中图分类号：TP753　　论文编号：1028704 20-B023
学科分类号：081002

博士学位论文

基于低秩稀疏分解与注意力机制的红外小目标检测

研究生姓名　　戴一冕
学科、专业　　信号与信息处理
研 究 方 向　　图像处理与目标检测
指 导 教 师　　吴一全 教授

南京航空航天大学
研究生院　电子信息工程学院

二〇二〇年十二月

Nanjing University of Aeronautics and Astronautics
The Graduate School
College of Electronic and Information Engineering

Infrared Small Target Detection Based on Low-Rank Plus Sparse Decomposition and Attention Mechanism

A Thesis in

Signal and Information Processing

by

Yimian Dai

Advised by

Prof. Yiquan Wu

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Engineering

December, 2020

承诺书

本人声明所呈交的博士学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南京航空航天大学或其他教育机构的学位或证书而使用过的材料。

本人授权南京航空航天大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本承诺书）

作者签名：_____

日　　期：_____

摘　要

小目标检测是红外探测系统的关键技术，其性能好坏对于能否及早发现和跟踪潜在目标、为预警系统或者精确制导系统赢得充足的反应时间至关重要。受限于红外成像的特点以及远距离成像的需求，感兴趣目标在红外图像中往往呈现出绝对尺寸过小、本征特征稀缺的特点。此外由于背景杂波干扰严重、图像场景变化复杂等因素，尽管经过多年的发展，红外小目标检测仍然在准确率和鲁棒性上面临着巨大的困难和挑战。针对该问题，本文沿着从模型驱动、到数据驱动、再到模型驱动的深度学习的发展脉络展开了相关研究，从建立能够更为准确刻画红外小目标的低秩稀疏模型开始，逐渐构建了多个像素级标注的小目标数据集，并将其用于探索新型的注意力机制及其在深度网络中更多样的应用模式，最后实现了深度神经网络与传统红外小目标检测模型的有机融合。主要工作如下：

（1）针对现有的稀疏约束无法有效区分真实目标与同样稀疏的强边缘残留这一问题，构建了重加权红外块张量模型，旨在通过对不同的图像局部结构施加不同的权重实现对稀疏元素的选择性抑制。首先，该模型通过由堆叠滑动窗口采样的图像块来构造红外块张量，将小目标与背景的分离问题建模为张量鲁棒低秩恢复问题，以更好地保存图像块内部的空间相关性。其次，借助结构张量，设计了逐元素的局部结构权重代替原始的全局参数，使得模型在迭代过程中能够根据局部结构权重自适应地调整收缩阈值，从而降低由背景中相对稀疏的干扰物引起的虚警。最后，依据目标块张量的稀疏度重新设计了模型迭代的终止条件，并采用稀疏性增强权重用于减少模型的迭代轮次。相比于其他低秩稀疏分解方法，对于复杂背景下的红外弱小目标，该模型能够在大幅提高检测速度的同时，更好地抑制强起伏云杂波的干扰。

（2）针对模型驱动的红外小目标检测方法判别能力不足、超参数对图像场景变化敏感等问题，设计了一个双向非对称的注意力调制模块并将其嵌入基准网络中，旨在以端到端的方式从标记数据中自动学习红外小目标的语义特征表示。首先，构建了一个单帧红外小目标检测的基准数据集，以五种不同的形式对其进行了标注，并在此基础上对红外小目标的稀疏性、尺度分布、亮度分布等特性进行了相应的统计分析。其次，为了克服特征分辨率与语义层次之间的矛盾，还构建了一个双向非对称的注意力调制模块以实现高层语义信息和目标细节信息的跨层交换。其中自顶向下的调制通路采用全局注意力模块，用于将网络高层特征的语义信息反馈到低层特征，编码目标上下文；而自底向上的调制通路则采用局部注意力模块，用于将低层特征的细节信息嵌入到高层特征中。相比于传统的模型驱动方法，采用该模块的网络能够显著提升红外小目标检测的效果。

（3）受注意力机制和激活函数两者相似性的启发，提出了一类能够根据上下文信息对特征进行选择性激活的注意力激活单元，旨在对神经网络中卷积提取的每一层特征进行逐层的动态精炼。考虑到激活单元的局部性需求，设计了一系列仅聚合局部上下文且也只作用于局部特征的轻量级注意力模块，包括局部通道注意力模块、局部空间注意力模块、混合注意力模块、多尺度混合注意力模块。通过将网络中原有的激活函数替换为注意力激活单元，可以构建出相应的全注意力网络。由于在低层网络中便开始抑制无关特征、强调相关特征，全注意力网络能够更为高效地编码高层语义。此外，出于在更大规模的数据集上验证小目标检测算法的目的，还构建了一个与红外弱小目标具有相似特性的弱小冰山检测数据集。在多个计算机视觉任务上的消融实验与对比实验表明，给定相同的宿主网络，相比于其他激活单元，注意力激活单元能够以较大的幅度提升各类网络的性能。

（4）针对现有特征融合方式忽视尺度不一致性这一问题，给出了一个注意力特征融合框架，旨在通过聚合多尺度的特征上下文实现对融合权重的动态合理分配。该框架统一了短跳连接、长跳连接等多种特征融合场景，并通过构建多尺度通道注意力模块，满足了不同尺度的物体对于各自相应大小的感受野的要求，避免了融合权重对于特定尺度的偏向性。此外，该框架还支持以迭代的方式持续优化注意力模块的输入特征，即采用另一层注意力特征融合来聚合待融合特征，从而进一步提升特征融合的性能。大量的消融实验与对比实验表明，相比于增加网络深度，将网络中原有的相加、拼接等线性融合方式替换为注意力特征融合从而赋予网络动态选择特征融合权重的能力，能够更加高效地提升网络性能。

（5）针对红外小目标本征特征稀缺的问题，结合数据驱动的深度卷积网络和模型驱动的局部对比度度量方法，提出了注意力局部对比度网络，旨在通过同时利用标记数据和领域知识来提升检测性能。借助特征图的循环移位技巧，模块化后的局部对比度度量方法作为深度网络中具有特定物理机理的非线性特征变换层，能够显式打破有效感受野的限制、捕获局部特征与区域上下文之间的交互关系。通过同层的并行多分支架构以及深度网络内生的特征金字塔，传统方法中小目标的最佳对比度选取问题可以转换为一个两阶段的多尺度特征融合问题，从而网络可以通过自底向上的局部通道注意力调制模块进一步优化跨层的多尺度对比度特征。从模型驱动方法的角度看，注意力局部对比度网络其实是将传统模型中过于简单的均值、最大值等特征替换为网络从标注数据中学到的语义特征，从而大幅提高了对于红外小目标的检测能力。

关键词：小目标检测，红外图像，低秩稀疏分解，注意力机制，深度学习

ABSTRACT

Infrared small target detection is the key technology of infrared searching and tracking system, and its performance determines the response time for early warning systems or precision guidance systems. Limited by the characteristics of infrared imaging and the needs of long-distance imaging, the infrared small target is scarce of intrinsic features. Besides, due to the heavy background clutter and complex image scenes, despite years of development, infrared small target detection still faces huge difficulties and challenges in terms of accuracy and robustness. To tackle these issues, this dissertation has carried out related research along with the development context of model-driven methods to data-driven methods and then model-driven deep learning. First, a low-rank plus sparse decomposition model is built to more accurately describe the infrared small target. Then to explore new attention mechanisms and their more diverse applications in deep networks, several pixel-level annotated small target datasets are constructed. Finally, an end-to-end model that combines the deep neural networks and traditional model-driven methods is proposed for infrared small target detection. Detailed contributions are summarized as follows:

(1) Aiming at the problem that the existing sparse constraints cannot effectively distinguish the real target from the same sparse strong edge residue, a reweighted infrared patch-tensor model is constructed to selectively suppress sparse elements by applying different weights to different image contents. First, to dig out more information from the non-local self-correlation property in patch space, a new infrared patch-tensor model is constructed and the separation of small targets and background is modeled as a tensor robust low-rank restoration problem. Secondly, based on structure tensor, element-wise local structure weights are designed to replace the original global parameter, so that the model can adaptively adjust the shrinkage thresholds according to the local structure weights in iterations, thereby reducing the false alarms caused by the relatively sparse background components. Finally, a new stopping criterion is designed according to the sparsity of the target patch-tensor, and the sparsity enhancement weight is used to reduce the iteration rounds. Compared with other low-rank plus sparse decomposition methods, this model can greatly improve the detection speed while better suppressing the complex cloud clutter.

(2) Model-driven infrared small target detection methods suffer from the problem of insufficient discriminative ability and the sensitivity of hyperparameters to image contents. To tackle these issues, an asymmetric bi-directional attentional modulation network is designed to automatically learn the semantic features of infrared small targets in an end-to-end manner. First, a single-frame infrared

small target detection benchmark dataset is constructed and annotated in five different forms, followed by a statistical analysis of the characteristics of the infrared small target. Secondly, to overcome the contradiction between feature resolution and semantic level, an asymmetric bi-directional attention modulation mechanism is proposed to achieve a cross-layer exchange of high-level semantic information and target detail information. The top-down modulation pathway adopts the global attention module, which is used to feedback the semantic information of high-level features to the low-level features, and encodes the target context; the bottom-up modulation path uses the local attention module to embed the details of the low-level features in the high-level features. Compared with the traditional model-driven methods, the proposed network can significantly improve the performance of infrared small target detection.

(3) Inspired by the similarities between the attention mechanism and the activation function, a novel type of activation units called attentional activation units are proposed, which can selectively activate features based on contextual information in a layer-wise manner. To meet the locality requirement, attentional activation units are a series of lightweight attention modules that only aggregate local feature contexts, e.g., the local channel attention module. By replacing the original activation functions in the network with the attentional activation units, a fully attentional network can be constructed, which can encode high-level semantics more efficiently since irrelevant features are suppressed in early stages. Besides, to verify the small target detection methods on a larger-scale dataset, a dim iceberg detection dataset is constructed which shares the similar characteristics of infrared small targets. Ablation experiments and comparative experiments on multiple computer vision tasks show that, given the same host network, the attentional activation unit can greatly improve the performance of various networks compared to other activation units.

(4) To tackle the scale inconsistency issue of feature fusion in deep networks, a uniform and general framework called attentional feature fusion is proposed, which is applicable for most common scenarios, including feature fusion induced by short and long skip connections as well as within Inception layers. To better fuse features of inconsistent semantics and scales, a multi-scale channel attention module is constructed, which addresses issues that arise when aggregating feature contexts given at different scales. It is also demonstrated that the initial integration of feature maps can become a bottleneck and that this issue can be alleviated by adding another level of attention, which is referred to as iterative attentional feature fusion. Given a comparable number of parameters, models with attentional feature fusion outperform state-of-the-art networks on multiple datasets, which suggests that more sophisticated attention mechanisms for feature fusion hold great potential to yield better results compared to their direct counterparts.

(5) To tackle the issue of minimal intrinsic characteristics, a novel model-driven deep network named attentional local contrast network is proposed for infrared small target detection, which combines discriminative networks and conventional model-driven methods to make use of both labeled data and the domain knowledge of local contrast prior. With the feature map cyclic shift trick, the modularized local contrast measurement method, as a nonlinear feature transformation layer with the specific physical mechanism, can explicitly break the limitation of the effective receptive field and capture the interaction between local features and their regional contexts. To highlight and preserve the subtle information of small targets, a bottom-up local attentional modulation module is adopted to dynamically encode the smaller scale details into high-level feature maps. From the perspective of the model-driven methods, the attentional local contrast network replaces the simple features such as mean and maximum values with the semantic features learned from the annotation data, thus greatly improving the performance of detecting infrared small targets.

Keywords: Infrared image, small target detection, low-rank plus sparse decomposition, attention mechanism, deep learning

目　录

图表清单

注释表

$\mathbb{R}^n$	n 维实数空间	$\boldsymbol{\mathcal{X}}^{(k)}$	第 k 次迭代的张量 $\boldsymbol{\mathcal{X}}$
$\boldsymbol{X}_{(i)}$	张量 $\boldsymbol{\mathcal{X}}$ 按照模式-i 展开的矩阵	$\text{fold}_i(\boldsymbol{X}_{(i)})$	将展开矩阵 $\boldsymbol{X}_{(i)}$ 恢复成原张量 $\boldsymbol{\mathcal{X}}$
$\text{rank}(\cdot)$	矩阵或张量的秩	$\text{CTrank}(\cdot)$	张量的 Tucker-秩
$\text{vec}(\cdot)$	矩阵或张量的向量化算子	$\mathbf{0}, \mathbf{1}$	全 0、全 1 矩阵或张量
$\|\cdot\|_0$	矩阵或张量的 ℓ_0 范数	$\|\cdot\|_1$	矩阵或张量的 ℓ_1 范数
$\|\cdot\|_{\text{F}}$	矩阵或张量的 Frobenius 范数	$\|\boldsymbol{X}\|_*$	矩阵 $\boldsymbol{X}$ 的核范数
$*$	卷积算子	∇	梯度算子
$\odot$	逐元素点乘	$\bar{\otimes}$	张量积
$\oplus$	带广播机制的逐元素相加	$\otimes$	带广播机制的逐元素相乘
$\exp(\cdot)$	指数运算	$\text{diag}(\boldsymbol{x})$	向量 $\boldsymbol{x}$ 的对角化矩阵算子
$\mathcal{S}_\mu(\cdot)$	参数为 μ 的阈值收缩算子	$\mathcal{D}_\mu(\cdot)$	参数为 μ 的矩阵奇异值收缩算子
$\mathbf{F}$	网络的特征图	$\mathbf{F}_{[c,i,j]}$	特征图上 $[c,i,j]$ 位置的元素
Concat	特征图拼接	$\uplus$	特征图融合
σ	Sigmoid 函数	δ	线性整流单元
$\mathcal{B}$	批标准化	$\sin(\cdot)$	正弦函数
$\boldsymbol{\mathcal{W}}_{\text{LS}}$	局部结构权重	$\boldsymbol{\mathcal{W}}_{\text{SE}}$	稀疏性增强权重
$g(\cdot)$	非线性门控函数	$\mathbf{M}(\cdot)$	多尺度注意力模块
$\mathbf{L}(\cdot)$	局部注意力模块	$\mathbf{G}(\cdot)$	全局注意力模块
$\widehat{\mathbf{L}}(\mathbf{X})$	特征图 $\mathbf{X}$ 的局部特征上下文	$\widehat{\mathbf{G}}(\mathbf{X})$	特征图 $\mathbf{X}$ 的全局特征上下文
PWConv	逐点卷积	DWConv	逐深度卷积
$\mathbf{D}^{(x,y)}$	(x,y) 方向的局部对比度特征	$\mathbf{S}^{(x,y)}$	沿 (x,y) 方向循环移位的特征图

缩略词

缩略词	英文全称	中文全称
ADMM	Alternating Direction Method of Multipliers	交替方向乘子法
BSF	Background Suppression Factor	背景抑制因子
BN	Batch Normalization	批归一化
DWConv	Depth-wise Convolution	深度卷积
FC	Fully Connected	全连接
FCN	Fully Convolutional Network	全卷积网络
FPN	Feature Pyramid Network	特征金字塔网络
GAP	Global Average Pooling	全局平均池化
GENet	Gather-Excite Network	聚集-激活网络
GFLOPs	Giga Floating Point Operations	10 亿次浮点运算
IPI	Infrared Patch-Image	红外块图像
IoU	Intersection over Union	交并比
LSNRG	Local Signal to Noise Ratio Gain	局部信噪比增益
MPCM	Multi-scale Patch-based Contrast Measurement	多尺度块对比度度量
NAG	Nesterov Accelerated Gradient	Nesterov 加速梯度
PWConv	Point-wise Convolution	逐点卷积
ReLU	Rectified Linear Unit	线性整流单元
ROC	Receiver Operating Characteristic	接收机工作特性
RPCA	Robust Principal Component Analysis	鲁棒主成分分析
ResNet	Residual Network	残差网络
SENet	Squeeze-and-Excitation Network	挤压-激发网络
SCRG	Signal to Clutter Ratio Gain	信噪比增益
SVD	Singular Value Decomposition	奇异值分解

第一章　绪论

1.1　立题背景阐述

遥感（Remote Sensing）是一种在远离感兴趣目标、非直接接触的情况下，利用目标自身辐射或反射的电磁波对目标进行测量、分析和判定的技术。在遥感成像的波段中，红外遥感主要利用目标的热辐射特性进行探测，可以在夜间以及雨、雾等多种天气状态下工作，抗干扰能力强，且能够探测到雷达探测不到的低空目标、弥补盲区，是合成孔径雷达（Synthetic Aperture Radar, SAR）图像和可见光遥感图像的重要补充。由于是被动成像，红外传感器具有体积小、重量轻、隐蔽性好的优点，通常可搭载于卫星、飞机、无人机等机动平台上，被广泛运用于国计民生的各个领域。例如，在民用领域，利用其较强的雨雾穿透能力，红外遥感常被用来监测森林火灾、预报天气以及辅助恶劣天气下的自动驾驶等。在军事上，红外成像技术可用于远距离目标搜索、跟踪、制导以及战场情报侦查、监视、预警等。

受限于红外成像性质、分辨率以及探测系统需要及早预警的需求，远距离感兴趣目标往往在视场中以缺少形状、纹理特征的小目标形态出现的。需要注意的是，虽然红外小目标（Small Target）和通用视觉任务中的小目标（Small Object）均被称为小目标，但实际上两者所面临的具体问题存在着不小的差异，重点表现在以下几点：

1. 红外小目标本征特征缺乏的问题更为严重：在 Microsoft COCO[1]、ImageNet[2] 这些可见光波段的通用视觉数据集中，定义小目标的标准通常被认为是小于图像大小的 1%[3]，大约 32×32 个像素。然而根据国际光学工程学会的定义，红外小目标的尺寸大约为 2×2 至 9×9 个像素，或者小于图像大小的 0.12%[4]。更小的目标尺寸，意味着更少的形状、纹理特征，也就更容易与背景中的杂波干扰一同被检测或抑制。其本征特征的严重缺乏也导致红外小目标检测通常且只能被建模为一个前景和背景的二分类问题，不再像通用视觉检测问题中进一步区分物体的种类。然而，天空、海空、陆空等不同场景下的小目标真正对应的物体及其特点并不完全相同，存在着较大的类内方差。
2. 红外小目标更为缺乏上下文信息：通用数据集大多在自然或者室内场景下采集，目标物体与周围背景之间存在着较强的共生关系和空间位置关系[5]。比如在人脸检测任务中，身体的其他部位能够为检测器提供高可信度的额外信息[6]。然而，在红外小目标检测场景中，目标物体通常为飞机、导弹、装甲车辆等非合作性目标，与周围环境缺少语义上下文上的强关联关系。
3. 真实的红外小目标数据极为稀缺。与 ImageNet、Open Images[7] 这些上百万乃至千万规模的可见光数据集不同，用于中波、短波红外成像的锑化铟、碲镉汞、砷化铟、锑化镓传

感器受到各国的严格管制和出口限制，成像器件获取不易。此外，红外小目标往往是各类高空高速的飞行器，数据获取难度同样远远大于可见光波段的小目标数据。因此，目前可公开获取的数据数量不足严重制约了红外小目标检测领域的发展。

经过多年的发展，红外小目标检测领域取得了长足的进步。一方面，学者们对该具体问题的认识，即领域知识也在不停地深化。从早期将红外小目标简单建模为平滑背景中的突出物[8,9]，到后来考虑了图像边缘结构、具有高局部对比度的物体[10,11]，再到近年来将其建模为低秩背景中的稀疏分量[12,13]，模型驱动的检测方法性能在逐渐地提高。然而，目标的局部对比度、稀疏性先验实际上是一种假设，作为归纳偏置引入，但是真实红外图像场景的复杂程度往往超出建模背后对于目标和背景的假设，比如背景中通常存在同样满足高局部对比度或者稀疏性的干扰结构，容易造成目标的漏检或者虚警。因此，如何构建更符合红外小目标特点的模型，特别是性能整体更好的低秩稀疏模型，使得模型假设能够更为吻合数据，对于进一步提升红外小目标检测方法的性能至关重要。

另一方面，随着深度学习在计算机视觉各个领域上的突飞猛进，一些学者也开始将其运用于一些小规模的私有红外小目标数据集[14]。然而，红外小目标既 ‘‘弱’’ 又 ‘‘小’’ 的特点决定了一味增加网络深度无法带来检测性能持续的提升，反而存在随着感受野不断扩大，小目标特征被背景特征淹没的风险。因此，在有限的网络深度下，如何增强网络中每一层所提取特征的表示能力，是一条提升红外小目标检测性能更有效的途径。受人类视觉系统具有从复杂多变的环境中快速且精准地选取感兴趣目标能力[15] 的启发，近年来，多种注意力机制模型[16,17] 被引入深度学习[18,19] 中，使得自然语言处理[20]、目标检测、图像分类[21] 等领域取得了长足的进步。得益于其根据长程或全局上下文对权重系数的分配，深度网络可以更高效地学习到数据的高层语义信息[22,23]，而无需一味地增加网络深度，非常适合红外小目标检测这种网络深度受限同时又对语义判别能力有较高要求的任务。

近十年来，在中美俄各国乃至印度的高超声速武器快速发展的背景下，伴随着目标和成像平台之间的相对运动速度在不停地加快、对于小目标检测的反应时间要求也在不停地提升，红外小目标检测的研究重心逐渐从传统需要时间积累的序列方法向单帧图像检测方法转变[12]，不仅要求检测方法要在缺少时空信息的情况下以高准确度检测出小目标，还要求其能够对复杂多变的图像场景具有较强的鲁棒性。然而，目前尚无一个符合国际光学工程学会定义、具有高质量标注且公开的单帧检测数据集，这严重制约了红外小目标领域的发展。

在以上的大背景下，一方面，在理论上如何更好地模拟人类视觉系统的信息加工和表征机制，特别是选择性注意机制，将图像处理、模式识别、机器学习与认知神经科学的理论学说进行有机的融合，是相关理论发展的必然要求；另一方面，如何构造相应的红外小目标单帧检测数据集，并在检测任务中运用低秩稀疏分解理论、注意力机制等建模工具，使得检测方法能够有效克服背景干扰，快速、精确且鲁棒的定位目标，对于提高精确制导武器的作战距离和反应

速度、增强信息化战争中的战略防御能力有着极为重要的实际意义。

1.2 国内外发展与研究现状

本节将针对低秩稀疏分解、注意力机制以及红外小目标检测三方面，分别回顾其国内外发展历程与研究现状，以此为本文的立题动机提供核心依据。

1.2.1 低秩稀疏分解的发展与现状

受稀疏表示[24,25]（Sparse Representation）和压缩感知[26–28]（Compressed Sensing）巨大成功的推动，低秩模型是近年涌现出来的、能够鲁棒且高效地处理高维数据的新理论。其中，矩阵的秩被重新阐释为数据的二阶稀疏性度量。在低秩矩阵恢复衍生的诸多模型中，低秩稀疏矩阵分解，也被称作鲁棒主成分分析[29]（Robust Principal Component Analysis，RPCA），旨在寻求满足一定约束条件且使某损失函数值尽可能小的低秩和稀疏矩阵（元素稀疏或列稀疏），在视频运动目标检测[30,31]、人脸建模、视频去噪[32]等领域具有广泛且重要的应用。

通常，矩阵低秩模型采用核范数[33]代替秩函数来刻画潜在的感兴趣矩阵，并且通过求解基于核范数的凸优化问题来得到相应的低秩解。虽然核范数是谱范数单位球上的最紧凸包[33]且易于求解，但两者之间存在很大的差别，核范数诱导低秩解的能力较弱。对于某些特定结构的矩阵低秩恢复问题，这类凸松弛方法往往无法很好地刻画待求解数据的特点，存在失效的风险。作为矩阵秩的多项式函数表示[34]，Max 范数是一种更紧的凸松弛约束[35]，也被运用于矩阵低秩稀疏模型中[36]，可获得比核范数更好的分解效果。为了取得更好的性能，更多的研究者将研究重点转移到基于非凸代理函数的近似逼近问题[37]。作为核范数更一般化的非凸推广，将 p 范数应用于矩阵的奇异值可以构建得到 Schatten-p 范数，以此作为低秩矩阵的正则项可以进一步取得更好的分解效果[38,39]。出于优化大奇异值并不会影响矩阵秩大小的考虑，Hu 等人[40]提出了截断核范数（Truncated Nuclear Norm），在迭代过程中仅衰减排序靠后的小奇异值，保留大奇异值。作为低秩正则项，截断核范数理论上能够比核范数更逼近秩函数，但其引入的用来截断奇异值的秩作为超参数需要在实际应用中不断调整。Gu 等人提出加权核范数最小化[41]（Weighted Nuclear Norm Minimization），根据奇异值数值大小来对其自身进行重加权，从而实现对不同奇异值进行不同程度的惩罚。Xie 等人更进一步，将重加权核范数与 Schatten-p 范数融合，得到了更加非凸的低秩项约束，即加权的 Schatten-p 范数[42–44]。

为了刻画更为精细的数据分布结构，Liu 等人利用样本之间的相互表示构建了低秩表示模型（Low-Rank Representation），将 RPCA 原有的单个低秩子空间假设进一步推广到了相互独立的多子空间假设[45]。Liu 和 Yan 进一步提出了隐式低秩表示[46]（Latent Low-Rank Representation），使用已观察到的数据和未观察到的隐藏数据共同来作为字典以应对数据不足或被严重污染的情形。RPCA 另一个不足在于其只能处理二维矩阵数据，然而在现实世界中，数据普遍以多维方

式出现，也称为张量，比如彩色图像和视频。如果使用 RPCA 来处理，需要先将这些数据的某些维度拉成向量，这势必会破坏数据原始的空间结构，导致这些数据的本质和内部特性无法被充分刻画和挖掘[47]。为此，作为低秩矩阵恢复问题在更高维度的推广，张量低秩恢复逐渐受到研究者们的关注。需要注意的是，计算张量的核范数是非确定性多项式时间-难（Non-deterministic Polynomial-time Hardness, NP-Hard）问题[48]，其秩的定义并不唯一，不同的定义可以导出不同形式的张量分解。目前较为流行的张量分解有 CANDECOMP/PARAFAC（CP）分解[49] 和 Tucker 分解[50]，分别是矩阵奇异值分解（Singular Value Decomposition, SVD）和主成分分析的高阶推广。Kilmer 和 Martin 在推导张量逆、伪逆、转置等概念的基础上，提出了一种新的张量分解方式，张量奇异值分解（Tensor SVD），可以将一个张量分解为多个张量乘积的形式[51]。基于张量奇异值分解，Lu 等人将二维 RPCA 推广到了更高维的张量鲁棒主成分分析[52]（Tensor Robust Principal Component Analysis）。Goldfarb 和 Qin 则是采用凸的 Tucker 秩作为低秩正则项，提出了单例模型（Singleton Model）、混合模型（Mixture Model）等多种鲁棒张量恢复模型[53]。由于 Tucker 秩的定义基于矩阵的秩，张量低 Tucker 秩优化问题本质上就是矩阵低秩优化问题，因此采用上一段中提到的非凸代理函数可以进一步提高张量低秩恢复的效果[54–56]。

除了建模数据本身，低秩分解在深度神经网络的压缩和加速领域也有着广泛的运用[57]。在网络中，不管是全连接层还是卷积层（im2col 向量化实现），本质均为矩阵相乘[58]，而其权重矩阵往往具有高度冗余性。利用低秩分解压缩和加速网络的基本原理是通过对网络中的全连接层或卷积核进行矩阵或张量分解，从而在不损失网络预测精度的同时，大幅减少网络的参数量和计算时间。对于 AlexNet 和 VGG 等早期的卷积神经网络来说，其网络参数大约有 90% 来源于全连接层，而其 90% 的计算量则主要消耗在卷积层，特别是浅层的卷积运算。Novikov 等人通过对 VGG 网络[59] 全连接层的密集权重矩阵进行 Tensor-Train 分解[60]，将其参数数量压缩为原来的两万分之一[61]，极大降低了 VGG 网络总体的参数量。Kim 等人使用 Tucker 分解对卷积核的权重张量进行低秩分解，将 VGG-16 的速度提高为原来的 3.6 倍[62]。对于卷积而言，另一种更为高效的压缩和加速方式是将大卷积核分解为多个小卷积核的串联[59]。某种程度上，这类分解可以看作是对原始卷积核施加了特定的正则化[59]，并在其间注入非线性。例如，一个 5×5 卷积层可以在有效感受野[63] 近似的情况下分解为两个 3×3 卷积层的叠加，从而将计算量和参数量减少 28%。Inception 网络[64] 更进一步，将这种低秩分解推广到非对称情形，即一个 $n\times n$ 的卷积核可以由两个串联的 $n\times1$ 卷积核和 $1\times n$ 卷积核近似。深度可分离卷积[65]（Depth-wise Separable Convolution）则可以看作是卷积在空间和通道维度上的低秩分解[66]。正是这些分解使得现代的卷积神经网络可以在低功耗的移动和边缘计算设备上进行部署[67]。

