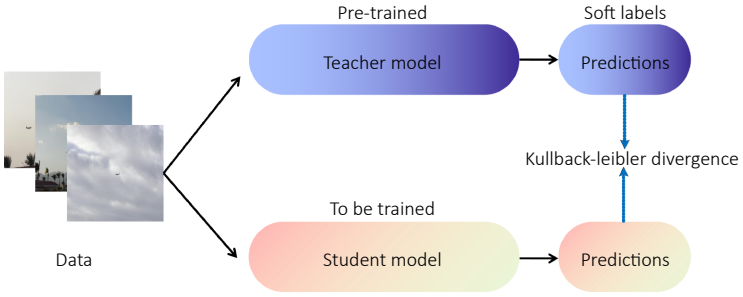


图 9 知识蒸馏架构

Fig. 9 Architecture of knowledge distillation

器各阶段的通道数削减至教师网络的一半，由于编码器的特征需通过跳跃连接与解码器拼接，解码器端的通道宽度也相应缩减，从而在维持网络拓扑结构的同时，显著降低了模型参数量与计算开销。

KL 散度 (Kullback-Leibler divergence) 可用于度量两个概率分布间的相对熵差异，其数值越小意味着分布越趋于一致，其数学表达式为

$$D_{\mathrm{KL}}(P\|Q)=\sum P\log\frac{P}{Q}, \tag{6}$$

式中：P 代表教师模型输出的概率分布；Q 代表学生模型的预测分布。

为了充分利用预先训练好的教师模型输出的软标签来指导学生模型，在训练中，本文将 KL 散度引入蒸馏损失L_{kd}，最小化此损失后，学生模型能够在每个像素点模仿软标签的概率分布，使学生模型学习到教师模型的判别能力与特征表达。学生模型在显著减少计算量的同时，仍能在跨域场景下维持较高的分割性能，为资源受限的实际部署提供了良好的平衡方案。

4 实验结果与分析

4.1 数据集收集和预处理

本研究使用来自 Kaggle 平台提供的数据集 Accurate plane shapes/segmentation[20]，该数据集由 MetaVision 团队制作而成。其包含了 9 个不同场景下的 11761 张 1280 pixel×720 pixel 的飞机 RGB 图片及其对应的像素级高质量标注，其中部分图片及其对应的标注信息如图 10 所示。数据集中飞机目标均为小尺寸，平均像素占比约 0.00082，每个场景图片的数量以及飞机标签区域占背景区域的平均大小比例如表 1 所示。本研究选取场景 9 图片作为目标域，包含 1210 张图片，剩下与之不同场景的 10551 张图片作为源域。源域所有带标注样本与目标域中 30% 无标注图片作为训练集，目标域剩余 70% 图片及对应的标注信息作为测试集。此外，对标注图进行二值化处理，像素值大于 0 的位置设为前景，其余位置为背景，最后将所有图片尺寸均调整为 256×256。

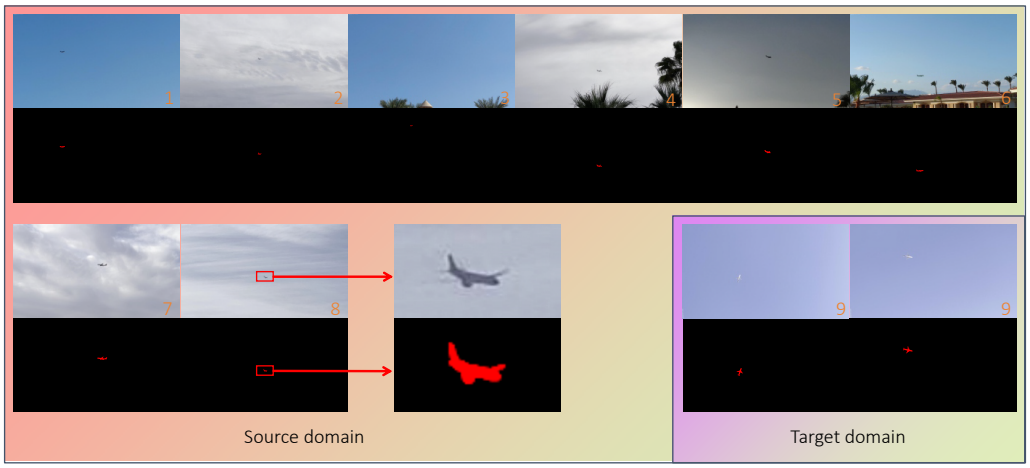


图 10 本研究采用的数据集样例

Fig. 10 Dataset samples used in this research

表 1 每个场景图片的数量及飞机标签区域占背景区域的平均大小比例

Table 1 The number of images per scene and the average ratio of mask size to background area

