

基于注意力机制的红外微小目标分割

王硕霆

智能医学工程 2201 班

U202215210

2963224321@qq.com

摘要—弱小目标检测在计算机视觉领域具有重要的研究价值，尤其是在遥感图像、医学成像和目标追踪等场景中。本文基于视觉认知机理，采用了一种改进的基于注意力机制的目标分割方法，结合 Attention-Unet 模型对弱小目标检测进行了深入研究。我们采用 MDvsFA 数据集进行训练，并在 MDvsFA 和 Sirst 数据集上进行测试评估，以 F1 Score 作为主要性能指标，量化模型的分割效果。实验结果表明，所提出的模型在目标分割的准确性上表现较为优异，验证了注意力机制在弱小目标检测任务中的有效性，展示了注意力机制与深度学习方法结合的潜力。代码在：<https://github.com/wstttttt/attention-U-net> 关键词——图像分割，红外小目标检测，注意力机制

I. 引言

A. 介绍

弱小目标检测是计算机视觉领域的重要研究方向之一，广泛应用于遥感图像目标检测、医学成像中病灶分割、以及目标追踪等场景。其核心在于如何准确检测和分割尺寸较小且信号较弱的目标。由于弱小目标在背景中信号微弱，极易受到噪声干扰，因此传统的目标检测方法在处理此类问题时往往难以获得令人满意的性能。近年来，深度学习技术，尤其是基于卷积神经网络的分割模型，极大地推动了目标检测技术的发展。随着技术的发展，基于深度学习的目标检测与分割方法取得了显著进展，尤其是以 U-Net 为代表的分割模型在医疗影像分割如病灶分割等任务中表现突出。所以在研究初期，我第一个想到的就是使用 UNet 网络来解决红外微小目标分割问题。然而经典的 U-Net 模型在处理弱小目标时仍然存在一些不足，如上下文信息丢失，背景噪声干扰等问题。本文结合视觉认知的注意力机制，采用一种改进的 Attention U-Net 模型，利用注意力机制增强特征提取能力和模型对目标区域的响应。

注意力机制作为模拟人类视觉认知的关键技术之一，能够很有效地分配计算资源至重要区域，提升特征

表达能力。故 AttentionU-net 模型以其较好的特征提取与定位能力，成为解决弱小目标检测问题的潜在方案。基于此，本文以 MDvsFA 数据集为基础，构建并优化了一种基于 Attention U-net 的弱小目标检测算法用于该任务。我们以 F1 Score 为主要性能指标，并在 Sirst 数据集上进一步测试算法的泛化能力，旨在研究注意力机制在弱小目标检测任务中的应用价值与潜力。

B. 相关工作

目标检测的核心任务是准确地定位图像中的目标并判断其类别。传统基于人工设计特征算子的目标检测方法计算复杂度较高，对变化多样的数据适应性较差，难以获取图像数据的高维特征。而深度学习与神经网络的兴起，通过自动从海量数据中学习高层语义特征，不仅提高了检测的准确性，还显著提升了计算效率。

基于深度学习的目标检测方法主要分为两类：单阶段检测网络和两阶段检测网络。单阶段检测网络以端到端的方式直接预测目标的位置和类别，代表性方法包括 YOLO (You Only Look Once) 系列和 SSD (Single Shot Multi-Box Detection)。两阶段检测网络则通过候选区域生成和进一步分类定位的方式进行目标检测，经典方法有 R-CNN、Fast R-CNN 和 Faster R-CNN。这些深度学习方法在自然图像目标检测中取得了显著成功，促使研究者将其引入红外小目标检测领域。

在红外小目标检测任务中，深度学习方法通过卷积神经网络自动提取特征表达，能够有效抑制背景杂波和传感器噪声，增强感兴趣目标的显著特征。例如，文献《Using deep learning to detect small targets in infrared oversampling images》设计了一种七层深度卷积网络用于小目标检测；文献《Interior Attention-Aware Network for Infrared Small Target Detection》提出了一种粗到细的内部注意力感知网络 (IAANet) 用于红外

小目标检测，通过建模像素间的内部关系来提高检测的准确性和鲁棒性；文献《Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images》基于漏检和虚警的平衡策略，提出了一种深度对抗学习网络（MDvsFA），通过对抗训练提升检测性能。

上述研究表明，深度学习方法已成为红外小目标检测领域的重要研究方向，并展现出显著优势。

II. 方法

A. Attention U-net 概述

Attention U-Net 是一种结合了注意力机制与经典 U-Net 架构的深度学习模型，专为图像分割任务设计，尤其适用于目标较小或背景复杂的情况。相比传统 U-Net，Attention U-Net 在跳跃连接中引入了自适应的注意力机制，使模型能够更加聚焦于图像中的重要区域，有效抑制背景噪声，从而显著提升分割精度。