1.2.2 注意力机制的发展与现状

注意力机制的早期研究主要集中于认知心理学和神经科学领域，其结果为后续选择性注意的模型化计算提供了可以借鉴的生理依据[68]和理论指导[69,70]。例如，Rodieck 在猫的视网膜上发现了中心-周边机制，并利用高斯差分（Difference of Gaussian）函数来描述这种同心圆型的感受野[71]。Koch 和 Ullman 提出了显著性图（Saliency Map）、赢者通吃（Winner Take All）、抑制返回（Inhibit of Return）等被后续研究广泛采用的概念与机制[72]。在人类视觉中，选择性注意是由两种加工机制相互制约、相互影响的：一种是自底向上（Bottom-Up）的刺激驱动，如视觉情境中与周围具有较强对比度或明显不同的显著物体对注意的捕获；另一种是自顶向下（Top-Down）的目标/记忆驱动，例如视觉工作记忆表征引导注意偏向到与之相同或相似的视觉刺激[73]。在计算机视觉中，通过算法模拟前一种注意加工机制、提取图像或视频中的显著区域的研究被称为视觉显著性检测（Saliency Detection）。

1998 年，Itti 等人提出了首个计算图像视觉显著性的模型[74]。该模型提取颜色、亮度、方向这三个底层特征，以多尺度的方式计算各自特征通道下的中心-周边对比度显著性图，最后对归一化融合后的特征图采用赢者通吃机制标注出图像中的显著区域。在 Itti 模型的基础上，通过采用不同的视觉特征[75,76]、不同的显著性度量方式[77,78]、不同的特征显著性图结合方式[79,80]，涌现出了大量的改进方法进一步提高了检测性能。不同于上述空域显著性检测方法，Hou 和 Zhang 基于图像显著区域往往对应于其对数频谱中突出部分的观察，提出了频谱残差（Spectrum Residual）方法[81]，首次在频域内进行显著性计算且实现简单。后续工作指出提取显著性区域的关键在于傅立叶变换后的相位谱而非幅度谱[82,83]，通过抛弃振幅信息，可以进一步简化计算。近年来，随着机器学习领域的飞速发展以及大规模公开数据集[84–86]的陆续出现，不管是人眼凝视点检测任务，还是显著性物体检测任务[84]，基于深度神经网络的方法逐渐成为主流。同语义分割任务类似，目前大多数基于全卷积神经网络（Fully Convolutional Network, FCN）的显著性物体检测方法主要致力于探索更为有效的网络结构，通过融合不同网络层的特征，使得网络输出既能够保留空间细节信息，也具有较强的语义判别能力。例如，Hou 等人通过向整体嵌套边缘检测器[87]（Holistically-Nested Edge Detector）的跳层结构中引入短跳连接，可以更为充分地利用 FCN 提取的多尺度特征，从而获得更好的检测性能[88]。Wang 等人提出的注意显著性网络（Attentive Saliency Network），利用人眼凝视点检测来指导显著性物体检测，在统一的神经网络中采用在较高层中编码的人眼凝视点特征图来辅助较低网络层上的显著性物体检测任务[89]。

近年来，受人类视觉注意力机制的启发，一种同样名为注意力机制的特殊网络结构在深度神经网络得到了广泛的运用，并在计算机视觉和自然语言处理领域的许多任务上表现出了较好的性能。这种神经网络中的注意力模块，可以被视为一种自顶向下与任务相关的注意力机制，通过端到端的隐式学习能够 ‘‘迫使’’ 神经网络关注文本或图像中与任务最相关的部分。在计算机

视觉中，由于卷积神经网络仍然占据着主导地位，其中主流的注意力机制大多表现为一些可以微分、用来让网络自己调节权重以适应不同任务的模块。典型的代表是挤压-激发网络（Squeeze-and-Excitation Network, SENet）[21]，通过显式建模特征通道之间的相互依赖关系来自适应地对特征进行加权，这类模型也被叫作通道注意力机制。后续研究表明[90]，避免降维对于学习通道注意力非常重要，采用一维卷积可以在显著降低模型复杂度的情况下，有效地实现不降维的局部跨信道交互。与通道注意力机制相对应的是空间注意力机制，既可以用来强调或抑制空间分布上的某些特征，也可以对特征图进行几何变换。空间变换网络[91]（Spatial Transformer Networks）使用输入特征自适应地生成仿射变换的超参数，结合双线性插值输出，可以实现对各个目标区域的特征图的对齐。可变形卷积网络（Deformable Convolutional Networks）根据输入特征动态调整反映每个像素贡献的有效感受野[92,93]，可以取得比传统刚性卷积核更好的效果。更进一步，Park 和 Woo 等人指出融合空间注意力和通道注意力可以让网络同时学到关注 ‘‘什么’’ 和 ‘‘哪里’’，从而获得更好的性能[22,94]。近年来，随着 Transformer 网络[20] 在机器翻译以及自然语言处理领域的大获成功，自注意力（Self-Attention）机制也逐渐开始在计算机视觉领域中得到应用。为了克服卷积核固有的局部性，Bello 等人提出了一种独立的二维相对自注意机制[95]，并将其抓取的全局语义特征与卷积特征结合以得到增强型的卷积。由于该机制能够在纳入相对位置信息的同时维持平移不变性，使得其适用于图像。Fu 等人提出的双注意力网络[96]（Dual Attention Network）在自注意力机制中实现了空间注意力和通道注意力的融合。Carion 等人将 Transformer 引入目标检测，将其重新建模为一个集合预测问题，简化了检测流水线，有效地消除了非极大值抑制、锚框生成等原需手工设计的组件[97]。然而，这类全局自注意力或者非局部网络[98]（Non-local Neural Networks）的共同缺点在于其空间复杂度非常高[99,100]。为此，Ramachandran 等人提出了一种局部自注意力模块[66]，将其用于替换网络中全部的卷积可以构建出全注意力网络，并且使网络参数量大幅减少。Zhu 等人[99] 和 Cao 等人[100] 则分别采用非对称金字塔模块和类 SENet 模块来降低非局部网络所需要的显存和计算量。

1.2.3 通用视觉任务中的小目标检测现状

虽然相较于通用视觉任务中的小目标（Small Object）检测，红外小目标（Small Target）检测所面临的挑战更加极端，但近十年来，作为计算机视觉领域的重要瓶颈之一，学术界对前者进行了大量研究，其成果对于红外小目标检测仍然具有相当的参考价值。因此，有必要对通用视觉任务中小目标检测的发展与现状进行一定的回顾。

在通用视觉任务中，检测小目标的一大难点在于有效样本的相对不足。例如在 COCO 数据集中，一方面，包含小目标的图像就相对较少，这使得检测模型在优化过程中会更倾向于提升中大型目标的检测性能而忽视小目标；另一方面，小目标本身覆盖的区域就远小得多，缺少位置上的多样性。面对该问题，最为直观的方式是针对小目标特点设计相应的数据增广（Data

Augmentation）策略。为此，Kisantal 等人提出利用过采样（Oversampling）来处理小目标图像样本不足的问题，而小目标覆盖不足的问题则通过在图像中多次复制粘贴小目标、增加小目标匹配的锚框（Anchor）数目来缓解[101]。此外，小目标样本不足的问题还可以通过设置更为稠密的锚框来增加匹配到的有效样本数量。基于该观察，Zhang 等人提出了一个等比例间隔原则用于保证不同尺度的锚框在图像中的密度大致相同，并且对小目标采用更为宽松的匹配标准，从而实现不同尺度目标数量的大致平衡[102]，使得网络能够更为充分地学习小目标的特征。

小目标检测的另一大难点在于由目标较小的尺寸导致的本征特征不足。为了保证目标定位精度、避免感受野过大导致的小目标特征丢失，单发检测器（Single-Shot Detector, SSD）采用卷积网络中具有较高分辨率的浅层特征检测小目标[103]。然而，浅层特征语义信息较少，判别能力较差，容易导致漏检和虚警。针对该问题，特征金字塔网络（Feature Pyramid Networks, FPN）利用自上而下连接和侧向连接融合网络不同层中不同尺度的特征，可以在基本不增加计算量和参数量的情况下，大幅提升小目标检测的性能[104]。类似的，U-Net 则采用编码器-解码器结构，通过长跳连接在解码端融合高层和低层特征，从而在保持语义信息的同时逐渐恢复目标的空间细节特征[105]。相较于特征金字塔利用内生的多尺度特征，图像金字塔通过构造多尺度的输入图像来检测不同分辨率的目标，特别是通过将小目标上采样可以获得更好的检测性能[106]，然而这也不可避免地带来了大幅增长的计算成本和存储开销。为此，Li 等人提出了感知生成对抗网络（Perceptual Generative Adversarial Networks, Perceptual GAN）利用 GAN 生成小目标对应的超分辨率目标并且叠加着原有小目标的特征图上[107]，从而缩小大小目标之间的表征差异、提高网络对小目标的检测能力。

相较于红外小目标，通用视觉任务中的小目标往往伴随着较强的场景上下文。在小目标自身特征较少时，图像中的上下文语义可以减少目标的模糊性，为检测提供额外的证据[108]。例如，在人脸检测任务中，通过扩大检测器的局部感受野，将人体的其他部位纳入检测，可以提高网络的检测精度[6]。然而，需要注意的是，由于红外小目标的主体往往为高空高速飞行器等非合作目标，与周围背景的语义联系较弱，贸然照搬通用视觉小目标检测增加上下文的方法未必会百分百奏效，存在更大的感受野中的背景特征淹没小目标特征的风险，因此需要在使用时根据红外小目标的特点进行有针对性的选取和设计。

1.2.4 红外小目标检测的发展与现状

传统的红外小目标检测算法主要分为检测前跟踪（Tracking Before Detection）和跟踪前检测（Detection Before Tracking）两大类[109]。前者基于小目标运动轨迹的时空连续性，利用粒子滤波[110] 等手段对小目标进行跟踪，在众多目标的潜在轨迹中检测真实目标[111]。这类算法的核心为跟踪算法[112]，对于信噪比较低的场景较为鲁棒，但是由于需要积累较多的帧数来提高信噪比[113,114]，其实时性较差，在实际工程中应用较少。跟踪前检测算法首先对图像进行逐帧

处理，然后在单帧图像中分割出潜在目标，最后在序列图像中根据目标的运动特点确认真实目标。该类算法的复杂程度较低、实时性较好，且易于硬件实现，被广泛应用于实际工程中。一方面，由于超高声速飞行器的快速发展，成像传感器和目标之间存在快速的相对运动。在成像背景快速变化时，3D 匹配滤波器[115] 这类依赖于静态背景假设的传统空时检测方法的性能会迅速降低[116]。另一方面，作为跟踪前检测类算法最为重要的基础，更为可靠的单帧检测效果也往往意味着更好的最终检测性能[12]。因此，近年来，单帧红外小目标检测逐渐受到研究者更多的关注。

早期的单帧红外小目标检测主要采用空域滤波方法，例如 Max-Median 滤波器[9,117]、双边滤波器[118,119]、二维最小均方差滤波器[120–122]、Top-Hat 滤波器[123,124] 等。这类方法假设图像背景缓慢变化且临近像素高度相关，而红外小目标则是破坏这种相关性的存在，旨在通过抑制（减去）背景实现红外小目标的增强和检测。其优点是计算量小、复杂度低、实时性好，但是这类假设往往过于简单。例如，海天交界线或重云杂波中的边缘等背景结构也会像目标一样破坏背景的一致性假设，从而造成大量虚警。虽然有研究者尝试改造这类滤波器试图使其能够抑制边缘结构的干扰[125,126]，但云杂波实际边缘形态的复杂度远超过理想边缘的假设，而且往往需要提前确定滤波器模板、结构元素这些难以获取的先验信息。在较为复杂的背景环境中，这类算法的性能会迅速恶化，容易出现检出率低、虚警率高的问题。为此，研究者们考虑将检测从空域转移到能够更好区分目标和背景干扰的各种变换域中，比如 Fourier 域[127]、模糊空间[128]、梯度矢量场[129]。在变换域方法中，通常将红外小目标看作是图像信号中的高频分量，而背景对应变换后的低频分量。但是背景中的部分噪声和强边缘往往也会出现在高频分量中，因此还需要在频域进行相关处理。为了能够更好地表征高维信号，变换域检测方法大多采用多尺度几何分析方法，如双树复数小波[130]、无下采样 Contourlet 变换[131]、对偶树复小波[132] 等，来分离具有方向性的边缘、保留各向同性的目标。然而，这些变换域方法的计算复杂度更高，很难被应用于实际场景中。

随着显著性检测方法在计算机视觉中的流行，近年来，受人类视觉注意力机制启发的红外小目标检测方法也陆续被提出。部分研究者采用一些接近小目标特点且模拟了人眼视网膜感受野特性的滤波器来生成红外小目标的显著性图，包括高斯差分滤波器[133]、高斯拉普拉斯（Laplacian of Gaussian）滤波器[134–136]、Gabor 差分滤波器[137]、二阶方向导数滤波器[138] 等。频谱残差方法[81]、频率调谐（Frequency-tuned）方法[139] 这些以速度快著称的显著性检测方法也被用来生成初始的显著性图，缩小目标可能存在的范围从而减少后续检测的计算量[140,141]。这些方法均利用了目标和背景的差异，但随着 Chen 等人提出局部对比度度量（Local Contrast Measurement, LCM）模型，利用图像块与其邻域之间的局部对比度来直接构建显著图检测小目标后[11]，越来越多研究者开始抛弃对感受野机制的模拟，致力于如何显式地构造更能够反映小目标特点、更具区分度的局部对比度度量方法[10]。加权局部差异度量[142]（Weighted Local Difference Measure）、多尺

度块对比度度量[10]（Multiscale Patch-based Contrast Measure）、基于熵窗口选择（Entropy-based Window Selection）的对比度度量[143]、基于导数熵的对比度度量[144]（Derivative Entropy-Based Contrast Measure）、导数差异度量[145]（Derivative Dissimilarity Measure）等复杂、更精细的局部对比度度量方案陆续被提出，分别从特征图融合权重、方向选择性、特征选择等方面对原有LCM 模型进行了改进。从某种程度上，这可以看作是空域滤波方法的一种复兴，因为该类方法中分拆出来的每个单一步骤大多都可以通过线性或非线性滤波快速实现[10]。但由于其对小目标局部对比度度量方法的构造更为灵活，可以嵌入并融合更多的领域知识，从而能够取得比传统空域滤波更好的效果。然而，对于高度复杂的实际场景，红外小目标极其微弱、有大量背景干扰时，由于目标和杂波的对比度差异很小，以上这些基于显著性假设的方法容易出现严重的误检。

近年来，低秩稀疏分解也被引入红外小目标检测领域中，并由于其较好的背景抑制性能得到了许多研究者的关注。传统的低秩稀疏分解多被应用于视频序列[30,47]或多张人脸集合中，其中数据矩阵的每一列都对应着向量化后的每一幅图像，而对于单帧红外小目标检测，问题在于算法的输入仅有一幅图像。针对这个问题，基于红外背景图像的非局部自相关性以及小目标的稀疏性，Gao 等人最早提出红外块图像（Infrared Patch-Image, IPI）模型[12]，通过将滑动窗采样的、相互重叠的图像块向量化并且拼接起来，获得满足背景低秩、目标稀疏假设的矩阵，由此可以将红外小目标检测问题转化为低秩稀疏矩阵分解问题，并由加速近端梯度（Accelerated Proximal Gradient）法求解[146,147]。在这种新思路的基础上，如何更好地构建低秩正则项和稀疏正则项来更加准确地刻画复杂的背景和目标成为后续研究者改进 IPI 模型的主要方向。一方面，低秩稀疏分解领域的发展为红外小目标检测提供了丰富的数学工具，比如对背景分量的非凸低秩约束可以有效减少分离后的目标分量中的背景干扰[148–152]、将矩阵形式的 IPI 模型推广到张量形式可以更为充分地抓取数据之间的联系[13,153,154]、将问题模型由仅刻画单个子空间的 RPCA 替换为低秩表示可以更为精细地刻画红外背景图像[155]、交替方向乘子法（Alternating Direction Method of Multipliers，ADMM）等新的优化方法的出现使得模型可以更快地分离出红外小目标和背景图像[156]。另一方面，研究者也针对红外小目标检测问题自身的特点，发展出了相应的目标和背景正则项[13,153,157,158]以及算法终止条件[13]。相较于滤波和显著性检测方法，基于低秩稀疏分解的方法通常能够取得较好效果，但其高昂的计算开销以及超参数对于图像场景变化的敏感性限制了其在实际工程中的应用。

需要注意的是，相较于机器学习，特别是深度学习在计算机视觉各个领域中的突飞猛进，近年来红外小目标领域的主流仍然为上述信号处理范式下的检测方法[145,153,159]。研究者早期主要采用字典学习来检测红外小目标[160]，一般通过使用二维高斯模型来预设超完备目标字典，通过图像块表示系数的稀疏度来判定是否为目标图像块[140,161]。然而，随着成像距离以及环境等因素的改变，小目标本身展现出的少量形状特征也会发生相应的改变，再叠加上复杂的背景干扰，

采用二维高斯函数直接模拟小目标很可能是无效的[12]。为此，Wang 等人和 Lu 等人分别提出通过刻画来源于真实图像的背景字典[162] 或同时刻画目标和背景字典[163] 来消除这种拟合函数与真实目标之间差距的影响。与通用目标检测（Generic Object Detection）不同，采用机器学习检测红外小目标的困难在于其本征特征稀缺。相较于采用方向梯度直方图（Histogram of Oriented Gradients）之类的通用特征[164]，研究者更倾向于针对红外小目标的特点单独设计手工特征[165]。Bi 等人采用局部对比度、滤波残差、边缘特征、熵特征构成特征向量，并由支持向量机（Support Vector Machine）来分类潜在目标[166]。即使采用自动学习特征的全卷积神经网络，往往还需要在网络对红外图像进行分割后，再利用小目标的显著性特征进一步抑制虚警[167,168]。Wang 等人将对抗学习引入红外小目标检测，将降低漏检率和虚警率分解为两个子任务，分别由两个经过对抗性训练的模型处理，每个模型仅着重于降低各自的漏检率或虚警率[14]。然而，这些方法均是在私有且规模较小的红外数据集上训练和测试，且部分方法采用序列数据而非单帧图像，除了目标位置不同外，训练集和测试集存在着严重的重叠现象，从机器学习的角度看来，模型有着过拟合的风险。因此，建立一个公开且符合机器学习规范的单帧红外小目标数据集对于推动该领域的发展具有十分重要的意义。

1.3 立题意义与核心问题

从上面的回顾可以看出，受机器学习、计算机视觉等领域突飞猛进的影响，红外小目标检测领域也在快速发展[12,13,148,153]。然而，低秩稀疏分解[29]、注意力机制[16]、深度学习[19] 这些理论和工具均是针对一般性的通用任务设计的，比如人脸对齐[169]、图像分类[21] 等。这些通用任务所面临的数据特点与实际红外小目标检测所面临的难点之间有巨大的鸿沟，如何考虑红外小目标的特点弥平这种鸿沟，更有针对性地利用和改进这些理论工具从而提高红外小目标检测的性能，是本论文立题的根本。更为具体的，本文主要围绕以下四个核心问题展开：

（1）构建更符合红外小目标特点的模型：低秩稀疏分解虽然为红外小目标检测提供了可以建模的数学工具，但是真实场景中背景和目标的复杂度远远超出了鲁棒主成分分析中的低秩和稀疏假设。此外，滑动窗口采样的图像块的对齐程度也远远低于静止不动的相机采集到的视频序列。因此如何构建更符合红外小目标特点的模型，使用更加符合真实场景的约束刻画目标和背景，对于提高红外小目标检测算法性能具有重要意义。

（2）构建红外小目标检测的基准数据集：在计算机视觉中，Pascal VOC[170]、ImageNet[2]、Microsoft COCO[1] 等大规模基准数据集（Benchmark Dataset）的出现，使得机器学习，特别是大规模深度网络的训练成为了可能，极大地推动了图像分类、目标检测、语义分割等各个领域的发展。然而，在红外小目标检测领域，目前还没有类似的公开基准数据集使得研究者可以训练深度网络，并在统一的标准下测试、比较各个模型的性能。这极大限制了这个领域的发展，也是目前深度学习还未成为红外小目标检测主流方法的首要原因。另一方面，构建红外小目标数

据集也可以提高基于显著性、低秩稀疏分解这些非学习方法的鲁棒性。不同于机器学习范式中的优化强调泛化性，即通过对现有‘‘经验’’（样本）的学习实现对未来未知样本的预测，在信号处理领域，如信号重构、图像修复、图像去噪等，其优化强调对于在给定的已知信号上取得满意的性能。过去的红外小目标检测方法通常在有限的几个序列上测试，其超参数容易过拟合，当图像场景大幅变化时，检测性能会快速下降，而通过在基准数据集验证集上调整超参数，可以增加这些方法对于图像背景变化的鲁棒性。因此，构建开放的红外小目标检测的基准数据集对于推动整个领域的发展具有重要的实际价值。

（3）探索新型的注意力机制及其在深度网络中更多样的应用：红外小目标的特点决定了一味增加网络深度无法带来检测性能持续的提升，反而存在随着感受野不断扩大，小目标特征被背景特征淹没的风险。因此，在有限的网络深度下，如何增强网络中每一层所提取特征的表示能力对于提升小目标检测的性能至关重要。无数的成功案例已经证明，注意力机制可以增强网络的表示能力，特别适用于深度和容量受限的网络[23]。然而在通用的计算机视觉任务中，目标物体通常在图像中占据比较大的比例[3]，注意力机制主要被用来精炼网络提取的特征图，往往出现在网络的中高层[21,171,172]乃至最后一层[96,173]，且偏好抓取全局[21,96,173]或大范围内[172]的相互依赖关系。但是，红外小目标的特点在于既‘‘弱’’又‘‘小’’，与大多强调长程依赖关系的注意力模块背后对目标物体的隐含假设不同，直接照搬针对通用任务设计的注意力模块很可能无法很好地加强小目标的特征。此外，传统卷积神经网络通常采用逐渐下采样特征图的方式来获取高层语义。因此，如何保存小目标的特征使其不在下采样过程中被背景特征淹没，不仅仅需要重新设计下采样方案，还需要根据红外小目标的特点，对注意力模块进行重新设计并且探索其在深度网络中更多样的应用范式。这有助于促进深度神经网络在红外小目标检测领域的应用，从而使得检测算法效果更好，并且更能够应对未来未知而多变的复杂场景。

（4）融合深度网络与传统模型：在低层视觉（Low-Level Vision）领域，模型驱动的深度学习[174,175]（Model-Driven Deep Learning）、深度展开[176]（Deep Unrolling）等工作已经展示了结合领域知识和标记数据对于解决图像恢复、图像去模糊、图像去噪这些问题上的有效性。红外小目标检测的特点决定了其并不完全是一个纯粹的高层语义任务，而是具有一定低层视觉特点的问题。对于红外小目标检测这种目标特征稀缺、数据集规模小的问题，完全抛弃领域知识（Domain Knowledge）、纯粹依靠网络端到端学习（End-to-End Learning）的策略是不明智的。传统方法中的数学模型反映了红外小目标检测的领域知识以及研究者对该问题的深刻认识，而仅仅依赖于数据让网络去学习这种这些先验知识背后的映射关系是相对低效的行为，特别是在样本数较少的情况下。因此，针对红外小目标检测问题，如何有效融合代表领域知识的传统模型和数据驱动的深度网络，设计出检测性能更好、可解释性更强的深度网络，对于该领域未来的发展具有理论和应用上的双重价值。

1.4 研究内容与全文结构安排

为了解决以上核心问题，本论文将围绕低秩稀疏分解和注意力机制对红外小目标检测问题开展具体研究，全文章节的内容安排如图 1.1 所示，其中第二章研究红外小目标与图像背景的低秩稀疏建模，第三章至第五章从注意力机制的角度出发分别针对特征金字塔、激活单元、特征融合模块等深度神经网络的基础结构展开探索，最后第六章研究了深度网络与传统模型的融合框架。与此同时，第三章和第四章还分别构建了人工像素级标注的单帧红外小目标检测数据集以及更大规模的弱小冰山检测数据集用以支撑深度网络的训练与测试。全文内容共分为七章，以下是后续章节的具体贡献：

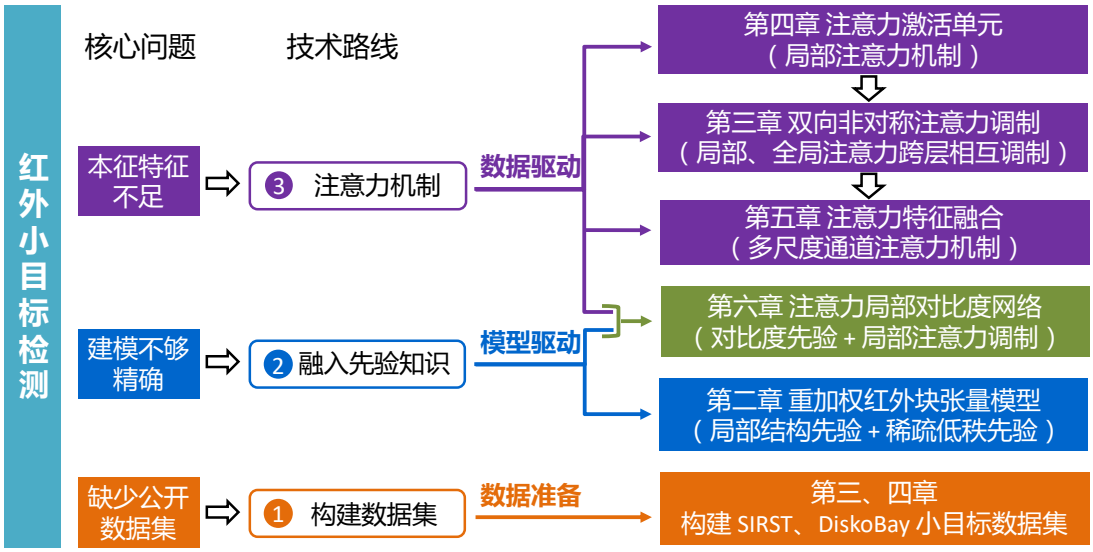


图 1.1　全文内容安排示意图

- 第二章构建了重加权红外块张量模型，旨在增强低秩稀疏分解方法对于稀疏的红外小目标与同样相对稀疏的强边缘残留的判别能力。该模型通过堆叠图像块构造红外块张量，将目标背景分离问题转换为张量鲁棒低秩恢复问题，从而更好地保留了图像块的空间相关性。在此基础上，借助结构张量，设计了逐元素的局部结构权重用来代替原本的全局参数，使得模型在迭代过程中能够自适应地调整收缩阈值，在保留小目标的同时有针对性地抑制强边缘结构。此外，考虑到小目标检测并不需要极小的重构误差这一现实情况，重新设计了模型迭代的终止条件，并采用重加权的稀疏性增强权重用以保证目标块张量稀疏度的稳定下降。相比于其他低秩稀疏分解方法，该模型能够在更好地抑制复杂云杂波干扰的同时，大幅降低优化求解所需的时间。
- 第三章针对模型驱动的红外小目标检测方法判别能力不足、超参数对图像场景变化敏感等

问题，在完成单帧红外小目标检测基准数据集构建与标注的基础上，设计了一个双向非对称的注意力调制模块并将其嵌入基准网络中，用于以端到端的方式从标记数据中自动学习红外小目标的语义特征表示。考虑到红外小目标本征特征不足、语义与分辨率矛盾更大的问题，该模块利用自顶而下的全局通道注意力调制与自底向上的局部通道注意力调制，在特征金字塔不同层之间互相交换高层语义信息和低层细节特征，作为彼此特征调制的指导信息。相比于特征调制方案，双向非对称注意力调制能够更为有效地保存红外弱小目标的特征，避免其被背景特征所淹没。