	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6	Scene7	Scene8	Scene9	All
Number	605	2361	693	1869	761	1441	1724	1097	1210	11761
Ratio	0.00074	0.00064	0.00058	0.00090	0.00082	0.00098	0.00088	0.00074	0.00091	0.00082

4.2 参数配置

由于数据集中均为小目标，为缓解前景背景不平衡问题，本研究在损失函数的设计中引入了 Dice Loss 和 Focal Loss。为了适配领域自适应任务并增强模型的泛化能力，在损失函数中添加了相应的 MMD 损失。最终损失函数 L 的数学表达式为

$$L = L_{\rm s} + L_{\rm d}\,, \tag{7}$$

$$L_{\rm s} = L_{\rm dice} + L_{\rm focal}\,, \tag{8}$$

$$L_{\rm d} = \lambda MMD^2\,, \tag{9}$$

式中：$L_{\rm d}$为领域自适应损失；$L_{\rm dice}$为 Dice Loss；$L_{\rm focal}$为 Focal Loss。

以 IoU 和 DICE 作为评价指标，来评价模型的分割性能，相应的计算公式为

$$IoU = \frac{|A \cap B|}{|A \cup B|}\,, \tag{10}$$

$$DICE = \frac{2\times|A \cap B|}{|A| + |B|}\,, \tag{11}$$

式中：A 是真实标注区域；B 是模型预测区域。

本研究在一块显存容量为 24 GB 的 NVIDIA GeForce RTX 3090 显卡上完成模型的训练和测试，代码实现基于 PyTorch 深度学习框架。此外为避免模型学习到数据集中连续图像的时序特征，训练集在输入前进行了随机打乱。由于源域图片数量远多于目标域，故通过循环复用目标域图片，直至取完所有源域图片。训练采用 Adam 优化器，且由于 FPN-SE-DA-UNet 模型较为复杂，训练过程对显存需求较高，因此本研究还引入了 PyTorch 自带的混合精度训练机制，以有效降低显存的占用。在训练过程中，经验证，设置超参数如表 2 所示，能够最大程度利用 GPU 显存并使分割精度达到最优，即训练共进行 30 个 epoch，batch size 设置为 16，初始学习率 ($L_{\rm r}$) 设为 0.0001，损失项中领域自适应权重系数设置为 0.008。

在后续进行知识蒸馏的过程中，本研究将损失函数 L设置为

$$L = L_{\rm s} + L_{\rm d} + L_{\rm kd}\,, \tag{12}$$

$$L_{\rm kd} = k \times R(Teacher, Student, T)\,, \tag{13}$$

式中：$L_{\rm kd}$为蒸馏损失；k 为蒸馏系数；R 为引入 KL 散度后，学生模型预测结果与教师模型生成的软标签之间的损失；T 为知识蒸馏的温度系数。

在训练过程中，k 设置为 0.1，T 设置为 2，除对模型中通道数进行缩减，其余超参数与 FPN-SE-DA-UNet 保持一致。

4.3 模型预测性能对比

为验证所提 FPN-SE-DA-UNet 模型的性能，本研究选取了 8 种具有代表性的语义分割模型作为对照组进行对比，包含 FCN[21]、SegNet[22]、AttentionUNet[10]、LWUNet、SegNeXt[23]、Swin-UNet[24]、PIDNet[25]、PP-LiteSeg[26]，采用 IoU 和 DICE 两项指标对各模型进行性能评估。为确保公平性，所有对照模型均使用相同的训练配置，包括混合精度训练策略、学习率、训练轮数等参数，均与 FPN-SE-DA-UNet 保持一致。各模型的性能测试结果如表 3 所示，部分实验可视化结果如图 11 所示。

从表 3 可以观察到，相比于 FCN、SegNet、AttentionUNet、LWUNet、SegNeXt、Swin-UNet、PIDNet、

表 2 超参数设置

Table 2 Hyperparameter settings

Epoch	Batch size	λ	$L_{\rm r}$
30	16	0.008	0.0001

表 3 模型间的性能测试结果对比

Table 3 Performance comparison of experimental results among multiple models

Evaluation index	FCN	SegNet	AttentionUNet	LWUNet	SegNeXt	Swin-UNet	PIDNet	PP-LiteSeg	Ours
IoU	0.1251	0.3076	0.4576	0.2254	0.3385	0.4043	0.3911	0.4337	**0.5229**
DICE	0.2196	0.4695	0.6276	0.3674	0.5054	0.5753	0.5622	0.6049	**0.6866**