Attention U-Net 的网络结构如下图（图 1）所示。编码器部分与 U-Net 的编码器基本相同，主要通过卷积和下采样操作提取特征。输入图像首先经过两次 $3 \times 3 \times 3$ 的 3D 卷积和 ReLU 激活函数处理，然后进行最大池化下采样操作。编码器中共有 3 组“卷积-池化”块，这些块提取了不同尺度的特征图。解码器部分是 Attention U-Net 的关键改进点。编码器的最后一层特征图被上采样后，与对应的编码器特征图通过注意力机制计算权重（即注意力门控操作）。经过注意力加权的特征图与上采样后的特征图合并，并进行进一步卷积处理。这样的上采样块重复三次后，最终生成分割输出图。

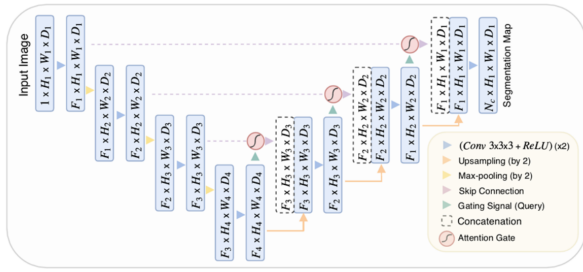


图 1. Attention U-net 网络架构。

B. U-net 网络

U-Net 是一种经典的卷积神经网络（架构，最早用于生物医学图像的分割任务。其特点是具有对称的编码

器（下采样路径）和解码器（上采样路径）结构，其中编码器负责提取图像的特征，而解码器则逐渐恢复图像的空间分辨率。同时，U-Net 采用了跳跃连接，可以将编码器中的低层特征直接传递到解码器中，其目的是保持图片的细节信息。这样的设计使得 U-Net 在图像分割任务中能够更好地恢复边缘和细节。

U-Net 网络架构如下图（图 2）所示。在这副最经典的 U-Net 结构图中，其架构呈现出一个对称的“U”字形，由编码器（Encoder）和解码器（Decoder）两部分组成，并通过跳跃连接（Skip Connections）相互连接。

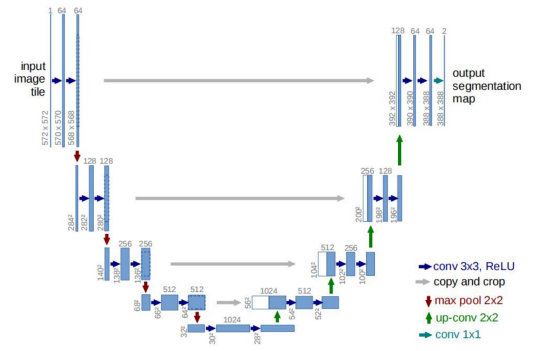


图 2. 经典 U-Net 网络架构。

C. 注意力机制

注意力机制近年来在深度学习领域取得了显著的进展，尤其在计算机视觉和自然语言处理任务中表现出了卓越的效果。注意力机制的核心思想是模仿人类的注意力分配能力，即在处理复杂信息时，根据输入特征的相关性动态调整对不同部分的关注程度。传统神经网络在处理输入数据时，往往对所有特征赋予相同的权重，而注意力机制通过学习生成加权分布，使模型能够集中关注最为重要的特征，同时抑制无关或次要的信息，从而提升模型的表达能力和泛化能力。

具体来说，注意力机制通过计算输入序列中每个特征的注意力权重，生成一个重要性分布。典型的实现方式是通过加性注意力（Additive Attention）或点积注意力（Scaled Dot-Product Attention）对特征间的关系进行建模。例如，在自注意力（Self-Attention）中，输入特征通过查询（Query）、键（Key）、值（Value）的相互作用生成注意力分数，并以此调整特征的权重。此外，多头注意力机制（Multi-Head Attention）进一步增强

了模型的特征捕获能力，通过在多个子空间中并行计算注意力来捕捉多粒度的信息。

D. 注意力门控

在 Attention U-Net 中，注意力门控 (Attention Gate, AG) 是关键模块。它的作用是在跳跃连接中对特征进行加权，使模型能够关注于分割目标的相关区域，抑制无关的背景干扰。通过注意力机制，模型能够自动识别目标区域的重要性，并动态调整特征映射的权重。