- 第四章在非线性门控函数的框架下，提出了注意力激活单元框架，作为传统激活单元与注意力机制两者的统一。除了在网络中引入非线性之外，通过聚合相应的通道、空间特征上下文信息，注意力激活单元还可以对特征进行有选择性的激活，实现对于特征图的重新校准。该框架提供了一条构建全注意力网络的途径，通过逐层替换网络中原有的激活函数，在低层网络中及早抑制无关特征、强调相关特征，全注意力网络能够更为有效地编码高层语义。为了在更大规模的数据集上验证小目标检测方法，还构建了与红外弱小目标具有相似特点的弱小冰山检测数据集。相比于其他激活单元，使用注意力激活单元的网络能够在保持相同性能的情况下，大幅减少网络层数和参数数量。

- 第五章针对深度网络中的特征融合问题，给出了一个能够根据特征上下文动态分配权重的通用特征融合框架——注意力特征融合。为了统一短跳连接、长跳连接等多种特征融合场景、克服特征之间语义和尺度的不连续性，构建了一个多尺度的通道注意力模块，通过在注意力模块内部聚合不同尺度的特征上下文匹配不同尺度的物体，以更好地实现特征的选择性融合。此外，该框架还支持以迭代的方式优化注意力模块的输入特征，从而进一步提高最终特征融合的质量。在图像分类、小目标分割等视觉任务上的消融实验与对比实验表明，通过替换原有的特征融合模块，注意力特征融合能够显著提高多种基准网络的性能。与增加网络深度相比，采用更为先进的特征融合方案是一种更加高效的提升网络性能的方式。

- 第六章提出了注意力局部对比度网络，将数据驱动的深度卷积网络和模型驱动的局部对比度度量方法统一到同一个框架下。为了克服红外小目标本征特征不足的问题，该网络首先将局部对比度度量方法模块化，作为特定的非线性特征变换层嵌入深度网络中，依据先验知识捕获局部特征与区域上下文之间的交互关系。然后利用同层的并行多分支架构与深度网络内生的特征金字塔，多尺度下小目标最佳的局部对比度选取问题可以转换为一个两阶段的特征融合问题，使得网络可以通过自底向上注意力特征调制获得与目标尺度最为匹配的局部对比度特征。相比于其他纯数据驱动或者纯模型驱动的方法，本章所构建的注意力局部对比度网络能够同时利用标记数据和领域知识，极大地提高了红外小目标检测的准确

率。

- 第七章对本文的主要工作进行了总结，并对各章节中存在的问题以及后续的研究进行了讨论和展望。

第二章　基于重加权块张量模型的红外小目标检测

红外块图像模型分离出的目标图像中往往残留有强起伏的云边缘，其根本原因在于刻画目标的稀疏约束无法区分稀疏的真实目标与同样相对稀疏的背景干扰物，使得一些目标和背景干扰物在低秩稀疏分离过程中被同时增强或者抑制，从而造成了分离图像中真实目标的丢失或者背景干扰物的残留。为了剔除云边缘等相对稀疏项的干扰，本章构建了一个能够同时利用目标局部结构先验与背景非局部自相关先验的重加权红外块张量（Reweighted Infrared Patch-Tensor, RIPT）模型，通过提取图像的局部结构信息自适应地调整每个像素的背景抑制力度，从而降低目标图像中的虚警成分。此外，考虑小目标检测的实际情况，本章还设计了新的终止条件和稀疏性增强权重以大幅减少算法所需的迭代次数。实验结果表明，RIPT 模型对于复杂背景中的红外弱小目标具有良好的检测效果。出于可重复研究的考虑，本章方法的代码可以从项目主页上获取[1]。

本章的具体内容安排如下：第 2.1 节阐释了红外图像背景的非局部自相关性先验与局部连续性先验之间的互补性，进而强调了本章工作的研究动机与意义；第 2.2 节详细描述了本章所提出的重加权红外块张量模型，并给出了模型求解的优化算法；第 2.3 节针对块张量、局部结构权重、稀疏性增强权重等在 RIPT 模型中所起的作用进行了有效性验证，同时还对比了其他多种红外小目标检测方法以验证 RIPT 模型的整体性能。第 2.4 节对本章工作的内容进行了小结。

2.1　引言

根据第一章的介绍，目前单帧的红外小目标检测方法大致可以分为两类，即基于背景局部连续性的方法[120,145,159] 和基于背景非局部自相关性的方法[12,148]。前者假设图像背景缓慢过渡且临近像素高度相关，而小目标则是破坏这种背景局部连续性的突出物体[124]，通过显式建模小目标与相邻区域之间的差异程度实现小目标检测[145]。这类方法相对简单、高效，但是对于目标尺度变化更为敏感，难以应对复杂多变的现实场景。此外，由于缺乏全局信息，图像中冗余但较为显著的背景结构容易被作为目标得到增强，从而淹没真实的弱小目标。后者以红外块图像（Infrared Patch-Image, IPI）模型[12] 为代表，假设所有的背景图像块均来自于某个或多个低秩子空间，而小目标则是不具有这种非局部自相关性的稀疏分量，从而将红外小目标检测问题转化为块图像矩阵的低秩稀疏分解问题。这类方法无需对目标尺度进行显式的假设，且算法迭代中的阈值收缩操作能够将绝大部分背景成分收缩至零，从而更好地抑制背景中的杂波干扰。然而，通过从单帧图像中采样图像块构造的红外块图像并非理想的低秩矩阵[148]，由于缺乏足够

[1]https://github.com/YimianDai/DENTIST

的相似图像块，部分稀有的背景结构具有与红外小目标相似的稀疏性，难以在低秩稀疏分离过程中将这二者分离，从而造成了不必要的虚警。此外，低秩稀疏分解方法往往需要几十次乃至上百次的迭代，实时性差。

事实上，对于红外小目标检测问题，背景的局部连续性先验和非局部自相关性先验并不完全等价，且具有一定的互补性，可以通过两者的结合来缓解各自方法的不足。例如，滤波方法或者局部对比度度量方法往往通过分析图像局部结构来区分同样破坏背景局部连续性的真实目标与背景干扰[10,125,126]。然而，这类方法通常只能抑制图像中的强边缘结构，而实际云杂波中边缘形态的复杂度远超过理想边缘的假设，导致最终的显著性图中往往仍有许多背景结构残留。与之相反的是，低秩稀疏分解方法通过阈值收缩操作能够抑制大部分形态复杂的云杂波。然而，由于强边缘的局部灰度差异较大，阈值收缩操作很难在保留真实目标的同时将其抑制，这导致了残留的强边缘成为目标图像中虚警的主要来源之一。基于上述观察，一个很自然的想法是能否将基于局部结构信息的边缘分析方法嵌入基于背景非局部自相关性的低秩稀疏分解方法中，使得结合后的模型能够更好地区分真实目标与同样稀疏的强边缘残留。

为此，针对目前低秩稀疏分解方法中，刻画目标的稀疏约束无法区分稀疏的真实目标与同样相对稀疏的背景干扰物这一不足，本章构建了一个新的红外小目标检测模型——重加权红外块张量（Reweighted Infrared Patch-Tensor, RIPT）模型，旨在同时利用图像中的局部结构先验与背景的非局部自相关先验。首先将传统矩阵形式的 IPI 模型推广为红外块张量（Infrared Patch-Tensor, IPT）模型，采用张量鲁棒低秩恢复来建模小目标与背景的分离问题，以更好地保留图像块之间的空间相关性。其次，为了抑制图像中的强边缘干扰，借助结构张量，设计了逐元素的局部结构权重用来代替原本的全局参数，使得模型在迭代过程中能够根据图像局部结构自适应地调整收缩阈值。最后，出于减少算法迭代次数的目的，以目标块张量的稀疏度变化情况为指标重新设计了算法的终止条件，并且采用稀疏性增强权重用于保证其稀疏度稳定且快速的下降。

2.2 重加权红外块张量模型

2.2.1 红外块张量模型

给定一幅红外图像 f_{F}，可以将其表示成背景图像 f_{B}、目标图像 f_{T} 和噪声图像 f_{N} 的线性组合：

$$f_{\mathrm{F}} = f_{\mathrm{B}} + f_{\mathrm{T}} + f_{\mathrm{N}}, \tag{2.1}$$

以一定步长，利用滑动窗口从左到右、从上到下依次遍历图像，并将得到的图像块堆叠成一个三维的立方体即可得到输入图像的块张量表示 $\boldsymbol{\mathcal{F}}$。相应的，式 (2.1) 可以被转换为

$$\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{T}} + \boldsymbol{\mathcal{N}}, \tag{2.2}$$

式中，$\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{N}} \in \mathbb{R}^{I \times J \times P}$ 分别表示背景块张量、目标块张量和噪声块张量。I 和 J 表示图像块的高度和宽度。

背景块张量 $\boldsymbol{\mathcal{B}}$：通常，红外图像背景被认为是缓慢过渡的，这意味着任意一个背景图像块不仅与其临近的图像块高度相关，也与非局部的图像块相关。以图 2.1 的第一列为例，尽管图像块 p_1, p_2, p_3 分布在图像的不同位置，但它们都是相似的。基于这种非局部的自相似性，IPI 模型通过拼接向量化后的图像块构建潜在的低秩矩阵，即红外块图像。事实上，块张量按照模式-3 展开后的矩阵便是块图像，因此可以将 IPI 模型看作是 IPT 模型的特例。图 2.1 展示了将块张量按照三种模式展开后各自的奇异值分布情况，从中可以看到，每个展开模式下大奇异值所占的比例都很小。因此，可以将红外图像的背景张量块 $\boldsymbol{\mathcal{B}}$ 视为低秩张量，且其展开后的矩阵也都是低秩矩阵，即：

$$\operatorname{rank}(\boldsymbol{B}_{(1)}) \leq r_1,\ \operatorname{rank}(\boldsymbol{B}_{(2)}) \leq r_2,\ \operatorname{rank}(\boldsymbol{B}_{(3)}) \leq r_3, \tag{2.3}$$

式中，rank 代表矩阵或张量的秩函数，$\boldsymbol{B}_{(1)}$、$\boldsymbol{B}_{(2)}$、$\boldsymbol{B}_{(3)}$ 分别为张量 $\boldsymbol{\mathcal{B}}$ 按照模式-1、模式-2、模式-3 展开后的矩阵，r_1, r_2 和 r_3 为反映背景图像复杂度的常数。

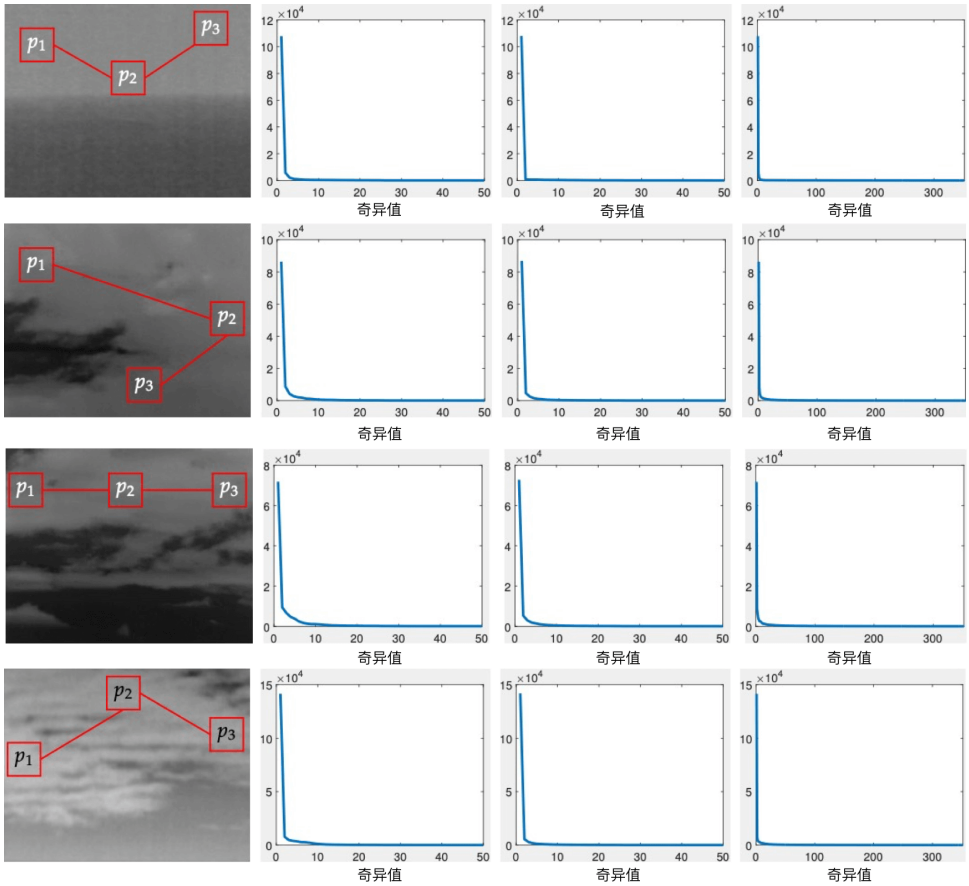


图 2.1　图像块非局部自相似性与块张量展开矩阵低秩性示意图

目标块张量 $\mathcal{T}$：由于小目标仅占据图像中的若干像素，因此可以将目标块张量视作一个稀疏张量，即

$$\|\mathcal{T}\|_0 \leq k, \tag{2.4}$$

式中，$\|\cdot\|_0$ 代表矩阵或张量的 ℓ_0 范数，定义为相应的非零元素个数。k 是一个由小目标面积和个数决定的正整数。

噪声块张量 $\mathcal{N}$：在本章中，噪声被假设为加性高斯白噪声且满足 $\|\mathcal{N}\|_{\mathrm{F}} \leq \delta_{\mathrm{N}}$，$\delta_{\mathrm{N}} > 0$ 即

$$\|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta_{\mathrm{N}}. \tag{2.5}$$

式中，$\|\cdot\|_{\mathrm{F}}$ 代表矩阵或张量的 Frobenius 范数。给定一个张量 $\mathcal{X}$，$\|\mathcal{X}\|_{\mathrm{F}} = \sqrt{\mathrm{vec}(\mathcal{X})^{\top}\mathrm{vec}(\mathcal{X})}$，vec 代表矩阵或张量的向量化。

基于上述约束条件，块张量模型的低秩稀疏分解可以被表示为：

$$\min_{\mathcal{B},\mathcal{T}} \ \mathrm{rank}(\mathcal{B}) + \|\mathcal{T}\|_0, \ \text{s.t.}\ \mathcal{B} + \mathcal{T} = \mathcal{F}. \tag{2.6}$$

然而，张量秩的计算是 NP-Hard 问题[48]。为此，Goldfarb 和 Qin 提出了鲁棒张量恢复[53]，采用凸的 Tucker-秩 $\mathrm{CTrank}(\mathcal{B})$ 和 $\|\mathcal{T}\|_1$ 来代替秩函数和 $\|\mathcal{T}\|_0$，使得松弛后的问题能够被求解。在单例模型中，张量秩的正则化项定义为所有模式展开后的核范数之和，即 $\mathrm{CTrank}(\mathcal{B}) = \sum_i \|\boldsymbol{B}_{(i)}\|_*$，$i = 1, 2, 3$，$\|\cdot\|_*$ 代表矩阵的核范数。通过松弛，带噪声的 IPT 模型可以通过求解以下凸问题实现：

$$\min_{\mathcal{B},\mathcal{T}} \sum_{i=1}^{3} \|\boldsymbol{B}_{(i)}\|_* + \lambda\|\mathcal{T}\|_1, \ \text{s.t.}\ \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta_{\mathrm{N}}. \tag{2.7}$$

式中，λ 是控制目标块张量稀疏程度的权重系数。越大的 λ 会将越多目标块张量中非目标的稀疏分量收缩为零，但也会使得真实的红外弱小目标被过度抑制。相反，较小的 λ 有利于保留弱小目标，但同时也会保留较强的云边缘。因此，采用全局恒定的权重系数 λ 对于复杂场景中的红外小目标检测并非最佳方案。

2.2.2 目标块张量的重加权系数

2.2.2.1 局部结构权重

结构张量能够较好地反映包括边缘方向在内的图像局部结构，且计算简单、快捷，在基于偏微分方程的图像处理方法中运用较为广泛[177,178]。结构张量的具体定义如下：

$$\boldsymbol{J}_{\alpha}(\nabla u_{\bar{\sigma}}) = G_{\alpha} * (\nabla \boldsymbol{u}_{\bar{\sigma}} \bar{\otimes} \nabla \boldsymbol{u}_{\bar{\sigma}}) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \tag{2.8}$$

式中，$\boldsymbol{u}_{\bar{\sigma}}$ 表示图像 $\boldsymbol{u}$ 高斯平滑后的结果，$\bar{\sigma} > 0$ 是控制平滑力度的高斯核标准差，∇ 和 $\bar{\otimes}$ 分别代表梯度算子和张量积，G_{α} 是用于计算方向均值、标准差为 α 的高斯核，$*$ 代表卷积。$\boldsymbol{J}_{\alpha}$ 是

一个大小为 2×2 的对称半正定矩阵，$J_{11}, J_{12}, J_{21}, J_{22}$ 为其中的四个元素。$\boldsymbol{J}_\alpha$ 的特征值 λ_1 和 λ_2 的计算方式为

$$\lambda_1, \lambda_2 = (J_{11}+J_{22}) \pm \sqrt{(J_{22}-J_{11})^2 + 4J_{12}^2}. \tag{2.9}$$

某种程度上，λ_1 和 λ_2 可以用来反映图像的局部结构。例如，在平坦区域，$\lambda_1 \approx \lambda_2 \approx 0$；在边缘区域，$\lambda_1 \gg \lambda_2 \approx 0$；在角点区域，$\lambda_1 \geq \lambda_2 \gg 0$。因此，可以用 $\lambda_1 - \lambda_2$ 作为边缘显著度量，其数值越大，表示对应像素越有可能属于边缘。

对于输入图像 f_{F} 中的每个像素计算式 (2.9)，可以得到两个与图像相同大小的矩阵 $\boldsymbol{L}_1$ 和 $\boldsymbol{L}_2$。然后采用同样的块张量构造方法，可以将其转换为相应的张量 $\boldsymbol{\mathcal{L}}_1$ 和 $\boldsymbol{\mathcal{L}}_2$。由此，局部结构权重的块张量可以定义为：

$$\boldsymbol{\mathcal{W}}_{\mathrm{LS}} = \exp\left(h \cdot \frac{(\boldsymbol{\mathcal{L}}_1 - \boldsymbol{\mathcal{L}}_2) - d_{\min}}{d_{\max} - d_{\min}}\right), \tag{2.10}$$

式中，exp 代表指数运算，h 是一个权重伸展因子，$d_{\max}$ 和 $d_{\min}$ 分别是 $\boldsymbol{\mathcal{L}}_1 - \boldsymbol{\mathcal{L}}_2$ 的最大值和最小值。图 2.2 展示了部分图像的边缘显著性图，可以看到基于结构张量的局部结构权重能够较好地辨别出边缘。

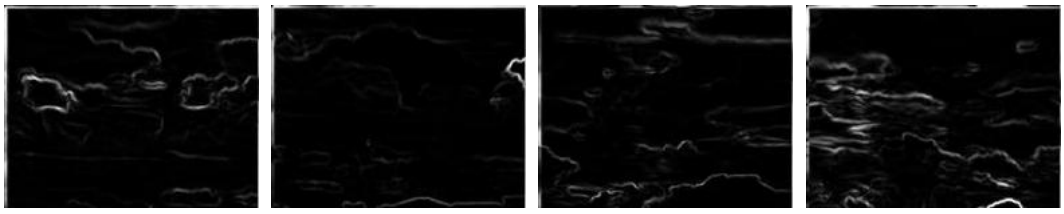


图 2.2　基于结构张量的边缘显著性图

通过将 $\boldsymbol{\mathcal{W}}_{\mathrm{LS}}$ 嵌入式 (2.7)，可以得到一个加权的 IPT（Weighted IPT, WIPT）模型：

$$\min_{\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{T}}} \sum_{i=1}^{3} \|\boldsymbol{B}_{(i)}\|_* + \lambda \|\boldsymbol{\mathcal{W}}_{\mathrm{LS}} \odot \boldsymbol{\mathcal{T}}\|_1, \ \text{s.t.}\ \|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{B}} - \boldsymbol{\mathcal{T}}\|_{\mathrm{F}} \leq \delta_{\mathrm{N}}, \tag{2.11}$$

式中，$\odot$ 表示逐元素点积。对于图像中的强边缘结构，其局部结构权重高于真实的小目标，相应的收缩阈值也更大。由此，WIPT 模型可以在保留小目标的同时实现对同样稀疏的强边缘结构的抑制。

2.2.2.2　稀疏性增强权重

相较于滤波方法和显著性检测方法，基于低秩稀疏分解的红外小目标检测方法计算时间较长，这是由于鲁棒主成分分析（Robust Principal Component Analysis, RPCA）类的算法往往选取重构误差作为迭代终止条件，只有重构误差小于某个很小的值时算法才会停止。事实上，在算法收敛之前，目标块张量中的非零元素个数通常已经停止改变，且真实目标所在的像素往往数

值最大，后续的算法迭代对于目标块张量的改动极其微小。基于该观察，考虑到基于低秩稀疏分解的红外小目标检测方法的最终目的是从分离出的目标图像中检测小目标，而非保证低秩矩阵恢复的精度，为了减少该类方法的计算时间，本章选用目标块张量的稀疏性代替重建误差作为算法终止条件，即 $\|\mathcal{T}^{k+1}\|_0 = \|\mathcal{T}^{k}\|_0$，一旦其 ℓ_0 范数不再变化就停止算法的迭代，k 为算法的迭代轮次。在新的算法终止条件之下，目标块张量的稀疏性对于减少计算时间至关重要。理想的情况是随着算法的不断迭代，目标块张量中的非零元素个数将逐渐减小，最终仅保留真实的目标元素。然而，在式 (2.7) 和式 (2.11) 中，刻画稀疏项的约束条件为 ℓ_1 范数，其并不具有良好的稀疏诱导性质。如图 2.7 所示，在 WIPT 模型的收敛过程中，随着算法迭代次数的增加，其目标块张量的 ℓ_0 范数反而存在上升的现象，无法大幅减少算法的迭代次数。

为了解决该问题，本小节采用重加权的 ℓ_1 范数代替原有的 ℓ_1 范数，以增强目标块张量的稀疏性。在迭代过程中，较大的权重有利于减少其非零元素的个数，而较小的权重会保留非零元素。此外，考虑到红外小目标的亮度通常高于周围背景区域这一先验知识，给定块张量中的位置 (i, j, p)，本章的稀疏性增强（Sparsity Enhancing, SE）权重 $\mathcal{W}_{\mathrm{SE}}^{k+1}(i, j, p)$ 具体定义如下：

$$\mathcal{W}_{\mathrm{SE}}^{k+1}(i, j, p) = \begin{cases} \frac{1}{\mathcal{T}^{k}(i,j,p)+\epsilon}, & \text{if } \mathcal{T}^{k}(i, j, p) > 0; \\ \infty, & \text{if } \mathcal{T}^{k}(i, j, p) \leq 0. \end{cases} \tag{2.12}$$

最后，通过结合局部结构权重 $\mathcal{W}_{\mathrm{LS}}$ 和稀疏性增强权重 $\mathcal{W}_{\mathrm{SE}}$，可以得到最终的重加权权重 $\mathcal{W}$：

$$\mathcal{W}^{k} = \mathcal{W}_{\mathrm{LS}} \odot \mathcal{W}_{\mathrm{SE}}^{k}. \tag{2.13}$$

借由 $\mathcal{W}$，本章设计的重加权红外块张量（Reweighted Infrared Patch-Tensor, RIPT）模型具体定义如下：

$$\min_{\mathcal{B},\mathcal{T}} \sum_{i=1}^{3} \|\boldsymbol{B}_{(i)}\|_* + \lambda \|\mathcal{W} \odot \mathcal{T}\|_1, \text{ s.t. } \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta_{\mathrm{N}}. \tag{2.14}$$

2.2.3 模型求解

本章使用交替方向乘子法（Alternating Direction Method of Multipliers, ADMM）求解重加权的鲁棒张量恢复问题，式 (2.14) 的增广拉格朗日函数定义为：

$$\mathcal{L} = \sum_{i=1}^{3} \|\boldsymbol{B}_{i,(i)}\|_* + \lambda \|\mathcal{W} \odot \mathcal{T}\|_1 + \sum_{i=1}^{N} \frac{1}{2\mu} \|\mathcal{B}_i + \mathcal{T} - \mathcal{F}\|_{\mathrm{F}}^2 - \langle \mathcal{Y}_i, \mathcal{B}_i + \mathcal{T} - \mathcal{F} \rangle, \tag{2.15}$$

式中，$\mathcal{Y}_i \in \mathbb{R}^{I \times J \times P}$，$i = 1, 2, 3$ 为拉格朗日乘子，μ 是一个正的保真项惩罚因子。ADMM 将 $\mathcal{L}$ 的最小化问题分解为 $\mathcal{B}_i$ 和 $\mathcal{T}$ 两个子问题进行求解：

$$\mathcal{B}_i^{k+1} = \arg\min_{\mathcal{B}_i} \left\|\boldsymbol{B}_{i,(i)}\right\|_* + \frac{1}{2\mu} \left\|\mathcal{B}_i - (\mathcal{F} + \mu \mathcal{Y}_i^{k} - \mathcal{T}^{k})\right\|_{\mathrm{F}}^2 \tag{2.16}$$

$$\mathcal{T}^{k+1} = \arg\min_{\mathcal{T}} \lambda \left\|\mathcal{W}^{k} \odot \mathcal{T}\right\|_1 + \sum_{i=1}^{N} \frac{1}{2\mu} \left\|\mathcal{T} - \left(\mathcal{F} + \mu \mathcal{Y}_i^{k} - \mathcal{B}_i^{k+1}\right)\right\|_{\mathrm{F}}^2 \tag{2.17}$$

$$\boldsymbol{\mathcal{Y}}_i^{k+1} = \boldsymbol{\mathcal{Y}}_i^k + \frac{1}{\mu^k}\left(\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{B}}_i^{k+1} - \boldsymbol{\mathcal{T}}^{k+1}\right),\ i = 1, \cdots, N. \tag{2.18}$$

式 (2.16) 和式 (2.17) 存在闭式解：

$$\boldsymbol{\mathcal{B}}_i^{k+1} = \text{fold}_i\left(\mathcal{D}_\mu\left[\left(\boldsymbol{\mathcal{F}} + \mu\boldsymbol{\mathcal{Y}}_i^k - \boldsymbol{\mathcal{E}}^k\right)_{(i)}\right]\right) \tag{2.19}$$

$$\boldsymbol{\mathcal{T}}^{k+1} = \mathcal{S}_{\frac{\mu\lambda}{N}\boldsymbol{\mathcal{W}}^k}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\mathcal{F}} + \mu\boldsymbol{\mathcal{Y}}_i^k - \boldsymbol{\mathcal{B}}_i^{k+1}\right)\right] \tag{2.20}$$

式中，$\mathcal{S}$ 为逐元素的阈值收缩算子，对于给定收缩阈值 μ，$\mathcal{S}_\mu(x) = \text{sign}(x)\max(|x| - \mu, 0)$，sign 为符号函数。$\mathcal{D}$ 为矩阵奇异值的阈值收缩算子，$\mathcal{D}_\mu(\boldsymbol{X}) = \boldsymbol{U}\text{diag}(\hat{\sigma})\boldsymbol{V}^\top$，其中 $\boldsymbol{X} = \boldsymbol{U}\text{diag}(\sigma)\boldsymbol{V}^\top$ 为矩阵 $\boldsymbol{X}$ 的 SVD 分解，$\hat{\sigma} = \mathcal{S}_\mu(\sigma)$。$\text{fold}_i$ 是将按照模式-i 展开的矩阵重新恢复成原张量的操作，即 $\boldsymbol{\mathcal{X}} = \text{fold}_i(\boldsymbol{X}_{(i)})$。RIPT 模型具体的优化步骤如算法 1 所示。

算法 1: RIPT 模型优化算法

输入: 红外块张量 $\boldsymbol{\mathcal{F}}$，局部结构权重张量 $\boldsymbol{\mathcal{W}}_{\text{LS}}$，超参数 λ

输出: 背景块张量 $\frac{1}{3}\left(\sum_{i=1}^{3}\boldsymbol{\mathcal{B}}_i^k\right)$，目标块张量 $\boldsymbol{\mathcal{T}}^k$