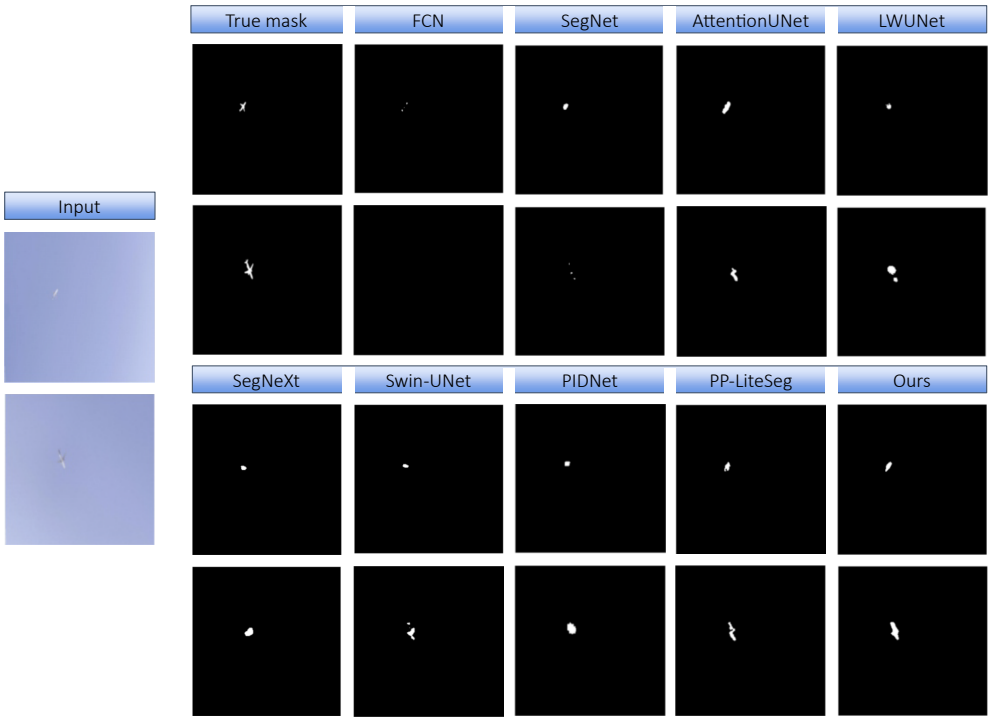


图 11 部分实验结果可视化

Fig. 11 Visualization of selected experimental results

PP-LiteSeg，本研究提出的 FPN-SE-DA-UNet 在 IoU 方面分别提升了 0.3978、0.2153、0.0653、0.2975、0.1844、0.1186、0.1318、0.0892，在 DICE 方面分别提升了 0.4670、0.2171、0.0590、0.3192、0.1812、0.1113、0.1244、0.0817。可见，FPN-SE-DA-UNet 在分割精度和小目标识别能力方面均具有显著优势。其具体原因：首先，研究所使用的数据集中，目标物体 (飞机) 尺寸较小，将图像尺寸调整为 256×256 后，飞机的宽度可能不足 10 个像素点。常规网络直接对输入图像进行多次下采样，造成大部分浅层的语义信息丢失，对于小目标而言，多次下采样后可能只剩下极少的像素。为此，本研究引入 FPN 模块，充分融合了不同层级的特征图信息，避免了下采样后小目标信息丢失的问题，在对小目标的预测上有更好的效果。其次，本研究引入注意力机制 SEBlock，使得模型关注更加重要的通道，增强了模型的特征表示能力。类似地，AttentionUNet 也通过添加了注意力门控单元，让模型能聚焦到关键区域，从而在测试中有较好的表现。最后，考虑到测试集中背景与训练集存在域间差异，本研究引入了领域特征对齐模块，使模型能够学习源域与目标域之间的共享特征，增强了模型的跨域泛化能力，在不同背景下依然保持良好的分割性能。

