

1. 摘要

红外图像弱小目标检测 (IRSTD) 是计算机视觉的重要问题, 基于红外小目标检测的技术被广泛应用于安防、军事、航空航天等领域。随着深度学习技术在计算机视觉领域的蓬勃发展, 使用卷积神经网络来进行红外小目标检测成为现阶段的主要趋势。在本题中, 我们选用一种基于 U-Net 网络的红外弱小目标检测方法 MSHNet, 其通过提取多尺度特征, 分别与标签比较得到损失来提高弱小目标检测的精确度。在此基础上, 我们在题给数据集上进行训练, 并分别在 MDvsFA 和 SIRST 数据集上测试, 得到 0.57 和 0.66 的 F1 Score。代码地址:

https://github.com/jerryfeng2003/IRSTD_homework

关键词——红外图片, 弱小目标检测, 卷积神经网络

2. 引言

红外图像即接收红外频段信号的相机或传感器收集得到的图片, 因为红外光对热辐射敏感故而可以从红外图像上观察的场景的热分布, 因此被广泛应用于各个领域, 包括安防、军事和航空航天等。红外图像弱小目标指的是在复杂场景下, 尺寸较小的红外目标物体, 因为图片整体噪音大且目标与背景差异小, 导致难以识别定位。使用目标检测或者图片分割技术对红外图片中的弱小目标进行检测是计算机视觉中的困难问题, 而现阶段深度学习卷积神经网络技术的发展, 大大加强了对这类问题的处理精确度和效率。由于现有红外图片数据集和相机通常采集小尺寸图片, 故而与通用目标检测任务不同, 应用卷积神经网络进行红外弱小目标检测一般直接预测实例分割的 Mask, 相比于预测检测框能更好地标定小目标的位置, 适用于小尺寸图片上的弱小目标。

我们选取了 Multi-Scale Head Network (MSHNet) 作为主要实现方法来进行红外弱小目标检测, 该方法于 CVPR 2024 会议上首次发表。文章中详细描述了其基于 Unet 方法构建了多尺度检测器, 并提出了针对弱小目标检测定位的 Scale and Location Sensitive Loss (SLSL) 来进行监督学习, 并通过实验和可视化对比来证明其方法的优越性。我们在方法框架总体不变的前提下, 对其训练过程进行一系列优化, 包括提前数据预处理和优化 SLSL 实现来加速训练, 并针对原实现代码中训练不稳定的问题, 使用 AdamW 优化器和引入余弦学习率调整策略来进行缓解, 最终得到了较为理想的实验结果。

3. 相关工作

在深度学习兴起之前, 主要采用传统方法实现红外弱小目标检测, 例如 Top-Hat, Max-Median, WSLCM, RIPT, MSLSTIPT 等。这些方法基于手工提取的特征, 往往应用于特定的、有限的场景下, 泛化性被限制。深度学习方法对红外弱小目标检测带来了极大的提升, 例如 MDvsFA 应用了条件对抗生成网络, 弥补了红外弱小目标检测中常发生的遗漏和错检问题。而

通过引入 U-net 架构，ACMNet 和 UIUNet 实现了检测精度的进一步提升。Transformer 架构逐渐被更多的计算机视觉模型采用，RKformer 是第一种使用了随机链接注意力的红外弱小目标检测器。不同于之前工作的创新点，本文选用了较为简单的架构，主要通过改进损失函数来提高性能，以此来维持训练时间、推理速度和检测精度之间的平衡。

4. 方法

MSHNet 论文的主要创新点在于首次在红外弱小目标检测中引入了基于 Unet 的多尺度检测器，并提出了 SLSL 作为损失优化目标。作者指出，现阶段的红外弱小目标检测模型在训练中以 IoU 和 Dice 损失为优化目标，对物体位置和尺寸的敏感性不足，可能是导致最终模型检测精度不佳的原因。