初始化: $\boldsymbol{\mathcal{T}}^0 = \boldsymbol{0}$; $\boldsymbol{\mathcal{B}}_i^0 = \boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{Y}}_i^0 = \boldsymbol{0}, i = 1, 2, 3$; $\boldsymbol{\mathcal{W}}_{\text{SE}}^0 = \boldsymbol{1}, \boldsymbol{\mathcal{W}}^0 = \boldsymbol{\mathcal{W}}_{\text{LS}} \odot \boldsymbol{\mathcal{W}}_{\text{SE}}^0$;
$\mu = 5 \cdot \text{std}(\text{vec}(\boldsymbol{\mathcal{F}}))$, $k = 0$

while $\left\|\boldsymbol{\mathcal{T}}^{k+1}\right\|_0 \neq \left\|\boldsymbol{\mathcal{T}}^k\right\|_0$ **do**

▷ 固定其他项，更新 $\boldsymbol{\mathcal{B}}_i$

for $i = 1$ to 3 **do**

$\boldsymbol{\mathcal{B}}_i^{k+1} := \text{fold}_i\left(\mathcal{D}_\mu\left[\left(\boldsymbol{\mathcal{F}} + \mu\boldsymbol{\mathcal{Y}}_i^k - \boldsymbol{\mathcal{E}}^k\right)_{(i)}\right]\right)$;

end

▷ 固定其他项，更新 $\boldsymbol{\mathcal{T}}$

$\boldsymbol{\mathcal{T}}^{k+1} := \mathcal{S}_{\frac{\mu\lambda}{N}\boldsymbol{\mathcal{W}}^k}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\mathcal{F}} + \mu\boldsymbol{\mathcal{Y}}_i^k - \boldsymbol{\mathcal{B}}_i^{k+1}\right)\right]$;

▷ 固定其他项，更新 $\boldsymbol{\mathcal{Y}}_i$

for $i = 1$ to 3 **do**

$\boldsymbol{\mathcal{Y}}_i^{k+1} := \boldsymbol{\mathcal{Y}}_i^k + \frac{1}{\mu^k}\left(\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{B}}_i^{k+1} - \boldsymbol{\mathcal{T}}^{k+1}\right)$;

end

▷ 按照式 (2.12) 和式 (2.13) 更新 $\boldsymbol{\mathcal{W}}^{k+1}$

▷ 更新 μ：$\mu^{k+1} := \mu^k/\rho$

▷ 更新 k：$k = k + 1$

end

图 2.3 展示了基于 RIPT 模型的红外小目标检测方法总体的示意图，其具体步骤如下：

步骤 1：给定一幅红外图像 f_{F}，按照式 (2.10) 计算局部结构权重构成的特征图 f_{LS}；

步骤 2：将 f_{F} 和 f_{LS} 转换为相应的块张量 $\boldsymbol{\mathcal{F}}$ 和 $\boldsymbol{\mathcal{W}}_{\text{LS}}$；

步骤 3：根据算法 1 将 $\boldsymbol{\mathcal{F}}$ 分解为背景块张量 $\boldsymbol{\mathcal{B}}$ 和目标块张量 $\boldsymbol{\mathcal{T}}$；

步骤 4：采用统一均值估计器[179]（Uniform Average of Estimators, UAE）将分离后的 $\mathcal{B}$ 和 $\mathcal{T}$ 重投影为背景图像 f_{B} 和目标图像 f_{T}；

步骤 5：计算目标图像分割的自适应阈值 t_{up}：

$$t_{\mathrm{up}} = \max(v_{\min}, \bar{f}_{\mathrm{T}} + k_{\mathrm{T}}\sigma_{\mathrm{T}}), \tag{2.21}$$

式中，$\bar{f}_{\mathrm{T}}$ 和 σ_{T} 分别为目标图像 f_{T} 的均值和标准差，k 和 $v_{\min}$ 为由实验决定的经验值。

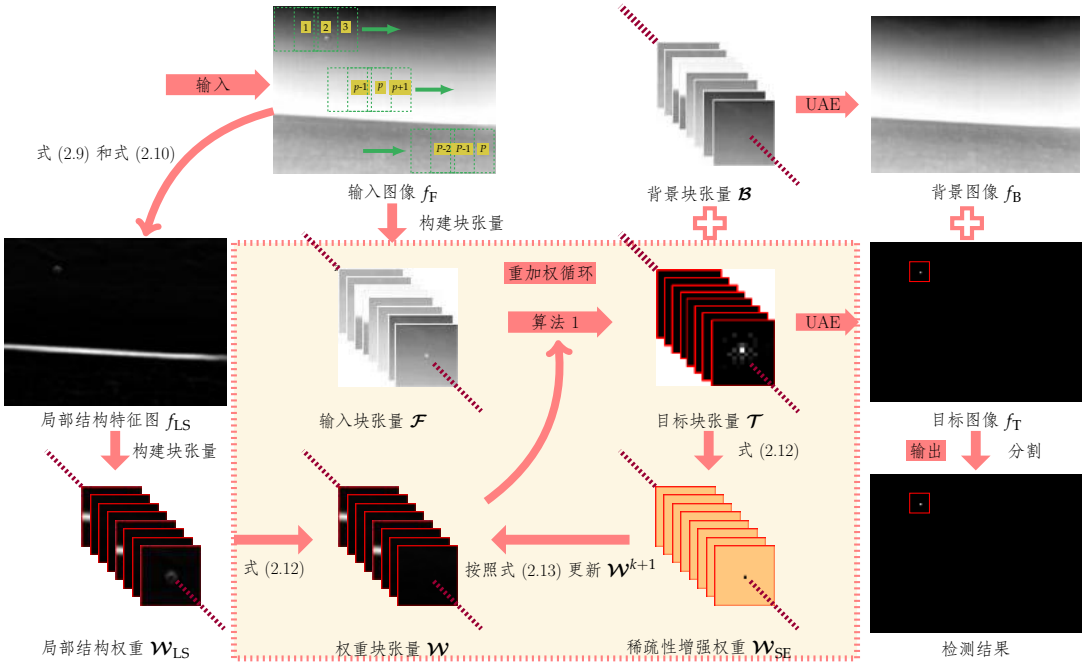


图 2.3　基于 RIPT 模型的红外小目标检测方法流程图

2.3　实验结果与分析

2.3.1　实验设置

本节将在真实场景的红外序列图像上验证所提出方法的有效性，图 2.4 展示实验所用序列中的代表性图像。从中可以看出，实验序列中既包含目标显著、背景平坦的简单场景，也有目标弱小、背景具有强烈云起伏的复杂场景。考虑到当前红外小目标检测的最大难点在于如何检测强杂波背景下的弱小目标，因此本节将重点关注以图 2.4 (a) - (d) 和图 2.4 (l) 为代表的高难度图像序列上。表 2.1 详细描述了这五个序列的特点。

为了充分评估所提出的 RIPT 模型，将其与其他 10 种方法进行横向对比，包括最大中值（Max-Median）滤波器[9]、Top-Hat 滤波器[8]、二维最小均方差（Two-Dimensional Least-Mean-Square, TDLMS）滤波器[180]，傅里叶变换相位谱（Phase Spectrum of Fourier Transform, PFT）方法[82]，多尺度块对比度度量（Multi-scale Patch-based Contrast Measure, MPCM）方法[10]、加权的

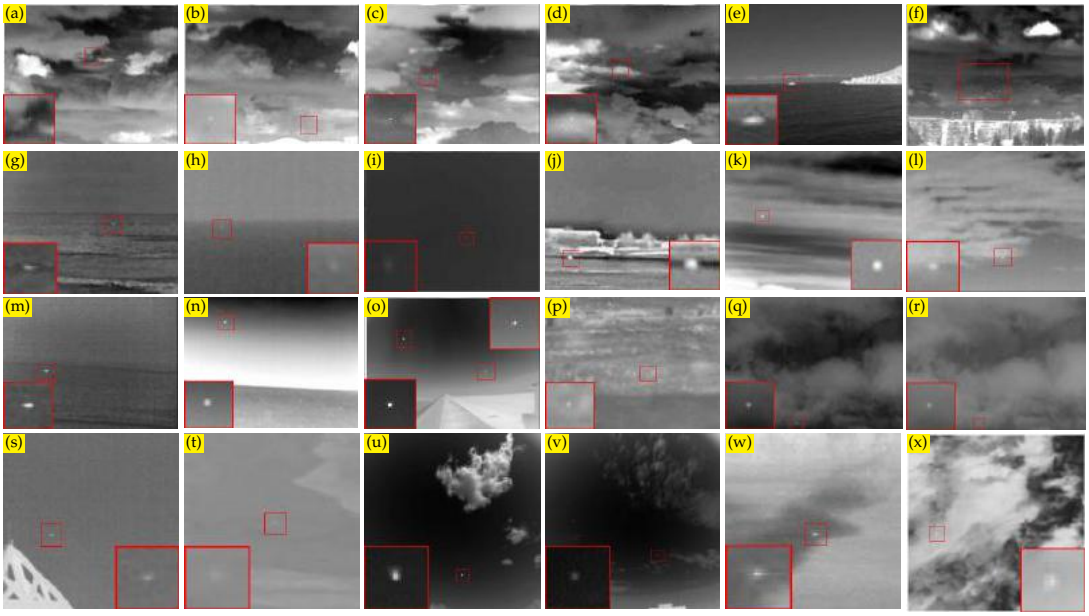


图 2.4　实验所用红外小目标序列的代表性图像展示

表 2.1　真实场景中红外序列的目标与背景特点

	帧数	图像大小	目标形状	目标特点	背景特点
序列 1-4	400	255 × 320	高斯	✧ 弱小、低对比度 ✧ 沿着云边缘运动	✧ 强起伏的云背景 ✧ 没有噪声
序列 5	30	200 × 256	矩形	✧ 目标较小 ✧ 亮度、对比度、尺寸变化大	✧ 存在强云杂波和部分噪声

局部差异度量（Weighted Local Difference Measure, WLDM）方法[142]、红外块图像（Infrared Patch-Image, IPI）模型[12]、块鲁棒主成分分析（Patch Robust Principal Component Analysis, PRPCA）方法[181]、加权的 IPI（Weighted IPI, WIPI）模型[157]、基于奇异值部分和最小化的非负 IPI 模型[148]（Non-negative IPI Model Based on Partial Sum Minimization of Singular Values, NIPPS）。表 2.2 中展示包括本章提出的 IPT 基础模型和 RIPT 模型在内一共 12 种方法的详细参数设置。

为了全面地评估本章方法，采用四个指标来衡量各类方法的背景抑制性能，包括局部信噪比增益（Local Signal to Noise Ratio Gain, LSNRG）、背景抑制因子（Background Suppression Factor, BSF）、信噪比增益（Signal to Clutter Ratio Gain, SCRG）、接收机工作特性（Receiver Operating Characteristic, ROC）曲线。其中，LSNRG 衡量局部信噪比（Local Signal to Noise Ratio, LSNR）的增益，定义为：

$$\mathrm{LSNRG} = \frac{\mathrm{LSNR_{out}}}{\mathrm{LSNR_{in}}}, \tag{2.22}$$

式中，$\mathrm{LSNR_{in}}$ 和 $\mathrm{LSNR_{out}}$ 分别为背景抑制前后的 LSNR 值。LSNR 的定义是 $\mathrm{LSNR} = P_{\mathrm{T}}/P_{\mathrm{B}}$，其

表 2.2　12 种方法的详细参数设置

序号	方法	参数设置
1	Max-Median	支撑区域大小：5×5
2	Top-Hat	结构元形状：正方形；结构元大小：3×3
3	PFT	圆盘半径：3
4	MPCM	$N=1,3,\ldots,9$
5	WLDM	$L=4, m=2, n=2$
6	TDLMS	支撑区域大小：5×5，步长：$\mu=5\times10^{-8}$
7	IPI	块大小：50×50，滑动步长：10，$\lambda=\frac{L}{\sqrt{\min(IJ,P)}}$, $L=3$, $\varepsilon=10^{-7}$
8	PRPCA	块大小：50×50，滑动步长：10，$\lambda=\frac{L}{\sqrt{\min(IJ,P)}}$, $L=3$, $\varepsilon=10^{-7}$
9	WIPI	块大小：51×51，滑动步长：10，平滑参数 $h=15$, $\varepsilon=10^{-7}$
10	NIPPS	块大小：50×50，滑动步长：10，$\lambda=\frac{L}{\sqrt{\min(IJ,P)}}$, $L=2$, $r=5\times10^{-3}$
11	IPT	块大小：50×50，滑动步长：10，$\lambda=\frac{L}{\sqrt{\max(I,J,P)}}$, $L=3$
12	RIPT	块大小：50×50，滑动步长：10，$\lambda=\frac{L}{\sqrt{\min(I,J,P)}}$, $L=1$, $h=10$, $\epsilon=0.01$, $\varepsilon=10^{-7}$

中 P_{T} 和 P_{B} 分别为目标和其邻域的最大灰度值。BSF 的定义为背景抑制前后目标临近区域的标准差 σ_{in} 和 σ_{out} 之比，

$$\mathrm{BSF}=\frac{\sigma_{\mathrm{in}}}{\sigma_{\mathrm{out}}}, \tag{2.23}$$

SCRG 为背景抑制前后的信噪比（Signal to Clutter Ratio, SCR）之比

$$\mathrm{SCRG}=\frac{\mathrm{SCR}_{\mathrm{out}}}{\mathrm{SCR}_{\mathrm{in}}}, \tag{2.24}$$

式中，$\mathrm{SCR}=|\mu_{\mathrm{t}}-\mu_{\mathrm{b}}|/\sigma_{\mathrm{b}}$ 反映了检测红外小目标的困难程度，μ_{t} 是平均目标灰度，μ_{b} 和 σ_{b} 是临近区域的平均灰度和标准差。上述三种评价指标均在一个如图 2.5 所示以小目标为中心的局部区域内计算完成，图中目标大小为 $a\times b$，d_{N} 是邻域宽度。

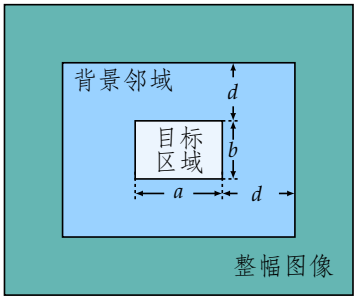


图 2.5　目标与背景邻域示意图

检测率 P_{d} 和虚警率 F_{a} 的定义如下：

$$P_{\mathrm{d}}=\frac{\text{检测到的真实目标数}}{\text{真实目标数}}, \tag{2.25}$$

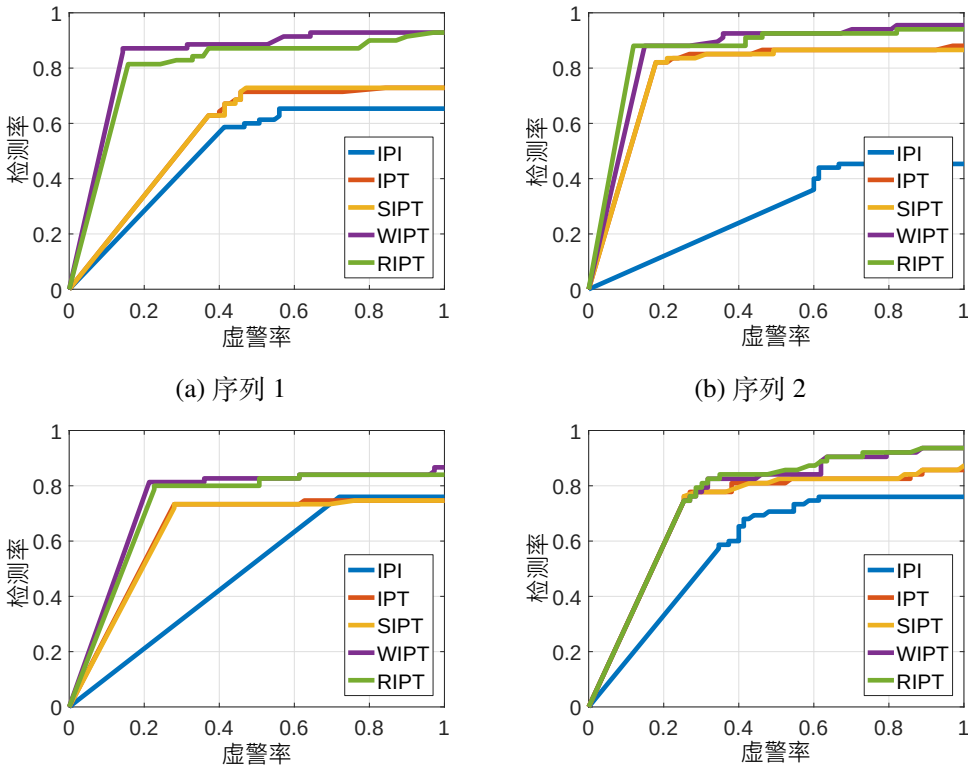
$$F_{\mathrm{a}}=\frac{\text{虚警目标数}}{\text{图像数}}. \tag{2.26}$$

2.3.2 模型框架的有效性验证

本小节首先研究了在所提出的 RIPT 框架中，构造块张量、局部结构权重、稀疏性增强权重这三者各自对于 RIPT 模型性能的影响，然后验证了 RIPT 模型对变化场景和噪声干扰的鲁棒性，并且给出了算法的计算复杂度和计算时间比较。

2.3.2.1 模型各组件对检测性能的影响

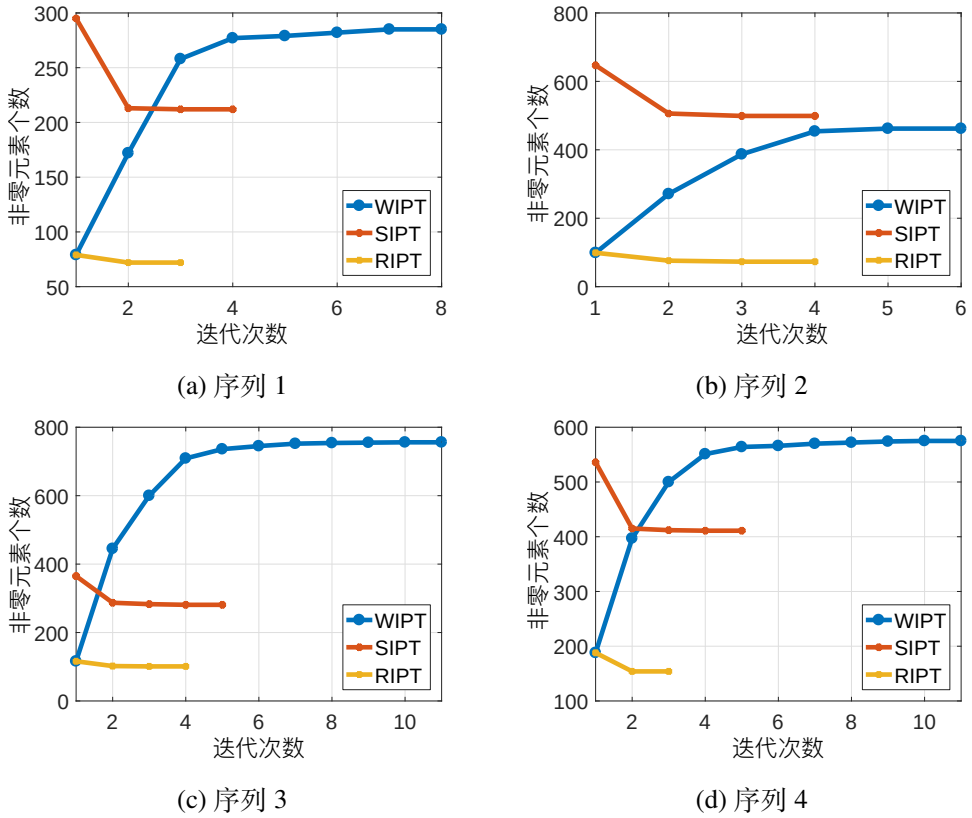
为了了解 RIPT 模型中各组件对最终检测性能的影响，图 2.6 展示了 IPI 模型、IPT 模型、使用稀疏性增强权重的 IPT 模型（Sparsity Enhanced IPT, SIPT）、WIPT 模型、RIPT 模型在四个红外序列上的 ROC 曲线。从中可以看出，1）四种基于块张量的模型效果都优于 IPI 模型，这表明张量的其他两种展开模式有助于检测；2）WIPT 模型的效果好于 IPT 模型，这表明在模型中引入局部结构权重可以提高检测性能；3）WIPT 模型与 RIPT 模型效果相当，这表明稀疏度增强权重并不会影响最终的检测效果。但是如图 2.7 所示，稀疏度增强权重可以显著减少新终止条件下的算法迭代次数。这些结果表明，将代表不同领域知识的各个模型组件组合在一起，可以使得红外块张量模型取得更好的性能。



(a) 序列 1　(b) 序列 2

(c) 序列 3　(d) 序列 4

图 2.6　RIPT 模型各组件对最终检测性能的影响分析



(a) 序列 1　(b) 序列 2

(c) 序列 3　(d) 序列 4

图 2.7　稀疏性增强权重对于目标图像稀疏性的影响分析

2.3.2.2　对场景变化和噪声干扰的鲁棒性

图 2.8 展示了图 2.4 中的红外图像经由 RIPT 模型分离出的目标图像。从中可以观察到，目标图像中只剩下真实的目标，而背景杂波已经被完全地抑制了。由于图 2.4 中包含了许多不同的场景，因此可以看到所提出方法对于不同场景都比较鲁棒。

为了检验模型对噪声的鲁棒性，图 2.9 展示了不同强度的高斯白噪声干扰下 RIPT 模型分离出的目标图像。其中，第一行为噪声标准差为 10 时的受污染图像，第二行为 RIPT 模型分离出的目标图像，可以看到，RIPT 模型很好地增强了目标并抑制了杂波和噪声。当噪声标准差增加到 20 时（第三行），RIPT 模型能够检测出大部分目标，但是会漏检第 3、4 列中的小目标。需要注意的是，此时在受噪声污染的图像中，由于对比度过小，其目标已经被噪声完全覆盖，人眼也无法区分。由此可见，模型受噪声的影响不仅取决于噪声本身的强度，还取决于目标原始的对比度。只要受污染的目标能够保持一定的对比度，RIPT 模型仍然将其能够检测出。

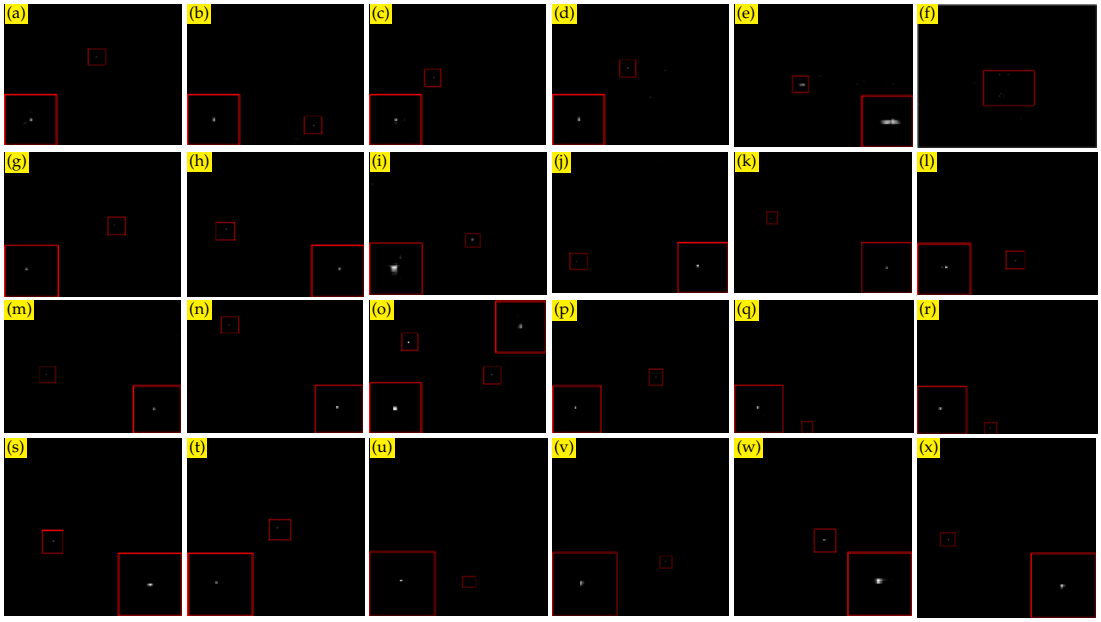


图 2.8　RIPT 模型分离出的目标图像

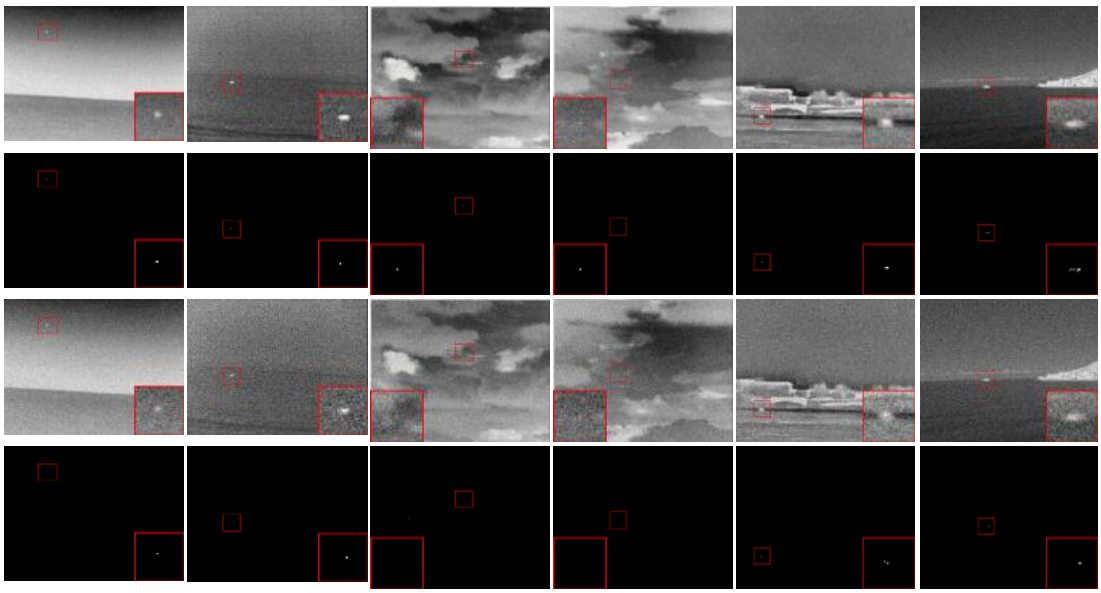


图 2.9　噪声干扰下 RIPT 模型分离出的目标图像

2.3.2.3　算法复杂度和计算时间

表 2.3 展示了一共十种方法的计算复杂度和计算时间，其中 $M \times N$ 为图像大小，$m \times n$ 为块图像或者块张量按照模式-3 展开后矩阵的大小。TDLMS、PFT、MPCM、WLDM 为典型的滤波方法和局部对比度度量方法，IPI、WIPI、NIPPS、IPT、WIPT、RIPT 等则是基于低秩稀疏分解

的方法。从中可以看出，由于需要通过反复多次迭代才能实现目标与背景的分离，在计算复杂度和计算时间上，基于低秩稀疏分解的方法普遍都大于滤波方法和局部对比度度量方法，这是该类方法的主要缺点之一。然而，就对背景杂波的抑制能力而言，基于低秩稀疏分解的方法则普遍好于其他两类方法。因此如何在保持背景抑制性能的同时，减少模型求解的迭代时间对于该类方法的进一步发展非常重要。从表 2.3 中可以看到，在所有基于低秩稀疏分解的方法中，本章构建的 RIPT 模型计算时间最少。特别是相比于 WIPT 模型，RIPT 能够将计算时间缩短 50% 以上，这表明在新的终止条件下，稀疏性促进权重能够有效地减少模型求解的迭代次数。

表 2.3　多种红外小目标检测方法的算法复杂度和计算时间比较

	TDLMS	PFT	MPCM	WLDM	IPI	WIPI	NIPPS	IPT	WIPT	RIPT
复杂度	$O(L^2MN)$	$O(MN\log MN)$	$O(L^3MN)$	$O(L^3MN)$	$O(mn^2)$	$O(mn^2)$	$O(mn^2)$	$O(mn^2)$	$O(mn^2)$	$O(mn^2)$
时间/秒	0.162	0.025	0.083	6.059	16.998	52.995	15.515	8.598	6.932	3.169