此外，如图 11 所示，FCN 深层特征图尺寸过小且上采样过程跨度太大，容易造成高分辨率信息丢失，在小目标分割能力上表现最差，几乎无法看出飞机所在位置。Swin-UNet 通过采用带有移动窗口的 Swin Transformer 搭建了 U 形编码器-解码器网络，该网络能够更好地捕获全局特征，因此在小目标分割任务中有较好的效果，能够在一定程度上看出飞机的大致轮廓。PP-LiteSeg 为轻量级分割模型，保留一个高分辨率分支专管边缘细节，避免主干下采样带来的边缘信息丢失，但轻量化设计导致的模型表示能力有限，在小目标场景中特征表达冗余度不足，而提出的 FPN-SE-DA-UNet 则能充分利用高分辨率特征信息，加强了对特征图中关键通道信息的学习，并且将训练集和测试集的特征分布对齐，该网络能够较为精确地锁定飞机所在位置，并对飞机机翼，头尾部的轮廓进行较好的分割。

4.4 消融实验

为了验证提出的网络中各个模块对模型性能的影响，在此数据集上进行消融实验。从初始的 UNet 架构开始，依次添加 FPN 模块，得到 FPN-UNet；添加注意力机制模块，得到 FPN-SE-UNet；添加领域自适应模块，得到最终的 FPN-SE-DA-UNet。分别对上述各模型进行 IoU 和 DICE 指标的测试，结果如表 4 所示，部分实验结果

展示于图 12。

从表 4 和图 12 中可以看出，UNet 由于结构过于简单，对于小目标的分割能力相当有限，仅能在飞机所在位置输出少量像素组成的“点”。在引入 FPN 模块后，增强了模型对小目标的分割能力，IoU 与 DICE 均有一定程度提升，输出的 mask 能够覆盖飞机更多区域。引入注意力机制后，FPN-SE-UNet 在 FPN-UNet 的基础上更加突出了对图片中关键特征的学习，IoU 与 DICE 进一步提升，输出 mask 也能包含较为模糊的轮廓信息。在引入领域自适应模块后，模型的跨域泛化能力得到增强，IoU 与 DICE 也有了明显提升，从 FPN-SE-DA-UNet 的输出 mask 来看，飞机能被较完整地分割，轮廓信息也能在一定程度上被预测，验证了所提各模块对模型性能提升的有效性。

4.5 注意力机制模块对比

为验证本研究引入的 SE 注意力机制在小目标飞机分割任务中的有效性，本研究选取了另外三种常见的注意力机制作为对照组进行对比，包含 CBAM[27]、ECA[28]、DAN[29]，采用 IoU 和 DICE 两项指标进行性能评估，结果如表 5 所示。其中 CBAM、ECA 是参数量较大的空间注意力机制，导致模型易发生过拟合，且这两种注意力机制对局部细节很敏感，容易受噪声影响，针对小目标分割任务效果不佳。DAN 包含两个注意力模块，分别处理全空间和全通道，当有一个模块没有关注到有效特征，就容易削弱特征表达或引起梯度不稳定。SE 注意力机制仅针对通道进行加权，额外参数极少，训练过程易收敛，且避免了对空间维度的过度干预，以极小的参数代价明显提升了模型对小目标飞机的分割性能。

4.6 超参数对比实验

在训练过程中，权重系数λ与学习率L_r是影响领域自适应任务的两个重要超参数，因此本研究探究了不同大小的λ、L_r对模型分割性能的影响，实验结果如表 6、表 7 所示。权重系数λ代表了训练过程中对分割任务与特征分布对齐任务的权衡，λ过大时模型更加专注于源域与目标域特征分布的对齐，λ过小时模型无法充分学习到的源域与目标域之间的共性特征，均会导致模型无

表 4 消融实验结果

Table 4 Ablation study results

Evaluation index	UNet	FPN-UNet	FPN-SE-UNet	FPN-SE-DA-UNet
IoU	0.2674	0.3366	0.3791	**0.5229**
DICE	0.4194	0.5014	0.5491	**0.6866**