如下图 (图 3) 所示，我们可以把 Attention U-Net 网络中的一个上采样块单独拿出来研究，其中 x^l 为来自同层编码器的输出特征图， g 表示由解码器部分的特征图，用于上采样，同时也作为门控信号参数 x^l 参与注意力计算，而 \hat{x}^l 即为经过注意力门控计算后的特征图，此时 \hat{x}^l 是包含了空间位置重要性信息的特征图，然后再将其与下一层上采样后的特征图进行合并得到该上采样块最终的输出。

在编码器中，第 l 层的特征图 x_i^l ，尺寸为 $F_l * H_x * W_x * D_x$ ，经过卷积后得到 $W_x^T x_i^l$ ，其中 W_x^T 为卷积核。上采样中上一层特征图 g_i 尺寸为 $F_g * H_g * W_g * D_g$ ，经过卷积后得到 $W_g^T g_i$ ，将 $W_x^T x_i^l$ 和 $W_g^T g_i$ 相加后进行 ReLU 激活函数处理得到 $\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)$ ，然后在经过卷积计算后得到 q_{att}^l ，最后对 q_{att}^l 进行 sigmoid 激活处理后即可得到最终的注意力得分 α_i^l 。加性注意力具体公式如下：

$$q_{att}^l = \psi^T (\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\psi$$

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l, g_i; \Theta_{att}))$$

E. 损失设计

在本研究中，为了提升模型在红外微小目标分割任务中的性能，我们采用了改进的 Focal Loss 作为损失函数。Focal Loss 是一种针对类别不平衡问题优化的损失函数，其通过对困难样本赋予更高的权重，减弱易分样本对损失的影响，从而引导模型更关注难以正确分类的样本。在红外微小目标分割任务中，由于目标通常占比较小且容易被背景噪声掩盖，类别不平衡问题尤为突

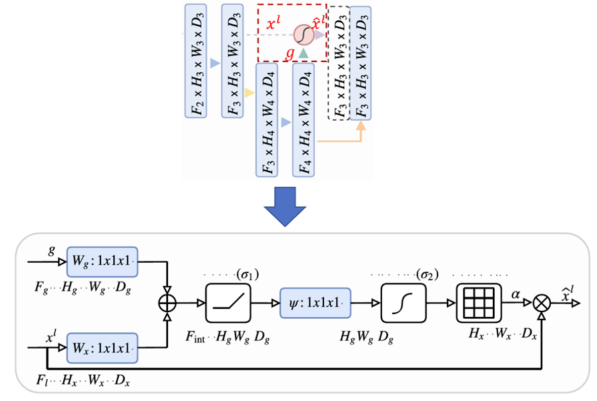


图 3. 注意力门控单元。

出，采用 Focal Loss 能够有效缓解这一问题。具体公式如下：

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t)$$

III. 实验

A. 数据预处理

在本实验中，针对数据预处理流程，我们设计了一系列步骤以提升模型训练的效果与鲁棒性。首先，将图像和掩码调整为固定尺寸以统一输入大小。随后，将图像和掩码转换为 PyTorch 张量格式并归一化至 $[0, 1]$ 范围以标准化数据分布。在数据增强方面，我们采用了随机水平翻转以提高模型的泛化能力，同时引入了 CLAHE，即对比度受限的自适应直方图均衡化算法以增强图像亮度与对比度，配合双边滤波 (Bilateral Filtering) 实现降噪处理。最后，使用 PyTorch 的 DataLoader 将数据加载为批量格式，并通过可视化检查预处理效果。该预处理流程为模型训练提供了高质量的数据支持，有助于提升弱小目标检测任务中的性能和鲁棒性。处理后的图片输出如下：

B. 实验设置

1) 训练数据集：在 Wang 等人提供的 2098 张图像中包含 9978 张大小为 128×128 的图像补丁的训练集上训练 attention U-Net 数据集。其中，我将数据集按 80% 和 20% 分成了训练集和验证集。

2) 测试数据集：在 MDvsFA 和 SRIST 数据集上测试模型的性能。

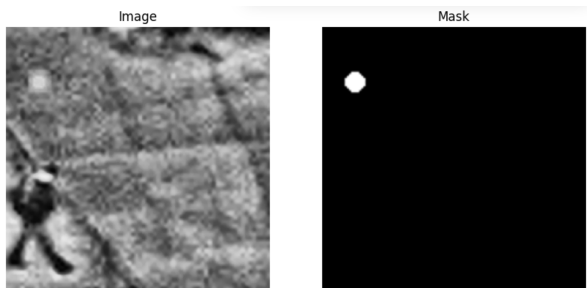


图 4. 预处理后图片。

3) 实验设备: 本研究所有实验是在 Autodl 平台上租的一台拥有 16 vCPU Intel(R) Xeon(R) Platinum 8481CCPU 和 RTX 4090D(24GB)GPU 的服务器上使用 Python 3.12(ubuntu22.04) 运行的。