2.3.3 模型对比实验与分析

表 2.4 展示了 RIPT 模型与其他 11 种方法的定量评价指标比较。从中可以看出：1）IPI、PRPCA、WIPI、NIPPS、IPT、RIPT 这些基于低秩稀疏分解的方法普遍好于其余的滤波方法和局部对比度度量方法，且部分方法中有出现无穷大 **Inf**。事实上，对于低秩稀疏方法而言，出现 **Inf** 其实非常常见，仅仅表明阈值收缩算子将小目标临近的背景区域的数值均收缩为零。LSNRG、BSF、SCRG 这些用于评估背景抑制效果的定量评价指标主要针对早期滤波方法设计，如表 2.4 的前六种方法所示，这些滤波方法几乎无法将背景元素彻底抑制为零，一般不会有出现 **Inf** 的现象。2）在所有实验场景和指标中，RIPT 模型取得了最好的效果，这表明其能够在保存弱小小目标的同时，对背景杂波进行较好的抑制。

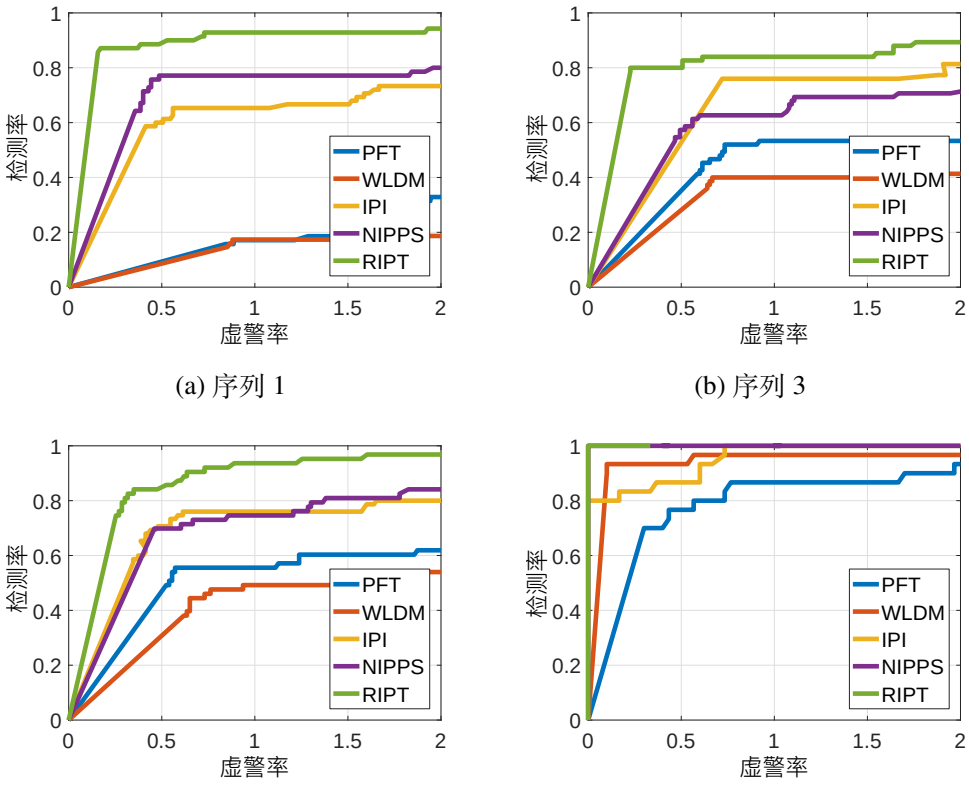
图 2.10 展示了 PFT、WLDM、IPI、NIPPS、RIPT 等方法在四个红外序列的代表性图像上的 ROC 曲线比较。从中可以看出在所有对比方法中，RIPT 模型能够在虚警率更低的时候取得更高的检测率，好于 IPI 和 NIPPS。这表明通过构造局部结构权重，RIPT 模型能够更好地抑制强边缘残留，使得虚警率更低。

最后，图 2.11 展示了 12 种方法在四个红外序列代表性图像上的可视化比较。从中可以看出，就背景抑制的视觉效果而言，TDLMS、Max-Median、Top-Hat、PFT、MPCM、WLDM 这些基于滤波或局部对比度度量的方法不如其余基于低秩稀疏分解的方法，残留的背景结构不管是亮度还是数量都远远多于后者。在所有基于低秩稀疏分解的方法，RIPT 模型则能够将背景抑制地更加干净。例如在图 2.11(b) 中，原图像右侧的云边缘在 IPI、PRPCA、IPT 这些方法所产生的目标图像中仍有残留，而 RIPT 模型则能较好对其进行抑制。

表 2.4　序列 1 - 4 代表性图像上的定量评价指标比较

方法	序列 1 第 65 帧			序列 2 第 52 帧			序列 3 第 53 帧			序列 4 第 56 帧		
	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF
Max-Median	5.49	10.69	12.10	1.87	5.46	7.37	2.96	6.21	11.27	7.54	9.81	16.66
Top-Hat	3.47	13.47	12.34	2.10	12.55	8.08	3.10	9.48	11.24	4.04	22.13	21.06
PFT	4.83	53.01	7.43	1.37	10.96	3.22	0.68	7.48	3.38	9.16	113.25	18.77
MPCM	1.48	7.69	3.46	1.62	11.84	15.98	0.38	1.91	2.15	1.88	14.17	4.68
WLDM	0.87	2.00	1.99	2.22	9.95	12.23	1.94	8.19	3.95	3.11	12.11	3.56
TDLMS	1.36	3.44	3.53	1.76	4.27	3.38	2.61	4.39	4.49	1.99	4.77	4.50
IPI	220.38	5215.82	19256.20	10.72	104.34	172.90	**Inf**	**Inf**	**Inf**	**Inf**	2788.19	4939.03
PRPCA	5.17	382.58	20179.12	1.30	26.88	2628.42	1.68	31.85	1982.80	2.83	267.31	20966.41
CWRPCA	2.67	36.77	602.45	4.62	40.59	58.23	7.69	98.57	201.89	52.92	441.65	2065.42
NIPPS	15.95	315.08	670.65	3.89	66.99	81.05	20.59	343.06	735.19	87.92	2280.13	3103.00
IPT	9.80	2096.70	87797.82	2.14	332.56	17488.27	3.22	**Inf**	**Inf**	2.86	**Inf**	**Inf**
RIPT	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**

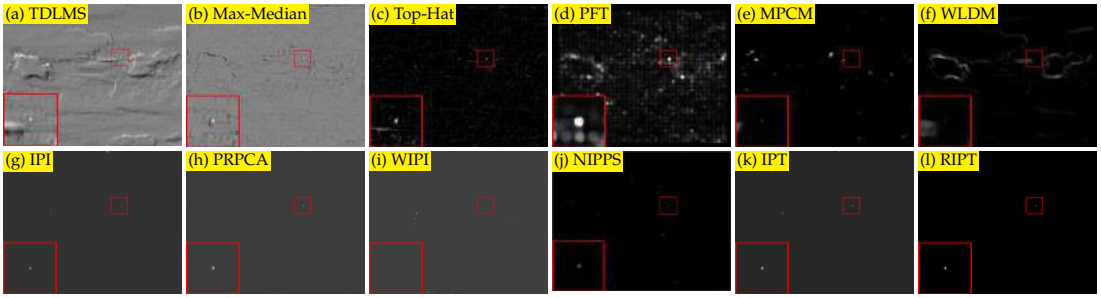
* 不同于滤波方法，在低秩稀疏分解方法中 **Inf** 非常常见，仅仅意味着目标临近区域被阈值收缩至 0 了。



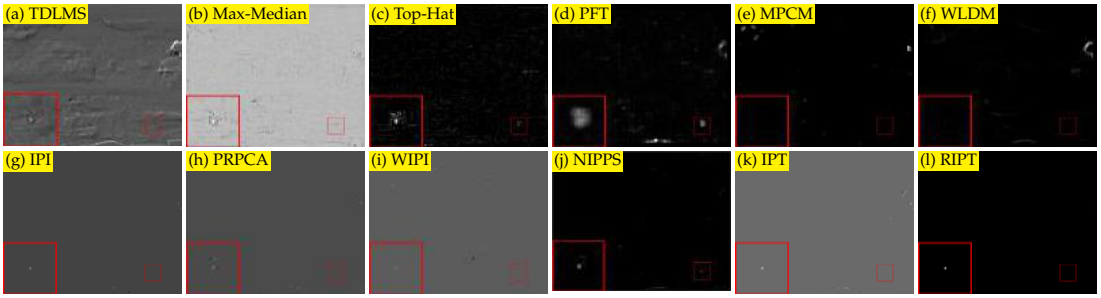
(a) 序列 1　(b) 序列 3

(c) 序列 4　(d) 序列 5

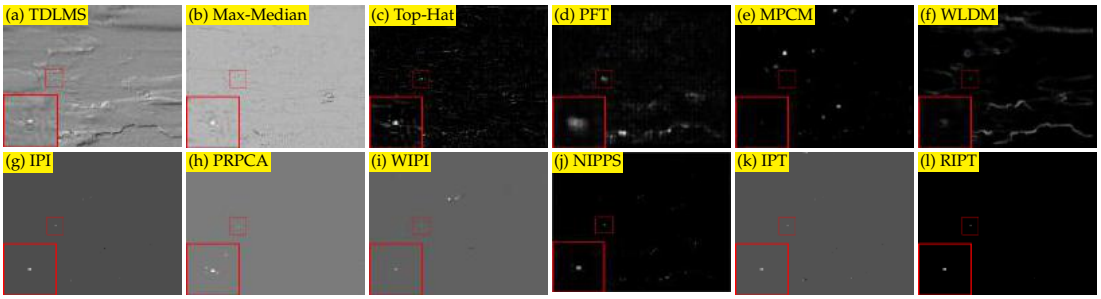
图 2.10　四个红外序列代表性图像的 ROC 曲线



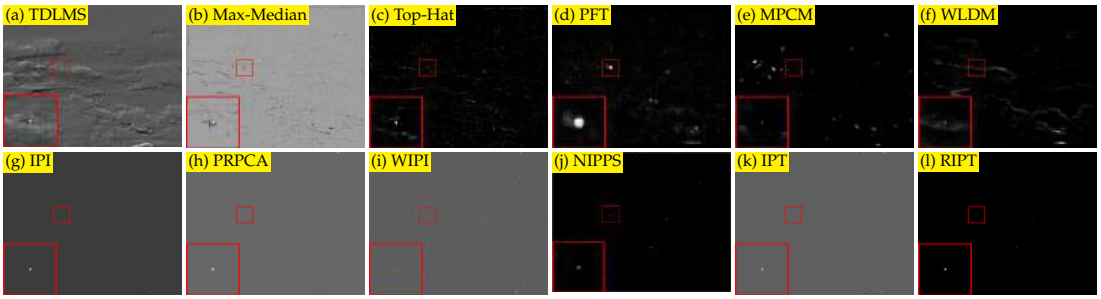
(a) 序列 1 第 65 帧



(b) 序列 2 第 52 帧



(c) 序列 3 第 53 帧



(d) 序列 4 第 56 帧

图 2.11　12 种方法在四个红外序列代表性图像上的可视化比较

2.4 本章小结

针对传统低秩稀疏分解方法无法有效抑制图像中的强边缘残留这一问题，本章提出了一个重加权红外块张量模型用于红外小目标的快速检测。为了更好地保留图像块之间的空间相关性，首先通过堆叠滑动窗口采样的图像块构造红外块张量，将小目标与背景的分离问题建模为张量鲁棒低秩恢复问题。其次，借助结构张量，设计了逐元素的局部结构权重用来代替原本的全局参数，使得模型在迭代过程中能够根据图像边缘的显著性大小自适应地调整收缩阈值，从而更好地抑制图像中的强边缘残留。最后，RIPT 模型构建了新的终止条件和稀疏性增强权重以大幅减少算法所需的迭代次数。大量实验结果表明，相较于其他低秩稀疏分解方法，该模型能够在保存红外弱小目标的同时，更好地抑制复杂的云杂波干扰，且所需的计算时间更少。

第三章　基于双向非对称注意力调制网络的红外小目标检测

前一章针对红外小目标检测问题，在低秩稀疏分解的框架下，通过采用更符合真实场景的稀疏正则项去刻画目标，以期获得更好地检测结果。然而，模型驱动方法在复杂多变的红外场景下都面临着如下一些问题：1）模型判别能力不足，容易混淆真实目标与其他高频背景干扰物；2）超参数的选择高度依赖于具体的图像背景，对变化多样的场景鲁棒性不足。

为了解决上述问题，本章首先构建了一个开放的单帧红外小目标（Single-frame InfraRed Small Target, SIRST）检测基准数据集，用于机器学习模型的训练、测试以及作为不同类型方法之间比较的基础。其次，为了以端到端的方式从标记数据中自动学习具有足够语义判别性的红外小目标特征表示，本章设计一个双向非对称的注意力调制（Asymmetric Bidirectional Attentional Modulation, ABAM）模块并将其嵌入基准网络中，构造出相应的双向非对称注意力调制网络（ABAM Network, ABAMNet）。在特征金字塔网络或者 U-Net 的基础上，除了自顶向下的反馈调制通路，ABAMNet 还额外添加了一条自底向上的特征调制通路。通过采用不同尺度的通道注意力模块，双向调制通路能够在网络不同层的特征之间互相交换高层语义信息和目标细节信息，提高了红外小目标的检测能力。出于可重复研究的考虑，本章的数据集、代码和训练好的模型参数可以从项目主页上获取[1]。

本章的具体内容安排如下：第 3.1 节梳理了模型驱动方法的不足以及采用数据驱动方法检测红外小目标的优点和难点，进而阐释了本章工作的研究动机与意义；第 3.2 节介绍了本章所构建的红外小目标检测数据集 SIRST，并且基于此数据集，对红外小目标的特性进行了统计分析；第 3.3 节详细描述了本章所提出的 ABAM 模块以及相应的网络实例；第 3.4 节针对两种 ABAMNet 的具体实例进行了消融实验以验证双向非对称注意力调制对于红外小目标的有效性，同时还对比了其他方法以验证所提出方法的检测性能。第 3.5 节对本章工作的内容进行了小结。

3.1　引言

在红外小目标检测领域，目前的研究大多集中于基于低秩稀疏分解[12,13,148,153] 或者局部对比度度量[10,11,142] 的方法。无论是将小目标视作破坏背景连续性的突出物体，还是低秩背景中的稀疏分量，这些方法大多通过建立模型，直接或间接地度量红外小目标与其周围区域之间的对比度以获得相应的目标显著性图，最后选取适当的阈值将潜在目标分割出来。这些模型驱动方法基于已知的物理机制和领域知识，不需要庞大的、带标注的数据集，计算上相对简单快捷，但是在实际应用中存在如下问题：

[1]https://github.com/YimianDai/open-acm

（1）在实际场景的红外图像中，红外弱小目标常常不满足这些方法所基于的显著性、稀疏性、高局部对比度假设，但与此同时，图像的复杂背景中却往往存在许多非目标的干扰物体满足这类假设。因此，通过简单的模型假设很难区分真实目标和少许杂波干扰，这会造成较高的漏检率或虚警率。

（2）这些方法依据对红外小目标及其背景的假设建立数学模型，其中部分超参数反映和承载了这些假设，例如滤波方法和局部对比度方法中对目标尺寸的假设、低秩稀疏分解方法中平衡背景低秩和目标稀疏程度的权重系数、目标图像分割的阈值等。这些超参数数值的选取大多与图像背景高度相关，且较为敏感，对于复杂多变的场景来说，其鲁棒性不足。

需要注意的是，模型驱动方法难以区分真实目标和背景相似干扰物的根源在于其所采用的特征过于简单、语义判别能力不足。例如，低秩稀疏分解方法往往直接使用原始图像的图像块构造冗余矩阵或张量，而局部对比度方法则采用图像块的均值、最大值、熵等简单统计量来计算局部对比度。然而，在信号杂波比较低的情况下，判别红外弱小目标还需要对更大范围的上下文乃至整体场景的高级语义理解，用于克服本征特征不足的问题。另一方面，模型驱动方法的超参数对图像场景变化的敏感性问题实质上反映了信号处理与机器学习在优化问题上的区别，即对模型泛化性能的关注与否。信号处理模型大多用于一些低层视觉（Low-Level Vision）任务中，比如图像去噪与去模糊[41]、图像修复与超分辨率等，优化目标在于使得模型能够在已知的给定信号上取得令人满意的效果。机器学习模型则强调对未来未知样本的预测性能，即泛化性能。当样本容量较小时，过分追求模型在已知样本上的效果，即经验风险最小化，容易导致模型过拟合，无法很好地预测未来样本。因此，采用机器学习方法来检测红外小目标，有助于提高对复杂多变场景的鲁棒性。

近年来，深度学习在图像分类、目标识别、语义分割等计算机视觉的诸多任务上都取得了令人瞩目、远远超越传统方法的成就。其能够以端到端的方式自动从数据中提取相关特征，学习到图像的高层语义，可以克服模型驱动方法特征语义判别能力不足的问题。然而，在目前的红外小目标检测领域，深度学习尚未成为主流，其主要原因在于：

（1）缺少公开的基准数据集：众所周知，深度学习需要大量数据用于训练，但是目前为止，尚无一个公开且具有高质量标注的单帧红外小目标检测数据集可以用来训练、测试深度网络。红外小目标数据获取难度较大，在没有公开数据集的情况下，应用深度学习检测红外小目标具有一定的数据门槛。

（2）分辨率与语义之间的极端矛盾：一方面，相较于通用视觉任务中的小目标（Small Object，以 Microsoft COCO 数据集为例，大小大约为 32×32 个像素），由于成像距离远，红外小目标（Small Target）在图像中通常仅占据几个像素，其特征更容易随着网络层数的增加而被周围的背景特征所覆盖。因此，其极小的尺寸要求网络能够在高分辨率的特征图上进行预测。另一方面，许多真实场景中，红外小目标往往被埋没在信号杂波比较低的复杂背景中，缺乏相应的纹理和

形状特征。在本征特征不足的情况下检测这些目标，需要网络具有对大范围的上下文、乃至整个图像场景的高层语义理解能力。然而，深度卷积网络通常采用逐渐减小特征图尺寸的方式来增大感受野，从而学习到更多的语义信息，其特征图的分辨率与语义层次往往是一对内生的矛盾。

为了解决上述问题，本章首先构建了一个单帧红外小目标（Single-frame InfraRed Small Target, SIRST）检测数据集，用于深度学习模型的训练、测试以及不同检测方法之间的比较。同时还分析了 SIRST 数据集中红外小目标的统计特性，并以此为准绳重新审视了众多模型驱动方法背后的假设。其次，考虑到红外小目标在特征分辨率与语义层次之间的极端矛盾，本章还构建了一个双向非对称的注意力调制机制用于特征金字塔中的跨层特征融合。其中自顶向下的调制通路采用全局通道注意力模块，用于将网络高层特征的语义信息反馈到低层特征，编码目标上下文；而自底向上的调制通路则采用局部通道注意力模块[1]，用于将低层特征的细节信息嵌入到高层特征中。不同于大多数仅采用自顶向下单向调制的同类工作，本章方法强调特征金字塔不同层之间高层语义信息和细节特征的互相交换，特别是基于局部通道注意力机制的自底向上通路，有助于在高层特征中保存和强调红外弱小目标，避免其被背景特征所淹没。

3.2 红外小目标数据集的构建与特性分析

目前，SIRST 数据集包含 427 幅来自不同场景、包含小目标的红外图像，一共 480 个小目标实例，其中 50% 的图像被选作训练集，20% 作为验证集，30% 测试集。由于短波和中波红外的视频序列非常难以公开获取，因此 SIRST 数据集中还包含部分波长为 950 纳米的近红外图像。图 3.1 中展示了 SIRST 数据集的一些代表性图像，从中可以看出，相当多的红外小目标本身非常暗淡，并且被淹没在杂乱无章的复杂背景中。即使对于人类视觉系统来说，检测它们也不是一件容易的事，这不仅需要集中注意力搜索，还要求对整体图像的场景语义具有一定程度的理解能力。

为了丰富红外小目标检测的建模形式，方便在未来将其从目前主流的图像分割范式拓展成更多元、更贴近计算机视觉发展趋势的方式，SIRST 数据集一共提供了五种图像标注类型，分别为图像类别标注（Category Labeling）、实例分割标注（Instance Segmentation）、边界框标注（Bounding Box）、语义分割标注（Semantic Segmentation）和实例发现标注（Instance Spotting），具体如图 3.2 所示。对于包含小目标的红外图像，在完成最为耗时的实例分割标注后，边界框标注、语义分割标注和实例发现标注均采用连通域分析和坐标转换得到。其中，每个红外序列只选择一幅代表性图像，因此人工神经网络无法通过机械记忆训练集中的目标和背景来检测验证集和测试集中的目标。

[1]局部通道注意力最早被设计用于第四章图像分类任务中特征的自我精炼，不同的是，本章将其用于自底向上的跨层特征调制。

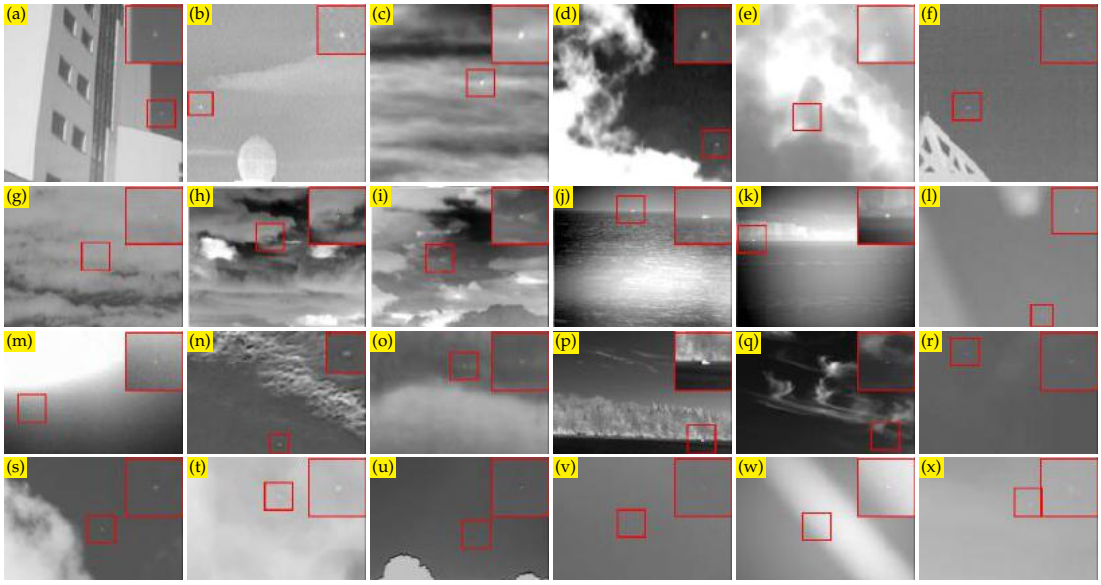
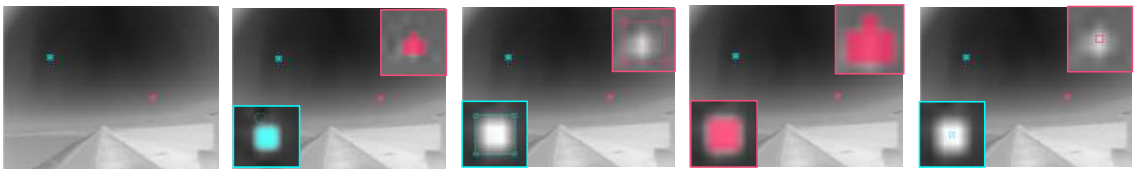


图 3.1　SIRST 数据集的部分代表性图像

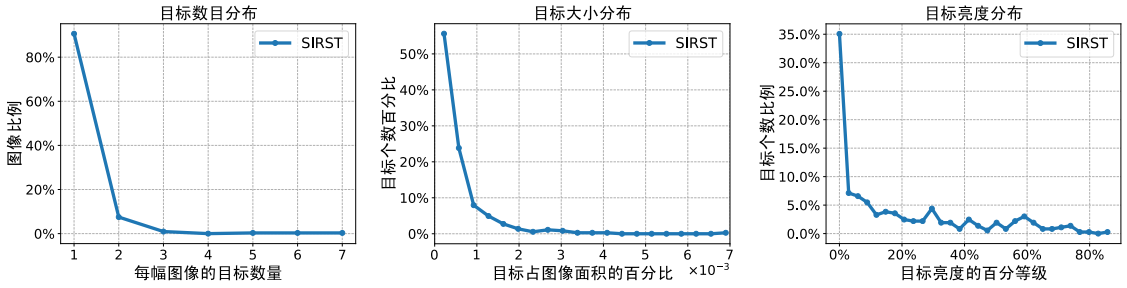


(a) 类别标注　(b) 实例分割标注　(c) 边界框标注　(d) 语义分割标注　(e) 实例发现标注

图 3.2　SIRST 数据集中不同标记类型的示意图

在完成对 SIRST 数据集的标注后，可以利用相应的标注信息对数据集的部分特性进行统计分析，以此作为对众多红外小目标检测方法背后假设的重新审视。图 3.3(a) 展示了每幅图像中红外小目标数量的分布情况，从中可以看到，大约 90% 的图像仅包含一个小目标。这个事实是众多显著性检测方法、局部对比度方法对于小目标全局唯一性假设的基础。然而，不可忽略的是，在 SIRST 数据集中大约仍有 10% 以上的图像包含有多个目标，仅仅假设目标是图像中唯一最显著的区域会导致算法忽略其他目标，从而造成大量漏检。图 3.3(b) 展示了红外小目标占图像整体大小的比例分布情况。从中可以看到，大约有 55% 的目标仅占到图像大小的 0.02%，给定一幅 300×300 的图像，目标大小仅为 3×3 个像素。通常，越小的物体越难以被网络识别，而且需要更多的上下文来辅助推理，而红外小目标将这种困难推到了一个极端的情形。红外小目标这种尺寸上的特点也决定了无法直接采用为 ImageNet 数据集上的图像分类任务设计的深度网络。为了更好地检测红外小目标，不仅应该重新设计骨干网络的下采样方案，还应该对于注意力模块中的特征上下文聚合方案进行重新设计，这也是本文第三章至第六章的核心出

发点。图 3.3(c) 展示了小目标亮度在图像中整体亮度中的分布情况，从中可以看到，仅有 35% 的小目标在图像中是最亮的。因此，对于 SIRST 数据集中的图像，采用阈值分割方法或者选择原始图像中最亮的像素仅能检测出 35% 的目标，存在大量漏检。此外，还需要注意的是，有 65% 的小目标，其亮度与背景非常相似，尤其是 16.7% 的小目标亮度甚至低于平均背景亮度。因此，红外小目标有很大可能并非图像中的显著区域。



(a) 目标数目分布　　(b) 目标大小分布　　(c) 目标亮度分布

图 3.3　SIRST 数据集的特性统计图

3.3　双向非对称注意力调制网络

本节将在特征金字塔网络（Feature Pyramid Networks, FPN）[104] 和 U-Net[105] 这两个基准网络的基础上，构建相应具体的双向非对称注意力调制网络（Asymmetric Bidirectional Attentional Modulation Network, ABAMNet）实例，即 ABAM-FPN 和 ABAM-U-Net，旨在解决以下两个红外小目标检测中的关键问题：1）对于本征特征稀缺的红外小目标，如何构造一个具有足够语义判别能力的深层网络；2）如何在编码高层语义信息的同时，保存红外弱小目标的细节特征。与上一章方法相同，本章遵从红外小目标检测领域的主流范式[12,153]，将其建模为单帧图像的语义分割问题。