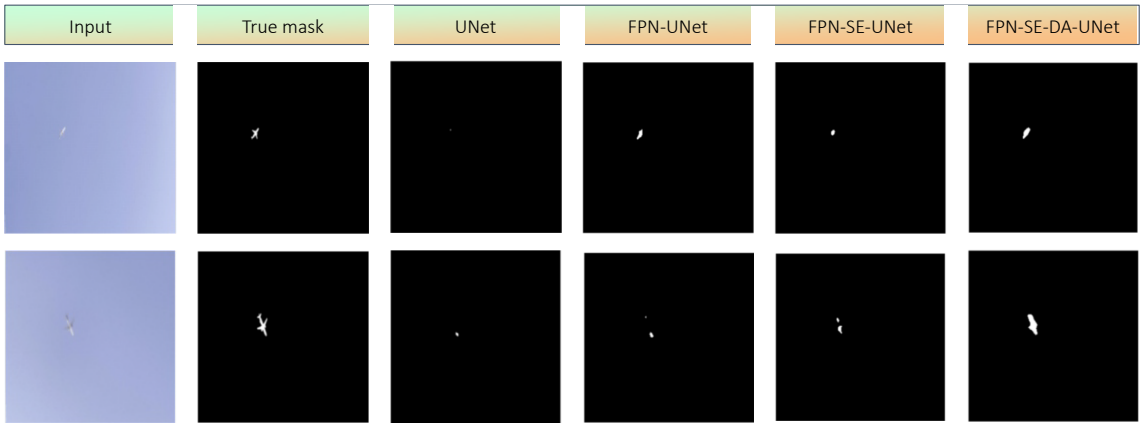


图 12 消融实验部分结果可视化

Fig. 12 Visualization of selected ablation study results

表 5 不同注意力机制测试结果对比

Table 5 Comparison of test results for different attention modules

Evaluation index	SE	CBAM	ECA	DAN
IoU	**0.5229**	0.3562	0.2799	0.1144
DICE	**0.6867**	0.5244	0.4359	0.2042

表 6 L_r=0.0001 时不同λ的实验结果对比

Table 6 Comparison of experimental results for different λ values at L_r=0.0001

Evaluation index	0.0001	0.008	0.001	0.01	0.1	1
IoU	0.4591	**0.5229**	0.5221	0.4893	0.2731	0.2258
DICE	0.6228	**0.6867**	0.6808	0.6568	0.4257	0.3681

表 7 λ=0.008 时不同 L_r 的实验结果对比

Table 7 Comparison of experimental results for different L_r at λ=0.008

Evaluation index	0.00001	0.0001	0.001	0.01	0.1
IoU	0.4798	**0.5229**	0.00338	0	0
DICE	0.6483	**0.6867**	0.00651	0	0

法达到最佳的分割性能。学习率控制每次梯度更新的步幅，直接决定模型训练的收敛速度、稳定性与最终精度。L_r过高会使梯度更新步幅过大，导致损失振荡甚至发散，模型难以收敛而出现欠拟合。L_r过低会导致每次更新步幅过小，模型在有限轮内难以充分逼近最优解，最终精度受限。通过实验证明，当L_r=0.0001 且λ=0.008 时，梯度更新步幅合适且在分割任务与特征分布对齐任务之间实现了最佳权衡，从而使模型分割性能达到最优。

4.7 蒸馏模型对比

本研究搭建的 FPN-SE-DA-UNet 有较高的检测精度，但模型参数量较大，推理过程较为复杂。为满足实际应用中实时性检测的需求，以训练好的 FPN-SE-DA-UNet 作为教师模型做知识蒸馏，得到了学生模型。另外，为验证知识蒸馏的有效性，本研究将 FPN-SE-DA-UNet 直接进行压缩，参数与知识蒸馏得到的学生模型保持一致，得到 Light-Teacher，并比较直接训练得到的 Light-teacher 与知识蒸馏得到的 Student 的分割性能，实验结果如表 8、图 13 所示。

经过知识蒸馏，可以观察到学生模型在总参数量、IoU、DICE 方面相较于教师模型分别降低了 9.3%、12.9%、8.8%，但在帧率方面提升了 49.1%，以较低的精度损失换取了较高的实时性提升。由于通道数的削减。从教师模型压缩而来的 Light-teacher 无法捕捉并表征更丰富的语义信息，因此相比于教师模型，在 IoU、DICE 方面均有明显下降。学生模型由于在训练过程中引入了教师模型生成的软标签，学习到了教师模型的判别能力与特征表达，即使通道数被削减，在输出 mask

表 8 教师-学生模型实验结果

Table 8 Teacher-student model experimental results

Model	Total parameter quantity	IoU	DICE	Frame rate
Teacher	40738434	**0.5229**	**0.6866**	39.9
Light-teacher	**36938722**	0.3985	0.5694	**59.5**
Student	**36938722**	0.4557	0.6260	**59.5**