4) 评价指标: 采取 F1score 作为模型评价指标, 1 Score 是一种结合了精确率 (Precision) 和召回率 (Recall) 的加权调和平均数, 能够在数据分布不均衡的情况下, 综合评估模型的预测能力。精确率衡量了模型预测为正类样本中有多少是正确的, 召回率则衡量了实际为正类的样本中有多少被正确预测。F1 Score 通过以下公式计算:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

红外微小目标分割是一项具有挑战性的任务, 通常面临目标占比极小、背景复杂、目标易被噪声干扰等问题, 导致数据分布存在严重的不均衡现象。因此在这种情况下, 传统的准确率可能无法有效反映模型的实际性能。例如模型即使完全忽略微小目标, 仅对背景区域进行正确分割, 也可能获得较高的准确率, 但无法实现对目标的准确检测和分割。因此, F1 Score 成为更适合该任务的评价指标。

C. 实验结果

实验共设定了 120 个训练周期, 批大小为 4, 学习率设置为 0.001。对于训练阶段, 每轮迭代 1996 个样本, 显示每秒处理的样本数和当前的损失及 F1 分数。测试阶段则每轮迭代 499 个样本, 显示每秒处理的样本数和当前的损失及 F1 分数。采用 F1 分数作为主要评估指标, 在训练和测试阶段分别记录。保存训练中最好的模型权重为结果, 用于测试。