3.3.1　自底向上的局部通道注意力调制

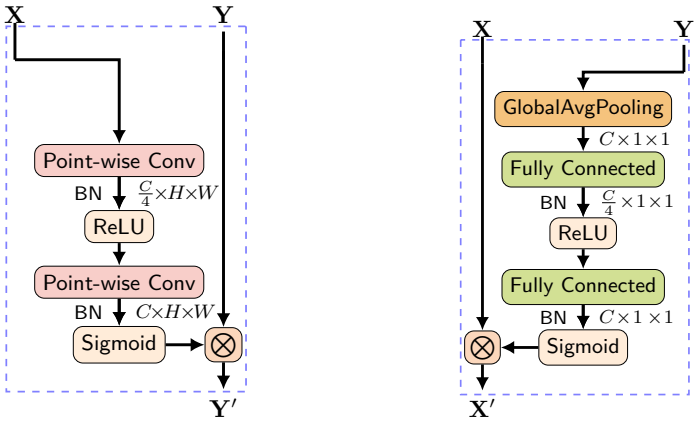
本小节介绍自底向上的局部通道注意力调制（Local Channel Attention Modulation, LCAM）。给定 $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{C\times H\times W}$ 为特征金字塔中相邻两层的输出，且已经经过了上采样和通道数变换使得两者具有相同的大小，默认 $\mathbf{Y}$ 为感受野更大、相对语义层次更高的特征图。为了保存并强调红外小目标的细节信息，在自底向上的调制通路中，局部通道注意力模块 $\mathbf{L}$ 在低层特征图 $\mathbf{X}$ 上以逐元素的方式聚合局部的通道特征上下文。自底向上的调制权重 $\mathbf{L}(\mathbf{X}) \in \mathbb{R}^{C\times H\times W}$ 的计算公式可以表示为：

$$\mathbf{L}(\mathbf{X}) = \sigma\left(\mathcal{B}\left(\mathrm{PWConv}_2\left(\delta\left(\mathcal{B}\left(\mathrm{PWConv}_1(\mathbf{X})\right)\right)\right)\right)\right) \tag{3.1}$$

式中，PWConv 代表逐点卷积[182]（Point-wise Convolution），σ 为 Sigmoid 函数，δ 为线性整流单元[183]（Rectified Linear Unit, ReLU），$\mathcal{B}$ 代表批标准化[184]（Batch Normalization, BN）层，r 是通道压缩比例。PWConv_1 和 PWConv_2 的卷积核大小分别为 $\frac{C}{r} \times C \times 1 \times 1$ 和 $C \times \frac{C}{r} \times 1 \times 1$。对于相对高层的特征图 $\mathbf{Y}$，其被局部通道注意力模块 $\mathbf{L}$ 自底向上调制后的特征图 $\mathbf{Y}' \in \mathbb{R}^{C \times H \times W}$ 的计算公式如下：

$$\mathbf{Y}' = \mathbf{L}(\mathbf{X}) \otimes \mathbf{Y} \tag{3.2}$$

式中，$\otimes$ 代表附加了维数对齐和广播（Broadcasting）机制的逐元素点乘。LCAM 模块的具体结构如图 3.4(a) 所示。



(a) LCAM 模块　　(b) GCAM 模块

图 3.4　LCAM 模块和 GCAM 模块示意图

3.3.2　自顶向下的全局通道注意力调制

本小节介绍自顶向下的全局通道注意力调制（Global Channel Attention Modulation, GCAM），其全局通道注意力模块采用挤压-激发网络（Squeeze-and-Excitation Network, SENet）中的注意力机制[21]。不同与 SENet 将其用作特征的自我精炼，GCAM 模块将其用于在低层特征中嵌入高层特征的语义信息。对于相对高层的特征图 $\mathbf{Y}$，首先采用全局平均池化来聚合全局上下文信息得到通道统计量 $\mathbf{y} \in \mathbb{R}^C$：

$$\mathbf{y} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{Y}[:, i, j]. \tag{3.3}$$

为了节省网络参数，GCAM 模块采用两个瓶颈结构的全连接（Fully Connected）层来计算自顶向下的全局调制权重 $\mathbf{G}(\mathbf{Y}) \in \mathbb{R}^C$：

$$\mathbf{G}(\mathbf{Y}) = \sigma\left(\mathcal{B}\left(\mathbf{W}_2 \delta\left(\mathcal{B}\left(\mathbf{W}_1 \mathbf{y}\right)\right)\right)\right), \tag{3.4}$$

式中，$\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ 表示第一个全连接层，用于降低通道维数，$\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ 为第二个全连接层，用于恢复通道数。对于相对低层的特征图 $\mathbf{X}$，其被全局通道注意力模块自顶向下调制后的特征图 $\mathbf{X}' \in \mathbb{R}^{C \times H \times W}$ 的计算公式如下：

$$\mathbf{X}' = \mathbf{G}(\mathbf{Y}) \otimes \mathbf{X} \tag{3.5}$$

式中，为了与 $\mathbf{X}$ 的维数对齐，在逐元素点积之前，$\mathbf{G}(\mathbf{Y})$ 会被重构和广播为 $C \times H \times W$ 大小。GCAM 模块的具体结构如图 3.4(b) 所示。

3.3.3 网络架构

在获得了 GCAM 和 LCAM 输出的特征图 $\mathbf{X}'$ 和 $\mathbf{Y}'$ 之后，融合后的特征 $\mathbf{Z} \in \mathbb{R}^{C \times H \times W}$ 可以由两者相加得到：

$$\mathbf{Z} = \mathbf{X}' + \mathbf{Y}' = \mathbf{G}(\mathbf{Y}) \otimes \mathbf{X} + \mathbf{L}(\mathbf{X}) \otimes \mathbf{Y}. \tag{3.6}$$

双向非对称注意力调制（Asymmetric Bidirectional Attentional Modulation, ABAM）模块的具体结构如图 3.5 所示。从中可以看到，在特征金字塔跨层的特征融合中，两个轻量级注意力模块聚合的特征上下文尺度不同，特征图 $\mathbf{X}$ 和 $\mathbf{Y}$ 作为彼此特征调制的指导信息来源，互相交换了高层语义特征和低层细节信息。

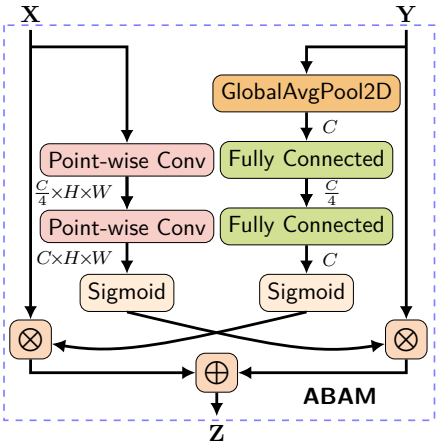


图 3.5　双向非对称注意力特征调制模块示意图

考虑到图 3.3(b) 中所展示的红外小目标的尺寸分布，本章在 ResNet-20-V2[185] 的基础上，重新设计了骨干网络的下采样方案，具体如表 3.1 所示。其中，b 是骨干网络每个阶段（Stage）的残差块（Residual Block, ResBlock）数量，通过改变 b 可以对网络进行缩放，得到不同深度的网络结构。当 $b = 3$ 时，则是标准的 ResNet-20。此外，还可以从表 3.1 中看到，为了保存红外弱小目标，网络中的 Conv-1 和 Stage-1 阶段并不进行下采样，因此 ABAMNet 最后的预测是在同输入图像相同大小的特征图上进行的。最后，为了展示 ABAM 模块的通用性，本章选取 FPN 和

U-Net 两种基准网络作为 ABAM 模块的宿主网络，并由此构建 ABAMNet 的两个具体实例，即 ABAM-FPN 和 ABAM-U-Net，相应的网络架构如图 3.6 所示。其中，绿色线表示特征图的下采样操作，红色线表示由双线性插值实现的特征图上采样，蓝色线则是由逐点卷积实现的通道数变换操作[182]。

表 3.1　ABAMNet 的骨干网络架构

网络阶段	特征图大小	通道与残差块数量
Conv-1	480×480	3×3 conv, 16
Stage-1 / UpStage-1	480×480	$\begin{bmatrix} 3 \times 3 \text{ conv}, 16 \\ 3 \times 3 \text{ conv}, 16 \end{bmatrix} \times b$
Stage-2 / UpStage-2	240×240	$\begin{bmatrix} 3 \times 3 \text{ conv}, 32 \\ 3 \times 3 \text{ conv}, 32 \end{bmatrix} \times b$
Stage-3 / UpStage-3	120×120	$\begin{bmatrix} 3 \times 3 \text{ conv}, 64 \\ 3 \times 3 \text{ conv}, 64 \end{bmatrix} \times b$

3.4　实验结果与分析

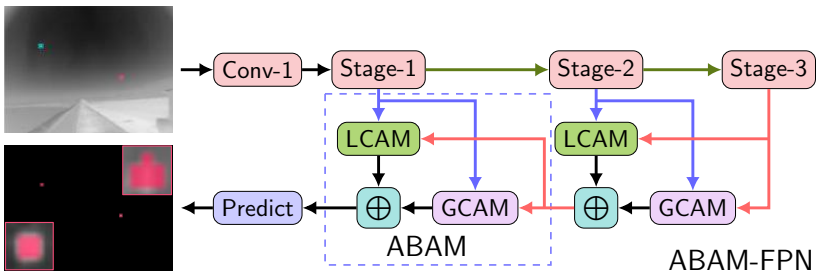
为了验证 ABAMNet 在结构设计上的合理性和有效性，本节将进行详细的消融实验和对比实验，具体探索以下问题：

（1）问题一：本章的核心出发点在于针对红外小目标的尺度特点，需要重新设计骨干网络以及相应的注意力模块。3.4.2 小节将首先研究骨干网络中的下采样方案对于最终红外小目标检测性能的影响。

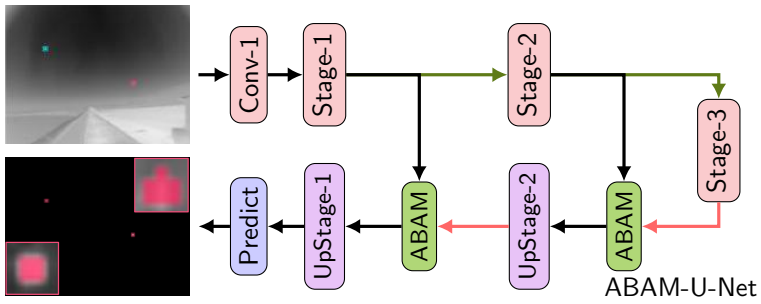
（2）问题二：与大多数只对低层特征进行自顶向下调制的网络不同，ABAMNet 还额外添加了一条反向自底向上的调制通路，用于将低层特征的细节信息嵌入高层特征中。3.4.2 小节将研究在给定相同网络参数的情况下，相较于其他注意力调制方案，双向调制机制能否获得更好的红外小目标检测性能。

（3）问题三：对于自顶向下和自底向上的两条调制通路，ABAMNet 以一种不对称的方式，分别采用了全局通道注意力模块和局部通道注意力模块来实现。3.4.2 小节还将研究给定相同网络参数的情况下，相较于其他对称的注意力调制机制，非对称注意力调制能否获得更好的红外小目标检测性能。

（4）问题四：3.4.3 小节将研究相比于其他模型驱动方法和其他深度网络方法，本章所提出的双向非对称注意力调制网络能否获得更好的的红外小目标检测性能。



(a) ABAM-FPN 架构示意图



(b) ABAM-U-Net 架构示意图

图 3.6　ABAM-FPN 和 ABAM-U-Net 的架构示意图

3.4.1　实验设置

本节所有实验在 3.2 节构建的 SIRST 数据集上进行。作为语义分割任务的默认评价指标，交并比（Intersection over Union, IoU）的定义如下：

$$\mathrm{IoU} = \frac{\sum_i^N \mathrm{TP}[i]}{\sum_i^N \mathrm{T}[i] + \mathrm{P}[i] - \mathrm{TP}[i]}, \tag{3.7}$$

式中，N 为测试集中图像的数量，TP 代表被模型预测正确的小目标区域，T 代表标注图像中的小目标区域，P 代表模型预测的小目标区域。然而，红外小目标的特殊性在于其较大的尺度变化，不同的目标之间像素数可能会相差 20 到 100 倍，因此 IoU 更容易反映模型在相对较大的红外小目标上的分割情况，而非所有目标。此外，大部分模型驱动方法只关心能否将目标标识出来，并不追求目标分割的完整性。因此，对于相对较大的小目标，模型驱动方法的 IoU 指标通常更低。为了能够更好地反映模型在红外小目标数据集上的性能，本章还根据红外小目标的特点构造了归一化的交并比（normalized IoU, nIoU）

$$\mathrm{nIoU} = \frac{1}{N}\sum_i^N \frac{\mathrm{TP}[i]}{\mathrm{T}[i] + \mathrm{P}[i] - \mathrm{TP}[i]} \tag{3.8}$$

相较于 IoU，nIoU 先计算每一幅图像上的交并比然后再在数据集上求平均，避免了较大的目标在 IoU 指标上覆盖较小的目标，也能够更好地平衡模型驱动方法和数据驱动方法。此外，本节

最后还采用了接收机工作特性（Receiver Operating Characteristic, ROC）曲线来评估各种小目标检测方法在滑动阈值下的检测性能。需要注意的是，不同于传统的滤波、显著性检测方法所产生的数值连续的目标图像，本章 ABAMNet 直接输出的是 0 或 1 的二值图。局部信噪比增益（Local Signal to Noise Ratio Gain, LSNRG）、背景抑制因子（Background Suppression Factor, BSF）、信噪比增益（Signal to Clutter Ratio Gain, SCRG）这些针对早期滤波方法设计的背景抑制性能指标，极其容易出现由于除零导致的正无穷大 **Inf**，因而不再采用。

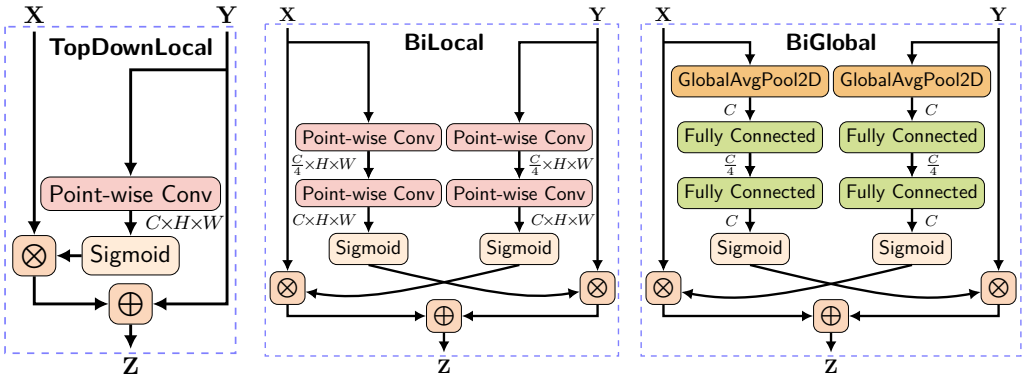
为了充分评估 ABAMNet 的性能，将其与多种方法进行了对比。其中，属于模型驱动方法的有局部对比度方法[11]（Local Contrast Method, LCM）、局部显著性方法[186]（Local Saliency Method, LSM）、基于小面核与随机游走（Facet Kernel and Random Walker, FKRW）的方法[159]、多尺度块对比度度量（Multi-scale Patch-based Contrast Measurement, MPCM）方法[10]、红外块图像（Infrared Patch-Image, IPI）模型[12]、基于奇异值部分和的非负 IPI 模型[148]（Non-Negative IPI Model via Partial Sum Minimization of Singular Values, NIPPS）、重加权红外块张量（Reweighted Infrared Patch-Tensor, RIPT）模型[13]、非凸秩逼近最小化 Non-convex Rank Approximation Minimization (NRAM) 方法[152]。此外，对于数据驱动方法，还选取了特征金字塔网络[104]（Feature Pyramid Network, FPN）、U-Net[105]、TBC-Net[187]、选择核（Selective Kernel, SK）网络[188]、全局注意力上采样（Global Attention Upsample, GAU）网络[189]，并以 FPN 和 U-Net 作为宿主网络构造了相应的 SK-FPN、SK-U-Net、GAU-FPN、GAU-U-Net 作为对比方法。所有的网络均选择 Soft-IoU[190] 作为损失函数，自适应梯度[191]（Adaptive Gradient, AdaGrad）方法作为网络的优化算法，学习率（Learning Rate）为 0.05，权重衰减（Weight Decay）率为 0.0001，批大小（Batchsize）为 8，一共训练 300 轮。为了堆叠大小不同的图像，每幅红外小目标图像在读入时都被拉伸为 512×512 大小。在训练过程中，随机裁剪大小为 480×480。

表 3.2　模型驱动方法的具体参数设置

方法	参数设定
LCM	子块大小: 8×8;，滑动步长: 4; 阈值系数: $k=1$
LSM	阈值参数：$a=3, g=0.6$
FKRW	$K=4$，$p=6$，$\beta=200$，窗口尺寸: 11×11
MPCM	$N=1,3,...,9$
IPI	块大小: 50×50，滑动步长: 10，$\lambda=\frac{L}{\sqrt{\min(m,n)}}$, $L=4.5$，阈值系数: $k=10$，$\varepsilon=10^{-7}$
NIPPS	块大小: 50×50，滑动步长: 10，$\lambda=\frac{L}{\sqrt{\min(m,n)}}$, $L=2.0$, 约束比例: 0.11，阈值系数: $k=10$
RIPT	块大小: 50×50，滑动步长: 10，$\lambda=\frac{L}{\sqrt{\min(I,J,P)}}$, $L=0.001, h=0.1, \epsilon=0.01, \varepsilon=10^{-7}$，$k=10$
NRAM	块大小: 50×50，滑动步长: 10，$\lambda=1.0$

3.4.2 消融实验与分析

本小节在 SIRST 数据集上进行消融实验（Ablation Study）以验证本章所提出的双向非对称注意力调制的合理性和有效性，消融实验中对比的模块架构如图 3.7 所示。



(a) TopDownLocal (b) BiLocal (c) BiGlobal

图 3.7 ABAM 模块消融实验中所对比的模块结构图

3.4.2.1 调整下采样方案的必要性

首先，对于问题一，在表 3.1 所展示的骨干网络基础上，通过将 Conv-1 的步长（Stride）设置为 2 并且添加步长为 2 的最大池化（Max Pooling）可以相应构造出多四倍下采样的网络。为了与本章所采用的 ABAM-FPN 和 ABAM-U-Net 区分，将采用该常规下采样方案的网络称为 Regular-ABAM-FPN 和 Regular-ABAM-U-Net. 由于两者网络之间具有相同的参数数量，唯一的差别在于由于不同下采样方案导致的特征图大小不同。表 3.3 展示了相应的分割精度比较。从中可以看出，采用表 3.1 下采样方案的网络性能明显优于常规下采样方案，尤其是在网络具有一定的深度之后。该结果表明，对于红外小目标检测来说，将特征图维持在较高的分辨率对于最终的检测性能至关重要，否则过度的下采样将导致在高层网络中的小目标特征丢失。

表 3.3 ABAM 模块与其他多种调制模块的分割精度对比

调制方案	FPN 作为宿主网络								U-Net 作为宿主网络							
	IoU				nIoU				IoU				nIoU			
	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
TopDownLocal	0.595	0.648	0.693	0.713	0.635	0.662	0.688	0.703	0.648	0.710	0.713	0.718	0.673	0.692	0.694	0.697
BiGlobal	0.599	0.660	0.685	0.693	0.645	0.674	0.696	0.684	0.682	0.716	0.723	0.730	0.688	0.708	0.707	0.719
BiLocal	0.591	0.662	0.713	0.722	0.657	0.694	0.709	0.714	0.670	0.715	0.718	0.742	0.680	0.710	0.713	0.720
Regular-ABAM	**0.683**	**0.703**	0.711	0.711	0.661	0.671	0.680	0.675	0.684	0.700	0.692	0.692	0.637	0.650	0.646	0.643
ABAM	0.645	0.700	**0.714**	**0.731**	**0.684**	**0.702**	**0.713**	**0.721**	**0.707**	**0.732**	**0.741**	**0.743**	**0.709**	**0.720**	**0.726**	**0.731**

3.4.2.2 双向注意力调制的重要性

对于问题二，在给定注意力调制均采用局部通道注意力模块情况下，设计了图 3.7(a) 所示的 TopDownLocal 模块与图 3.7(b) 所示的 BiLocal 模块用于对比单向自顶向下调制与双向调制之间的性能差异。两者卷积层的参数数量总和均为 C^2。将 ABAM-FPN 和 ABAM-U-Net 中的 ABAM 模块替换为 TopDownLocal 或 BiLocal 模块即可得到相应的用于消融实验的网络。表 3.3 展示了 TopDownLocal 和 BiLocal 在网络深度加深的情况下的性能比较，b 是表 3.1 中骨干网络每阶段的残差块数量。从中可以看到，在绝大多数情况下，BiLocal 网络都取得了比 TopDownLocal 网络更好的效果。$b = 3$ 时的 BiLocal 网络在 IoU 上可以取得与 $b = 4$ 时的 TopDownLocal 网络相当的效果，在 nIoU 指标上更加胜出。这表明在给定相同网络参数数量的情况下，双向注意力调制是比单向自顶向下调制更加有效的跨层特征融合方式。

3.4.2.3 非对称注意力调制的重要性

对于问题三，在均采用双向调制的情况下，将图 3.5 所示的 ABAM 模块与图 3.7(b) 所示的 BiLocal 模块和图 3.7(c) 所示的 BiGlobal 模块进行了对比。三者网络参数的数量相同，差异只在于在双向调制通路上所采用的注意力模块。BiLocal 模块和 BiGlobal 模块采用对称的注意力调制方案，即自顶向下和自底向上通路均采用局部通道注意力模块（LCAM）或者全局通道注意力模块（GCAM）。ABAM 则采用非对称的方案，在自底向上通路中采用局部通道注意力模块，而在自顶向下通路中采用全局通道注意力模块。表 3.3 展示了三者在网络深度加深的情况下在 SIRST 数据集上的性能比较。从中可以看到：1）在所有实验设置下，ABAM 均取得了比 BiLocal 和 BiGlobal 更好的性能，这表明了非对称注意力调制对于红外小目标的重要性，能够更好地平衡高层语义信息与低层细节信息；2）相比之下，大多数情况下，BiLocal 在 IoU 和 nIoU 指标上均显著好于 BiGlobal，这表明对于红外小目标检测，相比全局语义信息，局部细节信息对于最后的检测性能更为重要。特别需要注意的是，在 $b = 4$ 时，BiGlobal 在 nIoU 指标上会出现下降，说明随着网络的加深，红外弱小目标被背景特征淹没的风险也在逐渐增大，而全局注意力机制对于保存弱小目标并没有帮助。

3.4.3 方法对比与分析

随着 SIRST 数据集的构建完成，采用深度学习方法检测红外小目标有了数据基础。对于问题四，表 3.4 展示了 ABAM-FPN 和 ABAM-U-Net 与其他 14 种方法在 SIRST 数据集上的定量评价指标比较，从中可以看出：1）在具有足够训练数据的情况下，所有深度网络方法在 IoU 和 nIoU 指标上表现得均比非学习的模型驱动方法好。其中，ABAM-FPN 和 ABAM-U-Net 在所有方法中均取得了最佳的性能，这表明了在特征金字塔网络中进行双向非对称注意力调制的有效性。2）比较模型驱动方法内部可以看到总体上，低秩稀疏分解方法能够取得比局部对比度度量

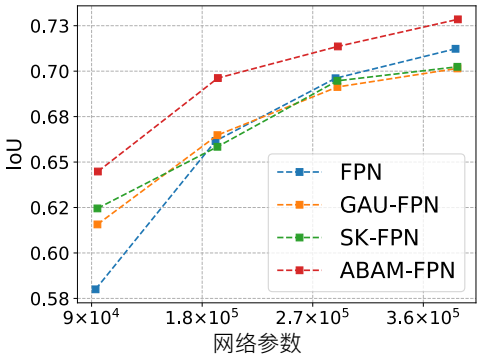
方法更好地检测性能。而在局部对比度度量方法中，MPCM 方法取得了最好的性能，接近低秩稀疏分解方法，这个观察也是本文第六章选择将 MPCM 方法网络模块化的主要原因。3）模型驱动方法的 nIoU 数值通常会高于其 IoU 数值，而深度网络模型则相反。这证实了构建 nIoU 指标的合理性，即 IoU 会倾向于反映模型在较大目标上的检测性能，容易忽略弱小目标，同时模型驱动方法由于在设计时忽视目标的完整性，其 IoU 指标也会偏低。因此，相较于 IoU，nIoU 是一个更符合红外小目标特点的评价指标。

表 3.4　ABAM-FPN 和 ABAM-U-Net 与其他 14 种方法的定量评价指标比较

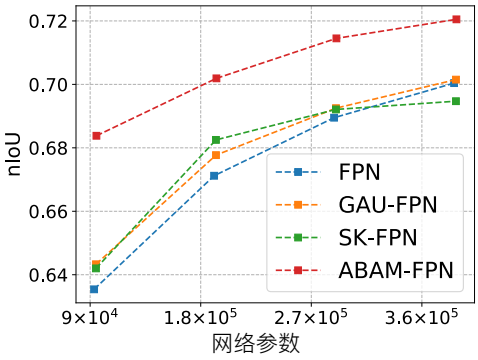
方法	模型驱动方法							数据驱动方法								
	局部对比度度量			低秩稀疏分解				FPN 作为宿主网络				U-Net 作为宿主网络				
	LCM	LSM	MPCM	IPI	NIPPS	RIPT	NRAM	FPN	SK	GAU	ABAM	TBC	U-Net	SK	GAU	ABAM
IoU	0.193	0.1864	0.357	0.466	0.473	0.146	0.294	0.720	0.702	0.701	**0.731**	0.734	0.733	0.708	0.718	**0.743**
nIoU	0.207	0.2598	0.445	0.607	0.602	0.245	0.424	0.700	0.695	0.701	**0.721**	0.713	0.709	0.699	0.697	**0.731**

图 3.8 展示了在网络深度逐渐增加的情况下，ABAMNet 与其他深度网络在 SIRST 数据集上的性能比较。从中可以看到，在网络参数数量相近的情况下，ABAM-FPN 与 ABAM-U-Net 的性能稳定地好于 FPN/U-Net、SK-FPN/SK-U-Net 和 GAU-FPN/GAU-U-Net。在 nIoU 指标上，$b = 2$ 时的 ABAM-FPN 可以取得与 FPN 在 $b = 4$ 时近乎相同的性能，却只需后者大约 50% 的网络参数。这表明对于红外小目标检测，一味地加大网络深度并不是最有效的策略，红外小目标并不需要特别复杂的特征表示，关键在于如何在获得高层语义信息的同时保留弱小目标的细节。对比 FPN 和 SK-FPN 两者的 nIoU 曲线也可以看出，在网络较浅时，比如 $b = 1$ 或者 $b = 2$，全局通道上下文信息有助于小目标检测。但是在网络相对较深时（$b = 4$），SK-FPN 性能低于 FPN，这表明不恰当地使用全局通道上下文反而会使得网络性能相对下降。

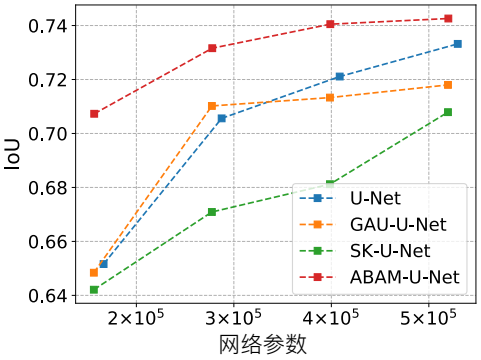
最后，图 3.9 比较了 ABAM-FPN 和 ABAM-U-Net 与其他五种方法的 ROC 曲线。属于数据驱动方法的 ABAM-FPN 和 ABAM-U-Net 两者效果均大幅好于 MPCM、IPI、RIPT、NIPPS 这些模型驱动方法，这再次显示了数据驱动方法在性能上的优势。另一个需要注意的是，尽管 RIPT 在 IoU 和 nIoU 性能指标上不如 IPI 和 MPCM，但其 ROC 曲线性能好于这两者。这是由于 IoU 和 nIoU 反映的是在固定阈值下的分割效果，相比 IPI 模型，RIPT 模型为了更好地检测红外弱小目标而设计的局部结构权重会导致分离出的目标更加的不完整，但是其能更准确地检测出红外弱小目标的存在。因此，在反映滑动阈值下分割总体情况的 ROC 曲线中，RIPT 模型表现相对更好。