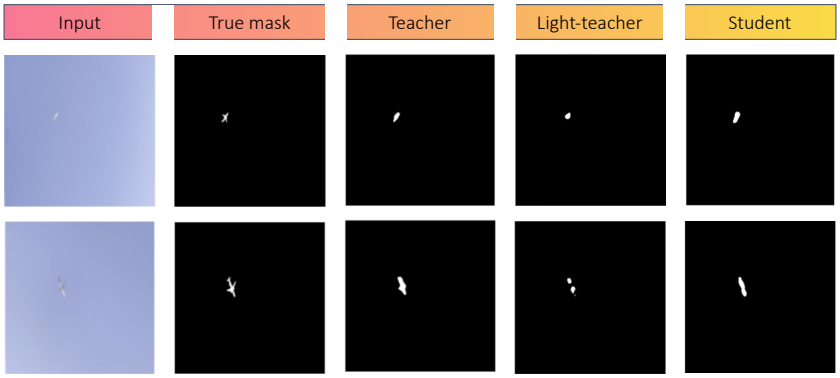


图 13 教师-学生模型部分实验结果可视化

Fig. 13 Visualization of selected experimental results from the teacher-student model

方面仍与教师模型较为接近，亦能够对飞机的整体以及轮廓信息进行较准确的预测。

5 结 论

本研究提出了一种基于领域自适应的语义分割模型FPN-SE-DA-UNet，旨在实现跨域情况下对小目标飞机的分割。通过引入 FPN 模块，充分利用不同尺度的特征信息，增强对小目标的捕捉能力。在此基础上引入 SE 注意力机制模块，突出模型中更加重要的通道特征。此外，引入领域自适应方法，对齐源域和目标域的特征分布，增强了模型的泛化能力。最后，通过知识蒸馏，得到了一个兼顾准确率与实时性的学生模型。经验证，FPN-SE-DA-UNet 相较于其他主流的语义分割模型具有更高的分割精度，通过知识蒸馏得到的学生模型在具有较高实时性的同时亦能保持较强的分割性能。在本研究的分割任务中，FPN-SE-DA-UNet 的性能虽优于其余分割模型，但数据集中飞机目标太小，部分图像甚至肉眼也无法分辨，导致无法分割出清晰的轮廓特征，在未来的工作中，可通过图像预处理、CRF 等算法进一步提升分割精度。