1) 模型在 MDvsFA 上的表现: 输出模型在 MDvsFA 测试集上的平均 F1 score, 并以柱状图形式显示

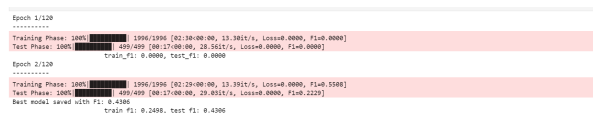


图 5. 训练过程。

每个 batch 上的结果, 如图 6 所示。可以看到模型在这

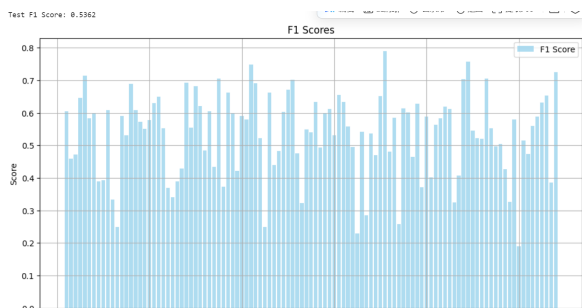


图 6. 模型在 MDvsFA 上的表现

个数据集上的平均 F1score 达到了 **0.5362**, 可以说有着较好的性能。

同时我也将预测结果可视化了, 随机选取了 4 张图片观察其准确性。如图 7-图 10

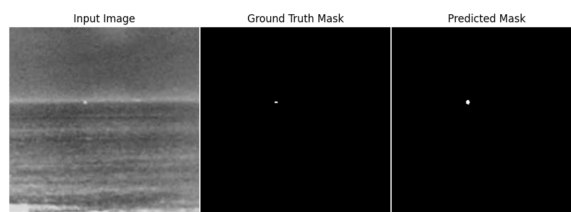


图 7. 1-1

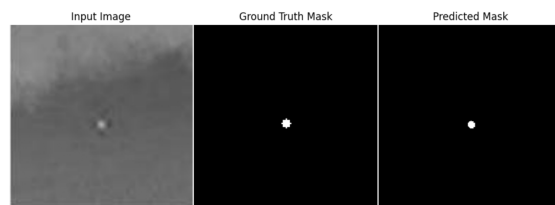


图 8. 1-2

可以看出模型在大部分图片上能取得较好的效果。

2) 模型在 SRIST 上的表现: 如图 11 可以看出模型在 SRIST 数据集上取得的 F1score 为 **0.6698**, 较为不错。

同样的, 我也随机抽取了 4 张图片进行了可视化研究。如图 12-图 15。

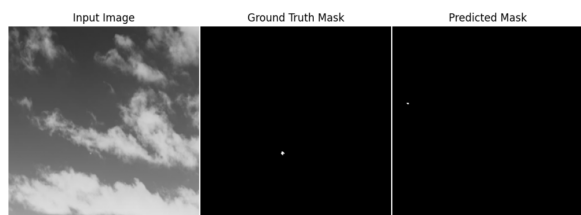


图 9. 1-3

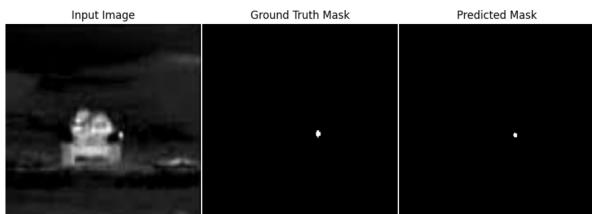


图 10. 1-4

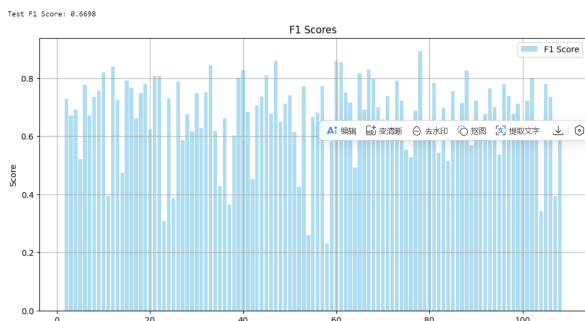


图 11. Enter Caption

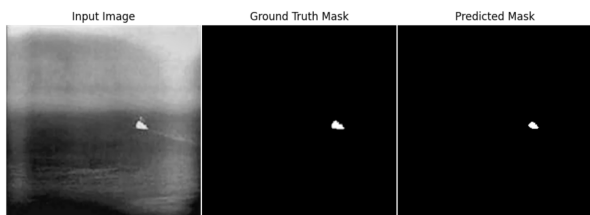


图 12. 2-1

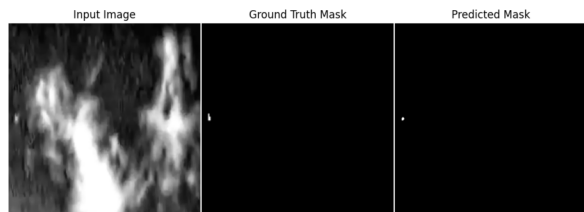


图 13. 2-2

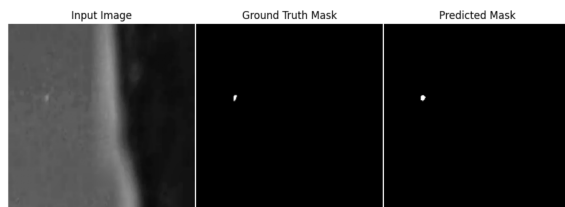


图 14. 2-3

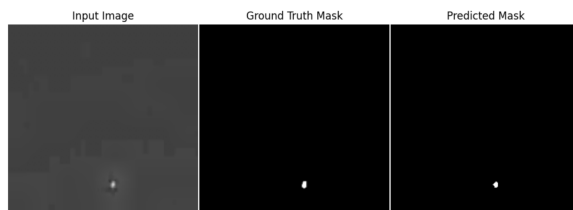


图 15. 2-4

IV. 结论

本文提出了一种基于注意力机制的分割网络 Attention U-Net，用于红外微小目标分割任务。在模型设计中，通过引入注意力门控模块，显著增强了编码器和解码器之间跳跃连接的特征表达能力，使网络能够更加聚焦于红外图像中目标区域的关键特征。具体而言，注意力门控的加入通过动态调整不同特征图的权重，有效抑制背景噪声和干扰信息，从而提高了微小目标分割的精度。

实验结果表明，所提出的模型在 MDvsFA 和 SRIST 两个公开红外微小目标分割数据集上均表现出色。在 MDvsFA 数据集上的 F1 Score 达到 0.5362，在 SRIST 数据集上的 F1 Score 达到 0.6698，相较于现有其它方法展现出一定的优势。

此外，通过可视化分析进一步验证了模型对目标区域的高效捕捉能力。在实际应用中，该方法可为红外微小目标的精准分割提供有效支持，特别是在高噪声场景下依然能够保持良好的性能。

由于 U-Net 模型过大，每次训练都要花费很久的时间，时间原因没有办法使用不同的学习率进行实验得到最佳的学习率。在未来，也会进一步考虑使用其它的 U-Net 网络进行对比实验，更进一步探究不同的事件机制在处理微小红外目标分割任务上的性能。