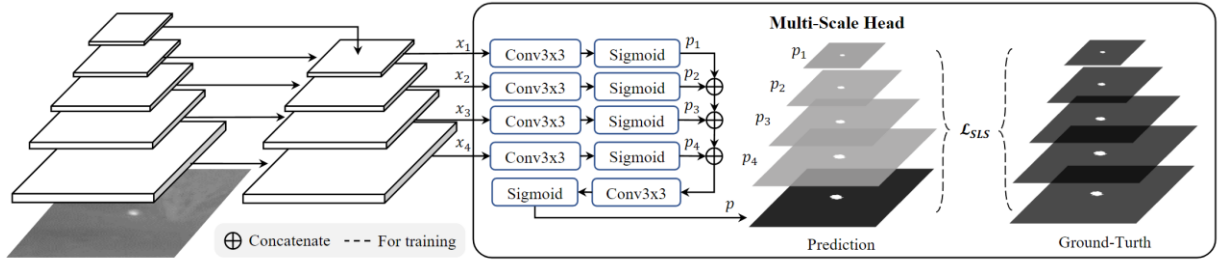


图 1. MSHNet 架构图

4.1. 多尺度检测器（Multi-Scale Head Detector）

U-net 是经典的高精度分割模型，通过对 U-net 模型思想的借鉴，作者提出了如图中的多尺度检测器 Multi-Scale Head Detector，即在反卷积生成预测 Mask 时生成多个尺寸与标签进行对比。通过在多个尺度上对预测 Mask 进行监督训练，能实现由粗到细的定位，从而弥补了敏感性。在本模型中，直接将输入红外图片池化为多个尺度，各自进行卷积后再进行特征融合，得到多尺度的预测输出。

4.2. 尺寸位置敏感损失（Scale and Location Sensitive Loss）

另一方面，文章提出了 Scale and Location Sensitive Loss (SLSL)，即尺寸位置敏感损失。进一步地，可以分为尺寸敏感损失 Scale Sensitive Loss (SSL) L_S 和位置敏感损失 Location Sensitive Loss (LSL) L_L 。而 SLSL 则为两者之和， $L = L_S + L_L$ 。

尺寸敏感损失实际上为对 IoU 损失的改进，通过对 IoU 损失增加和 pixel 位置相关的加权，来增大优化目标对尺寸差异的敏感性。

$$\mathcal{L}_S = 1 - w \frac{|A_p \cap A_{gt}|}{|A_p \cup A_{gt}|},$$

$$\text{s.t. } w = \frac{\min(|A_p|, |A_{gt}|) + \text{Var}(|A_p|, |A_{gt}|)}{\max(|A_p|, |A_{gt}|) + \text{Var}(|A_p|, |A_{gt}|)},$$

位置敏感损失则通过对预测 Mask 和标签 Mask 的 pixel 极坐标的计算，对像素偏移位置进行约束，从而增强了损失函数对位置预测的敏感性。

为了模型能够稳定收敛，作者在训练过程中设置 warm up epoch 时只使用尺寸敏感损失，在热身结束后则加入位置敏感损失，使得模型能在训练前期快速得到较好的分割权重，再进行精细调整。

$$d_p = \sqrt{x_p^2 + y_p^2},$$

$$\theta_p = \arctan(\frac{y_p}{x_p}).$$

$$\mathcal{L}_L = (1 - \frac{\min(d_p, d_{gt})}{\max(d_p, d_{gt})}) + \frac{4}{\pi^2}(\theta_p - \theta_{gt})^2,$$

5. 实验

5.1. 实验设置

原文使用 Adagrad 优化器，设置学习率固定为 0.05，batch size 为 4，训练 400 epoch，并对图片采用在线数据增强。为了加快训练并使模型稳定收敛，我们使用 AdamW 优化器，将 batch size 增加至 64，设置学习率为 0.01，并采用余弦学习率调整策略，设置初始学习率和最终学习率为 1e-3，使用离线数据增强，将预处理的数据储存好直接调用。同时，原论文中的 LSL 通过循环来计算每个 data batch 的损失，我们将过程转为矩阵计算，在维持结果一致的同时加快计算速度。为了展现加速效果，我们使用 python 资源分析工具 pyinstrument 展现了优化前后的 data loader 用时和损失计算用时，总体速度从原来的 5 min 每 epoch 加速至 2 min 每 epoch。