作者贡献

赵文峰：研究资料的收集、分析，论文写作；夏运霞、贾昊龙、包启亮：指导论文选题、论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Zhou H L, Ye Q, Liu W Q. Lightweight remote sensing military aircraft target detection in complex backgrounds[J]. *Opto-Electron Eng*, 2025, **52**(2): 240270.
周韩莲, 叶青, 刘文祺. 复杂背景下的轻量级遥感军用飞机目标检测[J]. 光电工程, 2025, **52**(2): 240270.
2. Wu M J, Zhang Y A, Lin S L, et al. Real-time semantic segmentation algorithm based on BiLevelNet[J]. *Opto-Electron Eng*, 2024, **51**(5): 240030.
吴马靖, 张永爱, 林珊玲, 等. 基于 BiLevelNet 的实时语义分割算法[J]. 光电工程, 2024, **51**(5): 240030.
3. Shaar F, Yılmaz A, Topcu A E, et al. Remote sensing image segmentation for aircraft recognition using U-Net as deep learning architecture[J]. *Appl Sci*, 2024, **14**(6): 2639.
4. Wang X Q, Zhang S W, Huang L. Aircraft segmentation in remote sensing images based on multi-scale residual U-Net with attention[J]. *Multimed Tools Appl*, 2024, **83**(6): 17855−17872.
5. Liu S T, Wang Q D, Zhang L, et al. Aircraft type recognition method by integrating target segmentation and key points detection[J]. *Natl Remote Sens Bull*, 2024, **28**(4): 1010−1024.
刘思婷, 王庆栋, 张力, 等. 联合目标分割和关键点检测的飞机型号识别[J]. 遥感学报, 2024, **28**(4): 1010−1024.
6. Cheng Y B, Lai X F, Xia Y C, et al. Infrared dim small target detection networks: a review[J]. *Sensors*, 2024, **24**(12): 3885.
7. Martinez C, Najera-Flores D A, Brink A R, et al. Confronting domain shift in trained neural networks[C]//*Proceedings of NeurIPS 2020 Workshop on Pre-Registration in Machine Learning*, 2020: 176–192.
8. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
9. Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: a nested u-net architecture for medical image segmentation[C]//*Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop*, Granada, Spain, 2018: 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.
10. Oktay O, Schlemper J, Le Folgoc L, et al. Attention U-Net: learning where to look for the pancreas[Z]. arXiv: 1804.03999, 2018. https://doi.org/10.48550/arXiv.1804.0399.
11. Huang H M, Lin L F, Tong R F, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//*Proceedings of ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, Spain, 2020: 1055–1059. https://doi.org/10.1109/ICASSP40776.2020.9053405.
12. Li R, Zheng S Y, Duan C X, et al. Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images[J]. *IEEE Geosci Remote Sens Lett*, 2022, **19**: 8009205.
13. Lin F Z, Yang J S, Shu J P, et al. Crack semantic segmentation using the U-Net with full attention strategy[Z]. arXiv: 2104.14586, 2021. https://doi.org/10.48550/arXiv.2104.14586.
14. Tzeng E, Hoffman J, Zhang N, et al. Deep domain confusion: maximizing for domain invariance[Z]. arXiv: 1412.3474, 2014. https://doi.org/10.48550/arXiv.1412.3474.
15. Smola A, Gretton A, Song L, et al. A Hilbert space embedding for distributions[C]//*Proceedings of the 18th International Conference on Algorithmic Learning Theory*, Sendai, Japan, 2007: 13–31. https://doi.org/10.1007/978-3-540-75225-7_5.
16. Sun B C, Feng J S, Saenko K. Return of frustratingly easy domain adaptation[C]//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*, Phoenix, USA, 2016: 2058–2065. https://doi.org/10.1609/aaai.v30i1.10306.
17. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[Z]. arXiv: 1503.02531, 2015. https://doi.org/10.48550/arXiv.1503.02531.
18. Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 936–944. https://doi.org/10.1109/CVPR.2017.106.
19. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018: 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
20. Metavision. Accurate plane shapes/segmentation[EB/OL]. [2025-05-14]. https://www.kaggle.com/datasets/metavision/accurate-plane-shapessegmentation/data.
21. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of 2015 IEEE Conference*

on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.

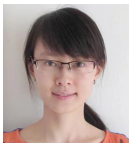
22. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(12): 2481−2495.
23. Guo M H, Lu C Z, Hou Q B, et al. SegNeXt: rethinking convolutional attention design for semantic segmentation[C]//*Proceedings of the 36th Conference on Neural Information Processing Systems*, New Orleans, USA, 2022: 1140–1156.
24. Cao H, Wang Y Y, Chen J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//*Proceedings of European Conference on Computer Vision*, Tel Aviv, Israel, 2022: 205–218. https://doi.org/10.1007/978-3-031-25066-8_9.
25. Xu J C, Xiong Z Z, Bhattacharyya S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023: 19529–19539. https://doi.org/10.1109/CVPR52729.2023.01871.
26. Peng J C, Liu Y, Tang S Y, et al. PP-LiteSeg: a superior real-time semantic segmentation model[Z]. arXiv: 2204.02681, 2022. https://doi.org/10.48550/arXiv.2204.02681.
27. Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, Germany, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
28. Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, 2020: 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.
29. Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, 2019: 3141–3149. https://doi.org/10.1109/CVPR.2019.00326.

作者简介



赵文峰 (1999-)，男，硕士研究生，研究方向为计算机视觉。

E-mail：zhaowenfeng23@mails.ucas.ac.cn



【通信作者】夏运霞 (1986-)，女，博士，研究员，研究方向为高精度视轴稳定控制。

E-mail：xyxhaha@163.com

A lightweight improved UNet model with domain adaptation for cross-domain segmentation of small aircraft targets

Zhao Wenfeng[1,2,3,4], Xia Yunxia[1,2,3,4*], Jia Haolong[1,2,3,4], Bao Qiliang[1,2,3,4]

[1] State Key Laboratory of Optical Field Manipulation Science and Technology, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

[2] Key Laboratory of Optical Engineering, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