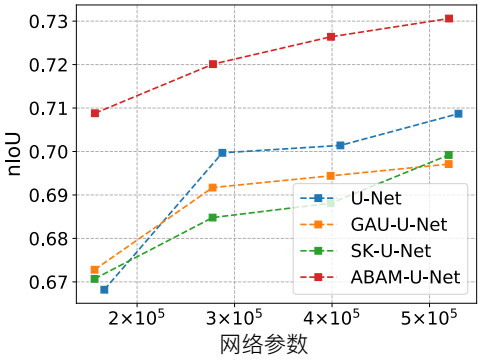
(a) 以 FPN 为宿主网络的 IoU 比较



(b) 以 FPN 为宿主网络的 nIoU 比较



(c) 以 U-Net 为宿主网络的 IoU 比较



(d) 以 U-Net 为宿主网络的 nIoU 比较

图 3.8　ABAM-FPN 和 ABAM-U-Net 与其他深度网络在 SIRST 数据集上的性能比较

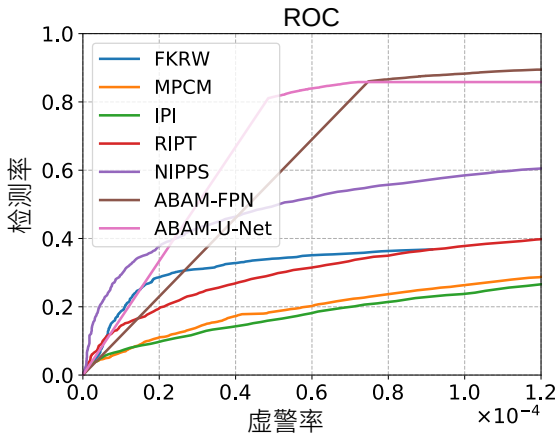


图 3.9　ABAM-FPN 和 ABAM-U-Net 与其他方法的 ROC 比较

3.5 本章小结

针对红外小目标检测任务，本章实现了从信号处理范式、模型驱动方法向机器学习范式、数据驱动方法的转变。为了能够训练和测试深度网络，首先构建了一个单帧红外小目标检测的基准数据集，并且统计分析了红外小目标的若干特点。其次，为了克服深度网络在检测小目标时面临的语义与分辨率之间的极端矛盾，在特征金字塔网络的基础上，设计了一个双向非对称的注意力调制模块。通过采用不同尺度的通道注意力机制，双向调制通路能够在网络不同层的特征之间互相交换高层语义信息和目标细节信息，从而在融合高层语义信息的同时保存红外小目标。详细的消融实验和对比实验结果表明，在 IoU、nIoU 和 ROC 等指标上，双向非对称注意力调制好于其他单向的或者对称的调制方法，同时也好于其余多种对比方法，能够有效地提升深度网络的检测性能。

第四章　基于注意力激活单元的图像分类与小目标分割

随着第三章中 SIRST 数据集的建构完成，采用机器学习检测红外小目标有了数据基础。特别是深度学习，能够从标记数据中自动学习特征表示，有望缓解红外小目标人工特征设计困难的问题。此外，注意力机制则可以为网络获取更加强大的特征表示能力，从而增强网络对于目标和背景干扰物的判别能力。然而，红外小目标的特点决定了在深度网络中必须及早使用注意力模块，甚至在第一个卷积层之后，这样才可以保存那些尺寸最小的小目标，避免其被周围的背景特征淹没。

在上述动机启发下，本章提出了一类注意力激活（Attentional Activation, ATAC）单元框架，采用轻量级且仅聚合局部特征上下文的注意力模块作为非线性激活单元。通过将网络中的线性整流单元（Rectified Linear Unit, ReLU）逐渐替换为 ATAC 单元，可以构建出全注意力网络（Fully Attentional Network），使得网络在仅增加很小比例参数的情况下可以获得更好的性能。ATAC 框架作为激活单元，具有通用性，因此为了证明 ATAC 框架的良好性能，本章还将其应用场景从小目标检测推广到更具一般性的图像分类任务。此外，考虑到红外目标序列较难获取、无法构建大规模数据集的问题，为了在更大规模的数据集上验证本章工作，还构建了与红外小目标具有相似特点的弱小冰山检测数据集（DiskoBay）以及基于 Microsoft COCO 数据集的 StopSign 数据集。出于可重复研究的考虑，本章方法的代码、训练好的模型参数、训练日志和 DiskoBay 数据集可以从项目主页上获取[1]。

本章的具体内容安排如下：第 4.1 节对于激活函数与注意力机制的相关工作进行了分析与讨论，并且阐述了本章的研究动机与意义。第 4.2 节首先介绍了激活函数与注意力机制形式上的相似性，然后给出了相应的 ATAC 单元实例，最后讨论了全注意力网络以及相应的参数量和计算量。第 4.3 节介绍了更大规模的 DiskoBay 数据集以及 StopSign 数据集。第 4.4 节首先对所提出的 ATAC 单元进行了详细的消融实验（Ablation Study），以验证注意力激活模块本身设计的合理性与应用方式上的有效性。同时，还将采用 ATAC 单元的网络与采用其他激活函数的网络以及使用其他注意力机制的网络进行了实验对比，以验证 ATAC 单元的性能表现。第 4.5 节对本章工作的内容进行了小结。

4.1　引言

近年来，深度学习领域的关键性技术进步包括新型的注意力机制[21] 和激活函数[183]。为了抓取长程的空间交互关系或者全局上下文，新的注意力模块往往被设计得结构更加精细、计算

[1]https://github.com/YimianDai/open-atac

更加复杂[22,95,172]。与此相反的是，虽然形式多样，但无论是人工设计[192]还是神经自动搜索[193]（Neural Architecture Search, NAS）得到的激活函数仍然保持着简单的标量函数形式。然而，在现代的网络架构中，尽管注意力机制和激活函数均被普遍使用，但这两个概念通常被视为网络中不同的组成部分，各自沿着不同的方向在演进发展，它们之间的相似性很少被讨论。

4.1.1 激活函数的研究进展

作为神经网络不可或缺的组成部分，在给定有限网络的情况下，更好的激活函数往往意味着网络更好的性能以及训练过程中更好的收敛性。线性整流单元[183]（Rectified Linear Unit, ReLU）是 AlexNet 取得巨大成功的关键之一[194]，由于其分段线性的特性，网络的梯度流可以在神经元的活跃路径上被良好地传递，这极大地减轻了诸如 Tanh 或 Sigmoid 这类饱和函数所带来的梯度消失问题。然而，由于 ReLU 在输入为负的区段导数恒为零，导致其对异常值特别敏感，反向传播中产生的大的梯度容易导致 ReLU 关闭，从而使得神经元死亡并且在后续训练过程中无法再恢复。为了缓解此问题，泄漏型线性整流单元[195]（Leaky ReLU）将输入负半轴的斜率设置为非零的超参数 α（例如 $\alpha = 0.1$），而随机泄漏型线性整流单元（Randomized Leaky ReLU, RReLU）则是更进一步，从一个连续性均匀分布中随机采样负输入值段的函数梯度。高斯误差线性单元（Gaussian Error Linear Unit, GELU）根据幅度对输入进行加权，而非像 ReLU 那样通过输入的符号[196]。另一个变体是可伸缩的指数型线性单元（Scaled Exponential Linear Unit, SELU），其通过引入自归一化性质来实现对特征的高级抽象表示[192]。不同于上述手工设计的激活函数，Ramachandran 等人采用自动搜索技术搜索最优的标量型激活函数，即输入是一个标量输出也是一个标量的函数，提出了 Swish 激活函数[193]（$x' = x \cdot \sigma(x)$，其中 σ 是 Sigmoid 函数），并在更深的网络模型和更多具有挑战性的数据集上取得了普遍好于 ReLU 的效果。需要注意的是，Swish 激活函数抛弃了 ReLU 分段线性函数的形式，而是采用非线性门控函数的形式。

增强 ReLU 的另一种方式是引入可学习的参数。例如，参数化的线性整流单元（Parametric Rectified Linear Unit, PReLU）可以在网络训练期间学习到负半轴上的响应斜率[197]。Yang 等人和 Agostinelli 等人则是将其推广为由更多可学习的分段线性函数组成的形式，其单元输出为一系列参数化的铰链函数之和[198,199]。不同于上述逐个像素都独立的激活单元，Kligvasser 等人提出了一种名为 xUnit 的空域激活函数[23]，通过逐深度卷积[65]（Depth-wise Convolution, DWConv）来学习局部邻域之间的空间连接以增强 ReLU。

4.1.2 注意力机制中的特征上下文聚合

自从在自然语言处理领域获得突破性的成就之后，注意力机制也被广泛地应用于计算机视觉的各个任务中[21,172]。这些注意力机制大多数是作为 ‘‘可插拔’’ 模块被集成到现有网络中，主要用途在于通过聚合特征图中的上下文信息来生成动态的权重以自适应地重新校正特征图[21]。

作为最后一届 ImageNet 挑战赛的冠军，挤压-激发网络（Squeeze-and-Excitation Network, SENet）采用瓶颈结构的两个全连接（Fully Connected, FC）层来显式地建模特征通道之间的相互依赖关系，从而对每个通道的特征图进行重新加权[21]。从特征上下文聚合尺度的角度，SENet 可以被称作全局通道注意力（Global Channel Attention）机制，因为其采用全局平均池化（Global Average Pooling, GAP）来生成每个通道的统计量。在此基础上，一种提高注意力模块性能的通用策略是在特征图上聚合更多、更精细的长程交互关系。其中，Woo 等人将全局通道注意力模块和全局空间注意力模块串联使得网络能够共同学习到需要加强的通道和空间位置[22]。聚集-激发网络（Gather-Excite Network, GENet）通过覆盖整个特征图的大卷积核（例如 56×56 的卷积核）聚合空域上的特征响应，并将合并后的信息重新分配给局部特征[172]。注意力增强的卷积神经网络（Attention Augmented Convolutional Network）采用全局的二维相对自注意机制作为图像分类的独立计算基元，并将其提取的特征与卷积特征拼接，从而克服了卷积的局部性缺陷[95]。不同于上述注意力机制均试图聚合全局特征上下文，Ramachandran 等人提出了一个局部自注意力层用来提取特征，通过将网络中的卷积全部都替换为局部自注意力层，可以构建出相应的全注意力视觉模型[66]（Fully Attentional Vision Model, FAVM）。该研究进一步指出，不同于其在后续网络核心块中的表现，在网络的初始层中使用局部自注意力层会导致比使用传统卷积更差的性能。表 4.1 在省略了瓶颈结构、批归一化、激活单元等细节结构的情况下，简短概括了注意力机制中各类特征上下文的聚合方案，其中 CAP 代表通道均值池化[22]（Channel Average Pooling, CAP），Conv 代表卷积（Convolution），PWConv 代表逐点卷积（Point-wise Convolution, PWConv）。

表 4.1　注意力模块中的上下文聚合方案

聚合尺度	聚合维度	聚合方式	参考文献
全局	空域	Conv (CAP($\mathbf{X}$))	[22,200]
	空域	DWConv($\mathbf{X}$)	[23,172]
	通道	FC(GAP($\mathbf{X}$))	[21,188]
	空域 + 通道	Conv(PWConv($\mathbf{X}$)) + FC(GAP($\mathbf{X}$))	[94]
局部	通道	PWConv($\mathbf{X}$)	本章
局部	空域 + 通道	PWConv(DWConv($\mathbf{X}$))	本章
局部多尺度	空域	$\text{DWConv}_1(\mathbf{X}) + \text{DWConv}_2(\mathbf{X})$	本章
局部多尺度	空域 + 通道	$\text{PWConv}(\text{DWConv}_1(\mathbf{X}) + \text{DWConv}_2(\mathbf{X}))$	本章
全局 + 局部	通道	FC(GAP($\mathbf{X}$)) + PWConv($\mathbf{X}$)	第五章

4.1.3　研究动机与意义

根据上文叙述可以得知，尽管目前已经有了一些可学习的激活单元，比如 PReLU 和 xUnit，但是这些激活单元仍然存在以下缺点：

（1）受限于 ReLU 的形式：当前可学习的激活函数主要致力于通过引入更加灵活的斜率[197]或者空间连接[23] 等方式来增强 ReLU。然而，以 Swish 函数为代表的网络自动搜索研究表明，放弃 ReLU 的形式可以获得性能和通用性更好的激活单元[193]。

（2）局部性不足：当前可学习的激活函数通常是对整个特征图施加全局相同的斜率[197]，比如 PReLU。然而，最近的研究表明，网络中的激活单元对于局部上下文和逐个元素各异的非线性激活方式具有明显的倾向性[23]。

为了解决上述问题，受注意力机制和激活函数两者相似性的启发，本章提出了一类动态且具有特征上下文感知特性的注意力激活（Attentional Activation, ATAC）单元。不同于大多数相关工作改进或者增强 ReLU 的思路，ATAC 单元的本质是通过设计一类仅聚合局部上下文且也只作用于局部特征的轻量级注意力模块，并将其用作网络中的激活单元。在非线性门控函数的框架下，ATAC 单元统一了激活单元和注意力机制。其中，传统的激活单元可以被看作是一个非上下文感知的注意力模块，而注意力机制则可以被看作是一个结构复杂的、上下文感知的激活单元。作为激活单元，ATAC 单元不仅可以在网络中引入非线性，还可以对网络中每一层的特征图进行上下文感知的重新校正（Recalibration）。通过在低层网络中及早抑制无关特征、强调相关特征，ATAC 单元可以使得网络能够更为有效地编码高层语义。

需要注意的是，与 FAMA 类似，ATAC 单元也提供了一种构建全注意力网络（Fully Attentional Network）的方式，即将网络中的激活函数（例如 ReLU）逐一替换为 ATAC 单元。但与 FAMA 不同的是，在网络初始层中使用 ATAC 单元并不会使得网络性能变差。在许多网络结构中，如果仅考虑增加的参数数量和所提升的性能，在初始层中使用 ATAC 单元比在网络中后期使用 ATAC 单元具有更高的性价比。这种差异可以用注意力机制不同的用法来解释。在 FAMA 中，自注意力层被用来代替卷积、学习如何提取有用的特征。在网络的初始层中，由于输入内容仅反映了图像原始的像素值，单个输入元素并不具备足够的信息并且在空间上高度相关，自注意力机制很难学习到边缘结构这样的有用的特征[66]。与之相反的是，在由 ATAC 单元构建的全注意力网络中，ATAC 单元仅仅负责激活和精炼特征，而特征抽取仍然使用卷积来完成。

值得指出的是，本文第三、四、五、六章或多或少都涉及到了‘‘通道注意力也应该具有尺度（Scale）属性’’这一想法，虽然出于论文总体架构安排的考虑，本章位于第三章之后，但是将尺度概念引入通道注意力机制最早源于本章，其他章节则在此基础上逐渐深化探索了更为多样的通道注意力模块和调制方式。具体而言，本章首先提出了局部通道注意力机制，挣脱了原先 SENet、SKNet 中聚合全局上下文的范式；随后，第三章进一步发展出了不同尺度的特征应当采用不同尺度的通道注意力这一想法，分别将全局和局部通道注意力用于自顶向下和自底向上的特征调制；最后，第五章为了应对待融合特征存在语义和尺度上的差异这一问题，探索了在同一个注意力模块内聚合多个不同尺度的通道特征上下文这一想法，构建了多尺度通道注意力模块。第六章则将局部注意力调制作为即插即用的模块用来提升红外小目标检测的性能。

4.2 注意力激活

4.2.1 注意力机制与激活函数的统一框架

给定一个通道数为 C、大小为 $H \times W$ 的特征图 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$，注意力机制的调制过程可以被抽象表示为

$$\mathbf{X}' = \mathbf{G}(\mathbf{X}) \otimes \mathbf{X}, \tag{4.1}$$

式中，$\otimes$ 代表附加了广播（Broadcasting）机制的逐元素点乘，$\mathbf{G}(\mathbf{X}) \in \mathbb{R}^{C \times H \times W}$ 是由注意力门控模块 $\mathbf{G}$ 产生的三维权重张量。给定一个具体的元素位置 (c, i, j)，式 (4.1) 的标量形式可以被表示为

$$\mathbf{X}'_{[c,i,j]} = \mathbf{G}(\mathbf{X})_{[c,i,j]} \cdot \mathbf{X}_{[c,i,j]} = g_{c,i,j}(\mathbf{X}) \cdot \mathbf{X}_{[c,i,j]}. \tag{4.2}$$

式中，g 是一个功能复杂的门控函数，对于给定元素位置 (c, i, j)，其负责聚合相关的特征上下文并产生对应的注意力权重。与此同时，激活函数也可以被统一表述成如下的门控函数形式[23]

$$\mathbf{X}'_{[c,i,j]} = g'\left(\mathbf{X}_{[c,i,j]}\right) \cdot \mathbf{X}_{[c,i,j]}. \tag{4.3}$$

式中，g' 表示一个简单的标量门控函数。例如，对于 ReLU，g' 是指示函数；对于 Swish 激活函数，g' 是 Sigmoid 函数；对于正弦表示网络（Sinusoidal Representation Network, SIREN）单元[201]，g' 则是 Sinc 函数（$\sin(x)/x$）。其他的激活函数也可以被表示成类似的形式，只是相应的 g' 不一定是常见的函数形式。

比较式 (4.2) 和式 (4.3) 可以看出，注意力机制和激活函数都可以被表述成非线性自适应的门控函数。尽管各自有着特定的形式，但两者的主要区别在于激活函数中的门控函数 g' 的输入是一个标量 $\mathbf{X}_{[c,i,j]}$，而注意力机制中的门控函数 $g_{c,i,j}$ 的输入则是整个特征图 $\mathbf{X}$。因此，激活函数可以看作是一个不具有上下文感知功能、输入输出均为标量、被极度简化的注意力模块。考虑到两者之间的这种联系，使用轻量级注意力模块作为激活单元，不仅可以实现在网络中引入非线性的基本功能，还能够逐层地对卷积输出的特征进行动态自适应、上下文感知的特征精炼。

4.2.2 基础注意力激活单元

主流的注意力模块大多被用于网络的中高层[96]，因为此时的特征图已经经过大幅的下采样，复杂的模块设计并不会太过影响网络总体的运行速度。但是，ATAC 模块是被用作整个网络结构中的激活单元，包括网络中下采样倍数很小的初始层，因此 ATAC 单元必须结构简单且计算量小。在此基础上，为了同时满足激活单元要求的局部性以及注意力机制所需的特征上下文，本小节中的基础 ATAC 单元分别采用逐点卷积（PWConv）和逐深度卷积（DWConv）来各自获取相应的跨通道相关性和空间依赖关系，即局部通道注意力模块和局部空间注意力模块。

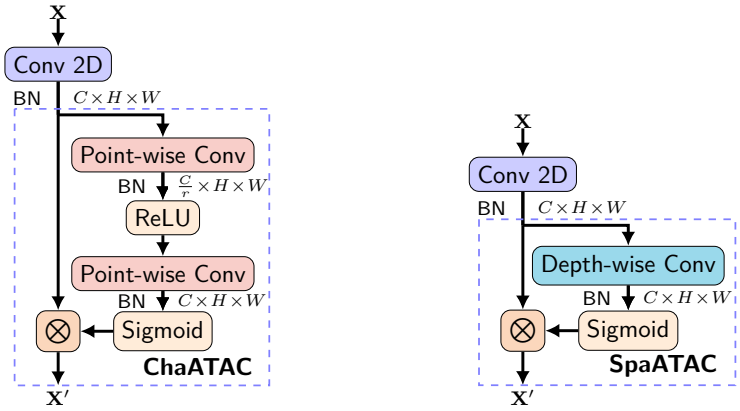
4.2.2.1 局部通道注意力单元

图 4.1(a) 描述了基于局部通道注意力模块的注意力激活单元（Channel Attentional Activation, ChaATAC），旨在使得网络能够逐一在每个像素点上根据跨通道相关性来激活和精炼特征。为了节省参数，给定特征图 $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$，ChaATAC 单元使用瓶颈结构来计算局部通道注意力权重 $\mathbf{L}^{c}(\mathbf{X}) \in \mathbb{R}^{C\times H\times W}$：

$$\mathbf{L}^{c}(\mathbf{X}) = \sigma\left(\mathcal{B}\left(\mathrm{PWConv}_2\left(\delta\left(\mathcal{B}\left(\mathrm{PWConv}_1(\mathbf{X})\right)\right)\right)\right)\right). \tag{4.4}$$

式中，δ 表示 ReLU，$\mathcal{B}$ 表示批标准化[184]（Batch Normalization, BN），r 是通道压缩比例。PWConv_1 和 PWConv_2 的卷积核大小分别为 $\frac{C}{r}\times C\times 1\times 1$ 和 $C\times\frac{C}{r}\times 1\times 1$。从某种程度上，ChaATAC 单元可以被看作是 SENet 模块的局部版本，其与全局的 SENet 模块的不同点在于：1）从结构本身上，ChaATAC 单元移除了 SENet 模块中的 GAP 层，并将 FC 层替换为 PWConv，着重强调局部注意力对于激活单元的重要性，这与主流聚合全局特征上下文的注意力模式不同；2）从设计初衷上，为了强调和保留精细结构，ChaATAC 单元利用局部跨通道上下文，对特征图中的每一个元素进行自适应的特征激活，而 SENet 模块对整个特征图切片施加相同的全局权重。因此 ChaATAC 单元的权重系数具有跟输入特征图一样的大小，即 $C\times H\times W$，而 SENet 模块的权重是一个长度为 C 的向量。3）从用法上，ChaATAC 被用作激活单元，对每一层卷积的输出进行激活和精炼，而 SENet 模块并不承担激活单元的功能，通常被用来对一个残差网络块的输出进行精炼，即 2 至 3 个卷积共享一个 SENet 模块。最后，激活后的特征图 $\mathbf{X}'$ 由 $\mathbf{X}$ 和 $\mathbf{L}^{c}(\mathbf{X})$ 的逐元素点积得到：

$$\mathbf{X}' = \mathbf{L}^{c}(\mathbf{X}) \otimes \mathbf{X}. \tag{4.5}$$



(a) ChaATAC 单元　　(b) SpaATAC 单元

图 4.1　基础注意力激活单元示意图

4.2.2.2 局部空间注意力单元

相对于小目标检测这类任务，类别判别并不是难点，其关键在于对于特征图上目标空间细节的保存。图 4.1(b) 描述了基于局部空间注意力模块（Local Spatial Attention Module）的注意力激活单元（Spatial Attentional Activation, SpaATAC），旨在使得网络能够对空间中特定区域的特征进行加强。SpaATAC 单元采用逐深度卷积在邻域上聚集局部空间信息，产生局部空间注意力图 $\mathbf{L}^s(\mathbf{X}) \in \mathbb{R}^{C\times H\times W}$。激活后的特征图可以被表示为

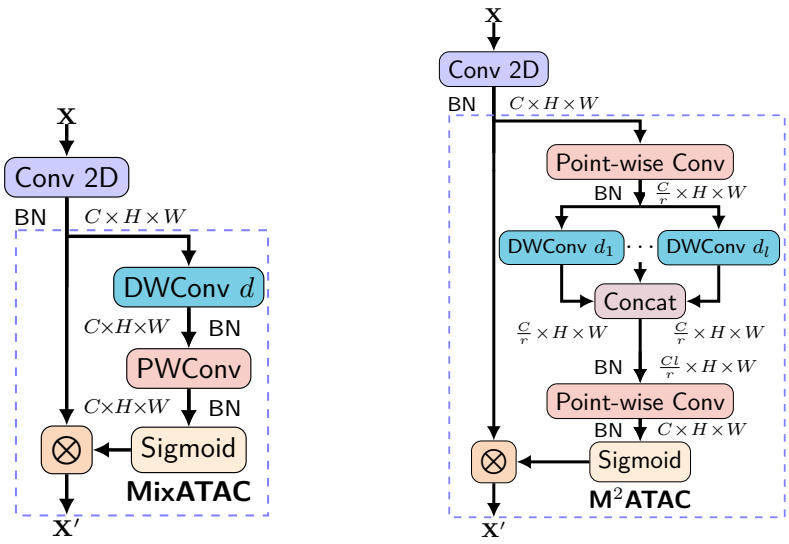
$$\mathbf{X}' = \mathbf{L}^s(\mathbf{X}) \otimes \mathbf{X} = \sigma\left(\mathcal{B}\left(\mathrm{DWConv}\left(\mathbf{X}, d\right)\right)\right) \otimes \mathbf{X}, \tag{4.6}$$

式中，d 是膨胀卷积[202]（Dilated Convolution）的膨胀因子。SpaATAC 单元和 xUnit[23] 的相似之处在于两者均采用逐深度卷积来聚合空间上下文，但两者的不同之处在于：1）从结构本身上，SpaATAC 单元移除了 xUnit 中的 ReLU，并且将 Gaussian 函数替换为 Sigmoid。xUnit 的思路在于采用逐深度卷积来增强 ReLU，而以 SpaATAC 为代表的注意力激活单元放弃了增强 ReLU 的思路，而是将整个注意力模块作为激活单元本身；2）相较于 xUnit 采用 9×9、15×15 这样较大的卷积核，SpaATAC 始终采用膨胀因子为 d 的 3×3 卷积核提取特征，对于抓取较为长程的空间关系（例如 $d = 32$）能够节省大量的网络参数和计算量。

4.2.3 混合注意力激活单元

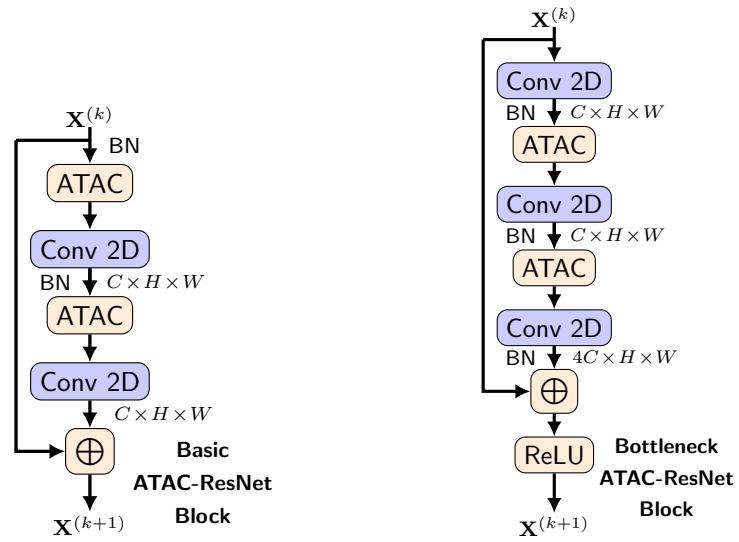
将上述的 SpaATAC 单元和 ChaATAC 单元组合在一起，可以获得能够同时捕获空间注意力和通道注意力的混合注意力激活（Mixed Attentional Activation, MixATAC）单元，如图 4.2(a) 所示。不同于卷积块注意力模块[22]（Convolutional Block Attention Module, CBAM）将空间注意力模块与通道注意力模块串联在一起、瓶颈注意力模块（Bottleneck Attention Module, BAM）将聚合的通道上下文和空间上下文相加，MixATAC 单元的通道上下文是在空间上下文聚合后的特征图上进行。需要注意的是，图 4.2(a) 中的 PWConv 也可以采用如图 4.1(a) 中的瓶颈结构。

使用 SpaATAC 单元和 MixATAC 单元的一个重要问题是如何选择逐深度卷积的膨胀因子 d。不同的数据集、不同的网络架构、网络中不同的层都有可能需要不同的膨胀因子。例如，网络的初始层不一定需要大的感受野，因为它们通常用来抽取较小的局部特征，但网络的中高层则受益于更大的感受野来提取语义特征。为解决该问题，图 4.2(b) 中所展示的多尺度混合注意力激活（Multi-scale Mixed Attentional Activation, M^{2}ATAC）单元采用多个不同感受野的并行分支来聚合多尺度特征上下文信息，以覆盖不同尺度的目标。



(a) MixATAC 单元　　(b) M^{2}ATAC 单元

图 4.2　单尺度、多尺度混合注意力激活单元示意图



(a) 基础 ATAC-ResNet 块　　(b) 瓶颈 ATAC-ResNet 块

图 4.3　两类 ATAC-ResNet 块示意图

4.2.4　全注意力网络

ATAC 单元提供了一种构建全注意力网络的方式，即将深度网络中的每个 ReLU 都替换为 ATAC 单元。本章选取残差网络（Residual Network, ResNet）作为 ATAC 单元的宿主网络，图 4.3 展示了采用 ATAC 单元的基础 ResNet 块和瓶颈 ResNet 块。如果忽略 BN 层，相比于每一层的