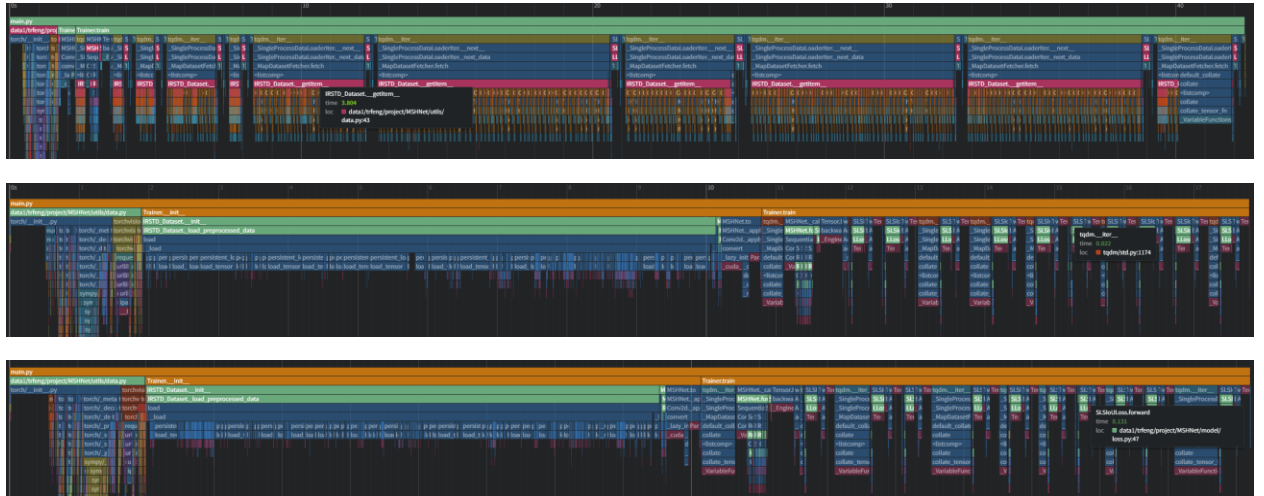


图 2. 不同训练策略下前 10 个 epoch 训练用时可视化

5.2. 红外弱小目标检测量化指标

根据题给要求，我们使用 F1 Score 作为分割结果的量化指标，以衡量模型性能。对模型进行测试分别在 MDvsFA 和 SIRST 得到 0.57 和 0.66 的结果。

5.3. 结果可视化

选取多个测试集样本，将输出图片二值化后 resize 为原图片的尺寸大小，以进行对比。

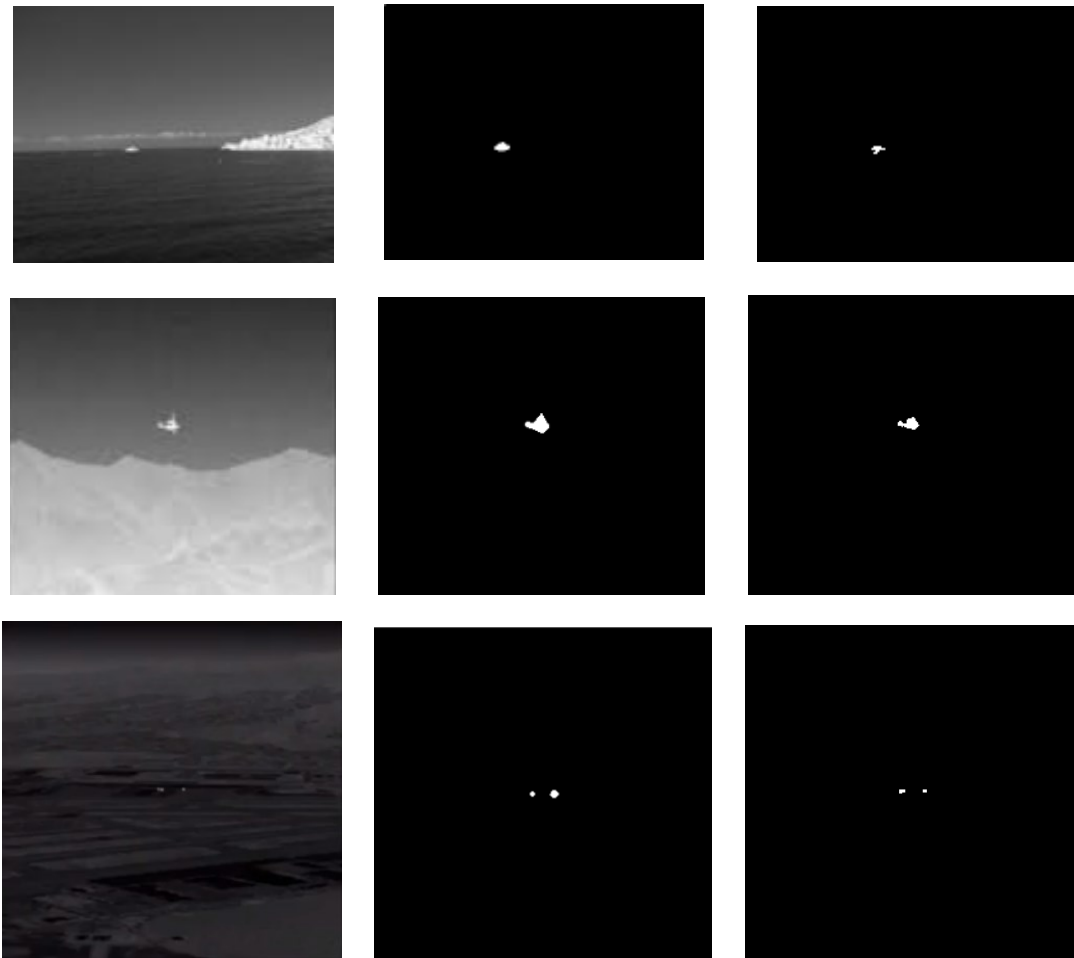


图 3. 检测结果可视化，从左至右依次为原图片，预测 Mask，标签 Mask

附录：Python 代码

本作业全部代码已上传至 github 并开源，代码地址：

https://github.com/jerryfeng2003/IRSTD_homework