[3] Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

[4] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

Objective With the rapid development of global air transportation system and unmanned aerial vehicle technologies, all kinds of aircraft are increasingly being applied for use in the security of civil and airspace monitoring and public safety. So, we need better and more correct aerial target recognition. Thanks to the development of deep learning and computer vision, Convolutional neural network based detection methods are being used for recognition of aircraft. Among them, the semantic segmentation can complete the localization of the pixel level target, therefore, it can provide much more detailed information to extract the aircraft contour and analyze the posture of the aircraft. However, compared to the medium and large aircraft, the small aircraft has fewer pixels in the image and its discriminative features are weaker, which is easily affected by the complex background and is hard to be precisely and steadily segmented by the current model. At the same time, in the practical application scenario, because of the difference in the use of imaging devices, the environment, and the distribution of data, it is easy to produce the domain shift phenomenon, and the model's generalization ability is greatly reduced. And all

*Correspondence: Xia Y X, E-mail: xyxhaha@163.com

these high-accuracy segmentation methods need very complex network structure and occupy lots of computational resources, which is hard to satisfy the real-time requirement of practical systems. So combining domain adaptation and light-weighting modeling approach to get both high accurate real-time small aircraft semantic segmentation of background. This is worth researching and putting into practical engineering.

Methods The paper focuses on the problem of poor segmentation accuracy of small aircraft in cross-domain situation, and carries out some model design and exploration within the scope of a deep learning-based semantic segmentation framework. The classical U-Net plus an additional feature pyramid network (FPN) which is a multi-scale feature fusion structure to improve the model's ability to represent small aircraft targets. In addition to this, a Squeeze-and-Excitation(SE)channel attention is added to the network bottleneck such that it can learn to adjust the weight of each feature channel adaptively, thereby increasing the models ability to react to important discriminative features. To solve the different feature distribution caused by the different scenes and imaging conditions, we proposed a method to align the deep feature distribution between the source and target domains by the MMD, which can also enhance the model's generalization to the target domain. On this basis, we use a teacher-student network and introduce a knowledge distillation approach, where the high-accuracy model with domain adaptation serves as the teacher. Using the teacher's soft label to supervise the training of a light student model, we can greatly reduce the number of parameters and computational cost of the student model, but still have good segmentation effect. Experiment is conducted on the multi-scene small-aircraft dataset with explicit source-target domain split. To validate the effectiveness and real-time performance of the above proposed method through comparison experiment and ablation experiment respectively.

Results and Discussions Small aircraft cross domain semantic segmentation performance, experimental results show that the proposed FPN-SE-DA-UNet is good at both accuracy and generalization. Feature pyramid network comes in to play in order to integrate information from different scale, and solve the loss of fine-grained features caused by the repeated down-sampling process. SE channel attention makes model better at noticing important discriminative features and leads to better target completeness and stability in segmentation results. Cross domain experiment shows that using the maximum mean discrepancy based domain adaptation to align the source and target domain feature can get much higher IOU and DICE score on the target domain test set, which is 0.5229 and 0.6866 respectively. Both is much better than all multiple semantic segmentation model. Also with the same model, it is possible to make a small student model by adding knowledge distillation. The student model is scored at 0.4557 IoU and 0.6260 Dice. The student model gets an average inference speed of 59.5 f/s on an NVIDIA GeForce RTX 3090, there is a slight drop in segmentation accuracy, this is about a 49.1% increase in real-time performance over the teacher model, it is a reasonable trade-off for accuracy vs. real-time performance. From the above we can see that the proposed method is able to obtain a stable small aircraft segmentation for cross-domain scenarios without sacrificing real-time application requirements.

Conclusions To solve the problems of low segmentation accuracy, poor generalization ability, and weak real-time performance in cross-domain semantic segmentation of small aircraft, this study proposes a domain adaptation based semantic segmentation model, FPN-SE-DA-UNet, it introduces the FPN module to obtain richer multi-scale feature information, which can enhance the ability to capture small targets. On the other hand, the SE attention module is used to highlight more informative channel information in the network, and it introduces a domain adaptation strategy to align the features of the source domain and target domain, which can improve the model's ability to generalize. On this basis, a lightweight student model is constructed with the help of a knowledge distillation strategy. It has good segmentation results and good inference efficiency, and the computational complexity has been greatly reduced. Experiments verify the effectiveness and practicability of the proposed method in cross domain semantic segmentation of small aircraft.

Keywords: semantic segmentation; attention mechanism; domain adaptation; knowledge distillation

Zhao W F, Xia Y X, Jia H L, et al. A lightweight improved UNet model with domain adaptation for cross-domain segmentation of small aircraft targets[J]. *Opto-Electron Eng*, 2026, **53**(1): 250145; DOI: 10.12086/oee.2026.250145

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文