3×3 卷积，ChaATAC 单元大约会产生 $\frac{2}{9r}$ 的额外参数量和计算量（r 通常为 4），而 SpaATAC 单元所消耗的额外参数量更少，仅为 $\frac{1}{C}$。虽然会增加一定的额外参数和计算开销，但 ATAC 单元使得在网络初期提早优化特征成为可能，甚至是网络的第一个卷积之后。由于在网络早期阶段便已经开始抑制不相关的低层特征、加强任务相关的目标特征，网络可以更高效地编码高层语义。相比于一味增加网络深度，在 ATAC 单元中花费这些额外的内存和计算资源更为值得。

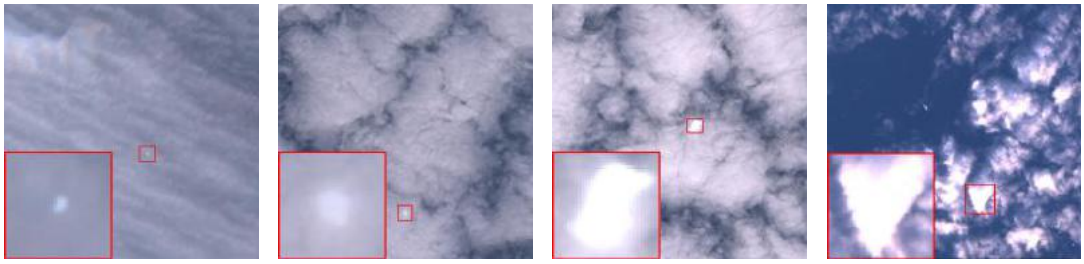
4.3 更大规模小目标数据集的构建

纯粹数据驱动的深度学习方法利用标记数据以端到端的方式学习输入与输出之间的映射关系，需要大量高质量的标记数据。然而，红外小目标数据的稀缺性决定了单帧检测数据集的规模较小，很大程度上限制了深度学习模型的潜力。此外，红外小目标的数据分布与 ImageNet 等大规模通用数据集中的物体数据分布差异过大，难以像部分视觉任务那样利用预训练模型实现模型迁移。事实上，在可见光遥感影像中，也存在诸如冰山之类的地物具有与红外小目标类似的特性，即本征特征不足、存在强起伏云背景干扰等，而且可见光遥感影像获取更为开放，更加易于构建较大规模的数据集。

为了能够在更大规模的数据集上验证小目标检测算法，同时也是为了通过迁移学习提高红外小目标检测性能创造条件，本节基于 Sentinel-2 卫星数据[203] 构建了一个名为 DiskoBay 的弱小冰山检测数据集。该数据集收集了 2017、2018、2019 三年来五月至十月期间格陵兰岛西海岸 Disko 湾区 1006 幅包含弱小冰山的可见光遥感图像，一共包含 2818 个冰山实例，其中 50% 的图像被选作训练集，20% 作为验证集，30% 测试集，而且每幅图像均由人工进行像素级的标注。此外，为了使得 DiskoBay 数据集中的目标特性与红外弱小目标更接近，同时也是为了加大算法检测的难度，DiskoBay 数据集主要选取了云含量较大的图像，以检验算法在复杂背景下的性能。

图 4.4 中展示了 DiskoBay 数据集的一些代表性图像，从中可以看出：1）类似于红外弱小目标，一些冰山本身相当昏暗，且往往被淹没在半透明的云层之下，对比度很小；2）受到传感器有限的分辨率以及冰山自身分布特性的影响，在 DiskoBay 数据集中，大小在 2×2 至 10×10 个像素之间的冰山占据了大部分，这些目标缺少相应的纹理和形状特征。表 4.2 展示了 DiskoBay 数据集目标尺度具体的分布情况，其中尺寸小于 12×12 像素的目标占据了目标总数的 83.5%，尺度变化范围与红外小目标基本相同。3）某些特定类型的云在光谱特性和外观上都类似于冰山，这些与目标相似的背景干扰会增加检测算法的虚警率。例如，对比图 4.4(c) 和图 4.4(d)，如果仅仅观察两者左下角的放大区域，较难区分出真实的目标和背景的云层。

此外，本章还利用 Microsoft COCO 数据集[1] 中的停车标志子类构建了一个 StopSign 数据集，用于进一步验证检测算法对于目标尺度变化的应对能力。该数据集的训练集含有 1734 幅图像、1983 个实例，验证集含有 69 幅图像、75 个实例。图 4.5 中展示了 StopSign 数据集的一些



(a) 极小尺寸　(b) 低对比度　(c) 大尺寸　(d) 相似的背景干扰

图 4.4　DiskoBay 数据集的部分代表性图像

表 4.2　DiskoBay 数据集目标尺度分布

冰山类型	长度	像素数	百分比	累计百分比
碎冰山	5 –15 米	1×1 至 2×2	1.7%	1.7%
小型冰山	15 –60 米	2×2 至 6×6	51.1%	52.8%
中型冰山	61 –120 米	6×6 至 12×12	30.7%	83.5%
大型冰山	121 –200 米	12×12 至 20×20	11.4%	94.9%
甚大型冰山	大于 200 米	大于 20×20	5.1%	100%

代表性图像，从中可以看出，1）与红外小目标类似，由于拍摄距离较远，相当多的目标占据图像的比例非常小，符合小目标的特点；2）大部分停车标志位于复杂的图像背景中，并不具有显著性，而且背景中存在着若干颜色、形状相近的干扰；3）不同于红外小目标和冰山的尺寸大多位于 2×2 至 10×10 个像素之间，停车标志的尺度变化更大。

4.4　实验结果与分析

为了验证 ATAC 单元结构设计上的合理性和有效性，本节将通过详细的消融实验与对比实验来具体探索以下问题：

（1）问题一：通常，计算机视觉领域的注意力机制往往致力于捕获长程乃至全局的特征上下文以解决语义判别上的模糊性，但是本章却使用局部注意力模块作为激活单元。4.4.2 小节将验证在 ATAC 单元中采用局部注意力机制的必要性，即在给定相同参数数量的情况下，使用局部注意力模块作为激活单元的网络能否取得比使用全局注意力激活单元的网络更好的性能。

（2）问题二：在深度学习领域中，众多研究已经证明在网络中嵌入参数量和计算量相对较小的微模块可以增强网络的判别性能[21,182]。网络中的网络[182]（Network in Network, NiN）模块、SENet 模块[21] 以及本章的 ATAC 单元都属于该改进方式的具体体现。4.4.2 小节还将研究这三者之中究竟哪种方式更加高效，即在给定相同参数数量的情况下，相比 NiN 模块和 SENet 模块，ATAC 单元能否为网络带来更多的性能提升。

（3）问题三：将部分 ReLU 替换为 ATAC 单元在提升深度网络性能的同时，也会增加网

图 4.5　StopSign 数据集的部分代表性图像

络的参数数量和计算开销。从节省网络参数和计算量的角度，关键在于随着越来越多的 ReLU 被 ATAC 单元代替，这种性能提升是否具有持续性，即是否存在特定的性能饱和点，当网络中 ATAC 单元的比例超过一定量后，其性能便不再提升。为此，4.4.2 小节还将对构建全注意力网络的必要性展开讨论。

（4）问题四：4.4.3 小节将研究相比于采用其他激活函数的网络以及使用其他注意力机制的网络，采用 ATAC 单元作为激活单元的网络是否能够获得更好的性能。

4.4.1　实验设置

考虑到不同计算机视觉任务对于通道、空间特征上下文侧重性的不同，为了有针对性地评估各类注意力激活单元，本节采用 CIFAR-10、CIFAR-100、ImageNet 等图像分类数据集验证 ChaATAC 单元，而在 DiskoBay 和 StopSign 数据集上对 SpaATAC、MixATAC、M^2ATAC 单元进行语义分割性能的评估。其中，CIFAR-10 数据集包含 10 个类别的 60000 幅 32×32 大小的彩色图像，每个类别 6000 幅，共分为 50000 张训练图像和 10000 张测试图像。CIFAR-100 数据集与之类似，差别在于其包含了 20 个超类，可以被进一步细分为 100 个子类别，每个子类别 600 幅图像，其中 500 幅训练图像和 100 幅测试图像。需要注意的是，在 MXNet / Gluon 框架中，默认的 CIFAR-100 数据集数据集设置为仅采用 20 个超类，本文中的所有剥离实验均在此基础上进行。ImageNet 数据集则包含了 128 万幅训练图像和 5 万幅测试图像，一共 1000 类[2]。

对于图像分类实验，分别选用 ResNet-20 和 ResNet-50 作为 ChaATAC 单元在 CIFAR-10/100 数据集和 ImageNet 数据集上的宿主网络。具体的网络架构如表 4.3 所示，其中 ResNet-20 结构中的 b 代表网络每个阶段的残差块数目，$b=3$ 时便是标准的 ResNet-20。通过变动 b 的数值，

表 4.3　用于图像分类实验的骨干网络结构

网络阶段	输出大小	ResNet-20	输出大小	ResNet-50
Conv-1	32×32	3× 3 conv, 16	112×112	7× 7 conv, 64
Stage-1	32×32	$\begin{bmatrix} 3\times3 \text{ conv, } 16 \\ 3\times3 \text{ conv, } 16 \end{bmatrix} \times b$	112×112	$\begin{bmatrix} 3\times3 \text{ conv, } 64 \\ 3\times3 \text{ conv, } 64 \\ 3\times3 \text{ conv, } 256 \end{bmatrix} \times 3$
Stage-2	16×16	$\begin{bmatrix} 3\times3 \text{ conv, } 32 \\ 3\times3 \text{ conv, } 32 \end{bmatrix} \times b$	56×56	$\begin{bmatrix} 3\times3 \text{ conv, } 128 \\ 3\times3 \text{ conv, } 128 \\ 3\times3 \text{ conv, } 512 \end{bmatrix} \times 4$
Stage-3	8×8	$\begin{bmatrix} 3\times3 \text{ conv, } 64 \\ 3\times3 \text{ conv, } 64 \end{bmatrix} \times b$	28×28	$\begin{bmatrix} 3\times3 \text{ conv, } 256 \\ 3\times3 \text{ conv, } 256 \\ 3\times3 \text{ conv, } 1024 \end{bmatrix} \times 6$
Stage-4			14×14	$\begin{bmatrix} 3\times3 \text{ conv, } 512 \\ 3\times3 \text{ conv, } 512 \\ 3\times3 \text{ conv, } 2048 \end{bmatrix} \times 3$
	1×1	GAP, FC, Softmax		

可以对网络进行伸缩，从而观察网络在不同深度下的性能表现。通道压缩比例 r 默认为 2。对于 CIFAR-10/100 数据集，网络训练过程采用 Kaiming 方法[197] 进行初始化，交叉熵作为损失函数，Nesterov 加速梯度（Nesterov Accelerated Gradient, NAG）方法进行优化，权重的衰减率为 0.0001，批大小（Batchsize）为 128，一共训练 400 轮次，初始学习率为 0.2 且在第 300 轮和 350 轮后各除以 10。数据增强（Data Augmentation）环节遵循标准做法，即以长度 4 对图像各边补零填充后，再对随机裁剪为 32×32 大小的图像进行随机水平翻转。对于 ImageNet 数据集，ChaATAC-ResNet-50 将 Stage-3/4 两阶段中的残差块替换为瓶颈 ChaATAC 残差块，学习率为 0.075 并按照余弦曲线衰减，一共训练 160 轮，其余的优化参数设置与 CIFAR-10/100 数据集相同。

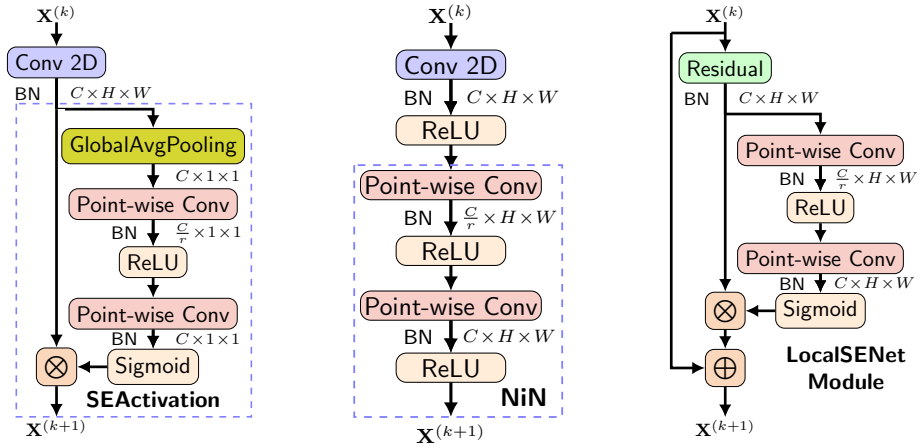
对于 DiskoBay 和 StopSign 数据集上的语义分割实验，选用基础上下文网络[202]（Context Network, ContextNet）作为宿主网络，其骨干网络架构如表 4.4 所示。DiskoBay 数据集采用交并比（Intersection over Union, IoU）作为评价指标，StopSign 数据集则将背景分割的准确率计算在内，采用平均 IoU（mean IoU, mIoU）作为评价指标。网络采用 Soft-IoU[190] 作为训练过程的损失函数，使用 AdaGrad 作为优化方法[191]，学习率为 0.1，权重衰减率为 0.0001，批大小为 10，一共训练 200 轮。除了随机裁剪的大小改为 480×480 之外，其余数据增强环节与上述分类实验设置相同。

表 4.4　用于小目标语义分割实验的骨干网络结构

层数 b	1	2	3	4	5	6
卷积大小	3×3	3×3	3×3	3×3	3×3	3×3
膨胀因子	1	1	2	4	8	16
感受野大小	3×3	5×5	9×9	17×17	33×33	65×65
输出通道数	C	C	C	C	C	C

4.4.2　消融实验与分析

为了更好地理解本章所提出的 ATAC 单元，本小节通过移除或者替换各个 ATAC 单元中的特定组件构建新的、用于消融实验的对比模块，从而在给定相同网络参数数量的情况下，验证本章所设计的注意力激活模块的有效性。以 ChaATAC 单元为例，图 4.6 展示了在其基础上构建的用于控制变量实验的模块。对于其他的 ATAC 单元而言，可以通过类似的方式构建出各自的消融实验模块。



(a) SEActivation　　(b) ChaNiN　　(c) LocalSENet

图 4.6　用于 ChaATAC 单元消融实验的模块结构示意图

4.4.2.1　特征上下文聚合尺度的重要性

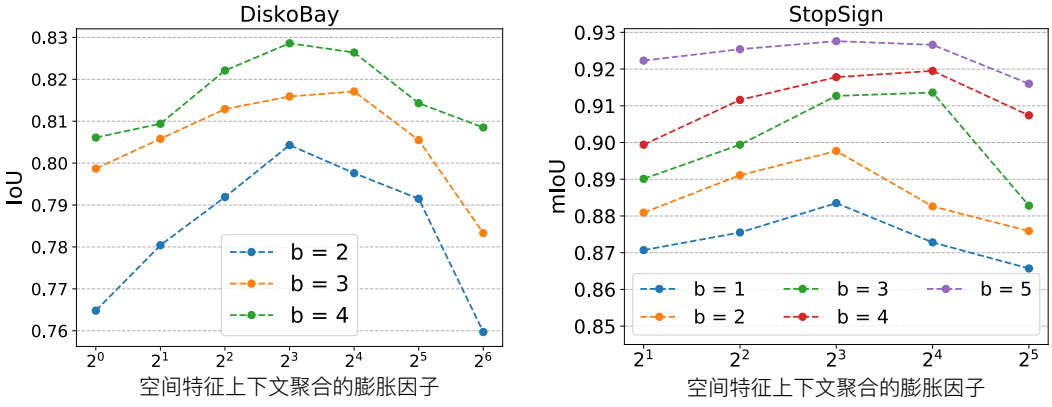
为了验证注意力激活单元使用局部注意力机制的必要性，首先针对通道注意力机制，将 ChaATAC 单元与图 4.6(a) 所示的 SEActivation 单元进行了比较。该单元使用了 SENet[21] 中的全局通道注意力模块，但是将该模块的用途从特征精炼改为了网络中的激活单元，用于代替 ReLU。相比 ChaATAC 单元，SEActivation 单元具有全局平均池化层，其输出的注意力权重大小为 $C\times1\times1$。因此，尽管具有相同的参数数量，但两者上下文聚合的尺度和注意力权重的应用范围不同。SEActivation 聚合了全局上下文，使得每个大小为 $H\times W$ 的特征图切片共享同一个

注意力权重。相比之下，ChaATAC 单元以逐点的方式抓取通道之间的关系，使得特征图中的每个元素都具有单独的权重，其权重张量大小为 $C \times H \times W$。表 4.5 给出了两者在网络深度逐渐增加的情况下在 CIFAR-10 和 CIFAR-100 数据集上的分类性能比较。从中可以看出，与 ChaATAC 单元相比，使用 SEActivation 单元的网络性能明显更差，这表明对于通道注意力激活单元，聚合上下文的局部性至关重要。

表 4.5　ChaATAC 单元与 SEActivation 单元的分类性能比较

激活单元	CIFAR-10				CIFAR-100			
	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
SEActivation	0.548	0.601	0.613	0.622	0.388	0.432	0.452	0.456
ChaATAC	**0.906**	**0.927**	**0.936**	**0.939**	**0.764**	**0.796**	**0.812**	**0.821**

图 4.7 展示了不同大小的膨胀因子下使用 SpaATAC 单元作为激活单元的基础 ContextNet 在 DiskoBay 和 StopSign 这两个小目标数据集上的语义分割性能。从中可以看出，对于采用空间注意力激活单元的网络来说，其预测性能与特征上下文聚合的尺度高度相关，尺度过大或者过小均会严重影响网络的性能。当尺度过大时，注意力模块所聚合的特征上下文大多为与当前目标无关的背景信息，其特征重加权过程相当于依据噪声调制原信号，使得网络性能大幅下降。当尺度过小时，注意力模块由于无法收集到足够的上下文信息，其性能将退化为如 Swish 激活单元这样无上下文感知的标量函数。然而，对于不同数据集上不同深度的网络而言，最佳的膨胀因子并不唯一，单一的聚合尺度很难捕获大小不一的目标特征。

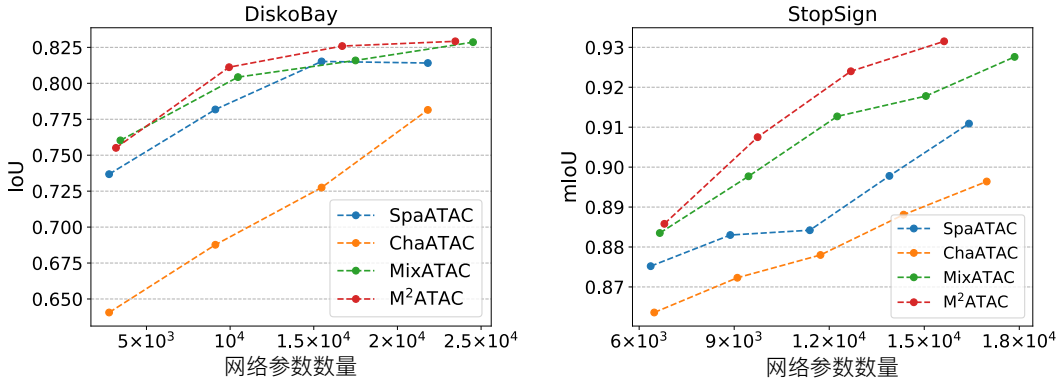


(a) DiskoBay　　(b) StopSign

图 4.7　不同膨胀因子下 SpaATAC 单元的语义分割性能比较

图 4.8 展示了在 DiskoBay 和 StopSign 数据集上，使用本章所构建的四种注意力激活单元的

网络的语义分割性能比较。从中可以看到：1）使用 ChaATAC 单元的网络性能最差，这是由于其只依据局部通道上下文对特征进行激活，没有聚合有利于消解目标语义模糊性的空间上下文信息。2）尽管如此，SpaATAC 单元与 MixATAC 单元的对比结果表明，在 DiskoBay 和 StopSign 数据集这样的二分类任务上，结合空间注意力与通道注意力能够进一步提高网络检测小目标的性能。虽然这种结合增加了单个注意力激活单元的参数数量，但其给予网络的性能增益使得网络可以在性能不变的情况下使用更少的层数，从而提升网络整体的效率。例如，在图 4.8(b) 中，$b = 3$ 时 MixATAC 网络的参数数量远少于 $b = 5$ 时的 SpaATAC 网络，但其 mIoU 指标却略好于后者。3）MixATAC 单元与 M^{2}ATAC 单元两者的对比结果表明，多尺度的空间特征上下文聚合可以进一步提升空间注意力激活单元的性能，而且提升效果与数据集中目标自身的尺度变化情况有关。DiskoBay 数据集的目标尺度较为单一，因此 MixATAC 单元与 M^{2}ATAC 单元两者之间的差异较小。StopSign 数据集的目标尺度变化大，相应的 M^{2}ATAC 单元的性能也明显好于 MixATAC 单元。



(a) DiskoBay　　(b) StopSign

图 4.8　四种注意力激活单元的小目标语义分割性能比较

4.4.2.2　微模块结构的选取

图 4.6(b) 和图 4.6(c) 分别展示了用于与 ChaATAC 单元对比的两个微模块结构，即 NiN 模块和 LocalSE 模块。其中，NiN 模块是在卷积层后引入逐点卷积来增强网络的判别能力，实质上是增加了网络的深度。与 ChaATAC 单元一样，LocalSE 模块也是去除了原始 SENet 模块中的全局平均池化，并将全连接层替换为逐点卷积。不同的是，ChaATAC 单元是作为网络中的激活函数被使用，而 LocalSE 模块仍然保持了 SENet 模块原始的用途，即精炼特征，并不被用来代替 ReLU。由于在 SENet 模块中，两个卷积层共享一个注意力模块，为了保持网络参数数量相同，将 LocalSE 模块的通道压缩比例 r 设置为 1。表 4.6 展示了三者在 CIFAR-10 和 CIFAR-100 数据集上的分类准确率比较，从中可以看出：1）NiN 模块的性能不如 LocalSE 模块和 ChaATAC

单元，这表明给定少量的额外参数和计算量，相比增加网络深度的 NiN 模块，通过注意力机制精炼特征是一种更为高效的提升网络性能的方式。2）LocalSE 模块和 ChaATAC 单元两者的差别在于对于每个残差块，LocalSE 模块只会进行一次特征精炼，而 ChaATAC 单元则是会对每个卷积层的输出都进行特征精炼。ChaATAC 单元的性能好于 LocalSE 模块，这表明在给定相同参数数量和计算量的情况下，对于注意力模块的使用方式，轻量级但次数更多的注意力激活方式好于更为复杂但次数较少的特征精炼方式。

表 4.6　ChaATAC 单元与其他两种微模块结构的分类性能比较

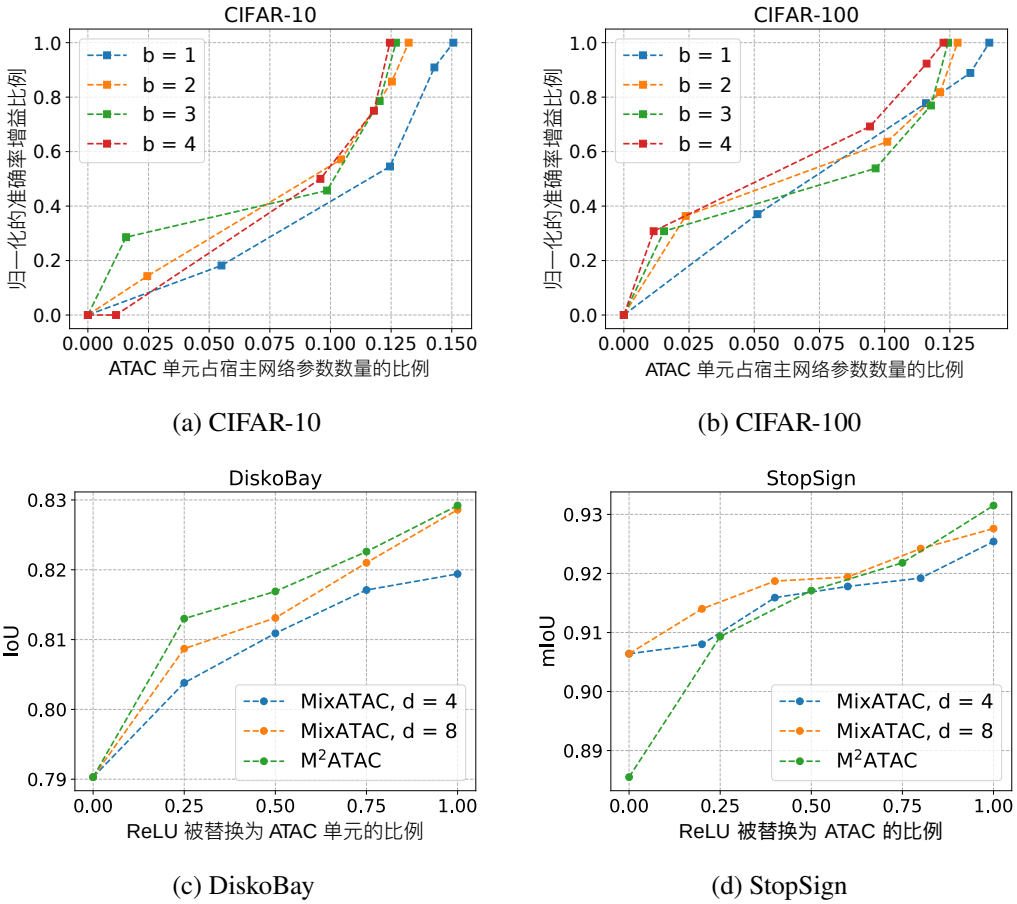
微模块类型	CIFAR-10				CIFAR-100			
	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
NiN	0.893	0.917	0.922	0.926	0.743	0.776	0.792	0.796
LocalSE	0.906	0.926	0.931	0.937	0.762	0.794	0.805	0.811
ChaATAC	**0.906**	**0.927**	**0.936**	**0.939**	**0.764**	**0.796**	**0.812**	**0.821**

4.4.2.3　全注意力网络的必要性

图 4.9 展示了从网络的最后一层开始逐渐将 ReLU 替换为 ATAC 单元时，网络性能相应的变化情况。横轴的起点代表完全不使用 ATAC 单元时的网络，横轴的终点表示所有 ReLU 均被替换成 ATAC 单元时的网络，即全注意力网络。图 4.9(a) 和图 4.9(b) 展示了在 CIFAR-10 和 CIFAR-100 数据集上，随着 ATAC 单元比例的上升，不同深度的 ResNet 各自分类准确率的增长情况。由于 ResNet 不同层的网络通道数不同，在不同层中将 ReLU 替换为 ATAC 单元所增加的网络参数数量也各不相同。为了更忠实地反映 ATAC 单元带来的额外开销与性能增益之间的关系，CIFAR-10/100 两图使用网络参数数量作为横坐标。此外，出于将不同深度的网络纳入同一张图的目的，CIFAR-10/100 两图采用归一化的性能增益作为纵坐标。从中可以看出，对于被用于 CIFAR-10/100 数据集上的 ResNet-20 来说，并不存在相应的饱和点。随着 ATAC 单元比例的增加，网络的性能在持续性地提升，直至变成全注意力网络。值得注意的是，增长折线的最右侧，其斜率在不少折线中最大，这表明在给定相同的网络参数下，在网络最底层将其中的 ReLU 替换为 ATAC 单元可以带来最多的性能提升。这一定程度上印证了本章的动机，即网络早期阶段的注意力调制通过抑制无关的低层特征、突出相关特征，可以使得网络能够更为高效地编码图像的高层语义信息。

图 4.9(c) 和图 4.9(d) 分别展示了在 DiskoBay 和 StopSign 数据集上，$b=4$ 时和 $b=5$ 时的网络随着 MixATAC 单元和 M^2ATAC 单元比例增加的性能增长情况。由于网络深度相同，小目标分割的性能相近，图 4.9(c) 和图 4.9(d) 的纵轴反映的是各个网络检测性能的绝对值，无需对其进行归一化。从中可以看出，与 ChaATAC 单元相同，深度网络的小目标分割性能随着 MixATAC

单元和 M^{2}ATAC 单元比例的增长而升高，且并未出现性能饱和的现象。由此可知，对于特定的网络，通过将其中的 ReLU 替换为相应的 ATAC 单元构建全注意力网络能够有效提升网络的性能。



(a) CIFAR-10　　(b) CIFAR-100

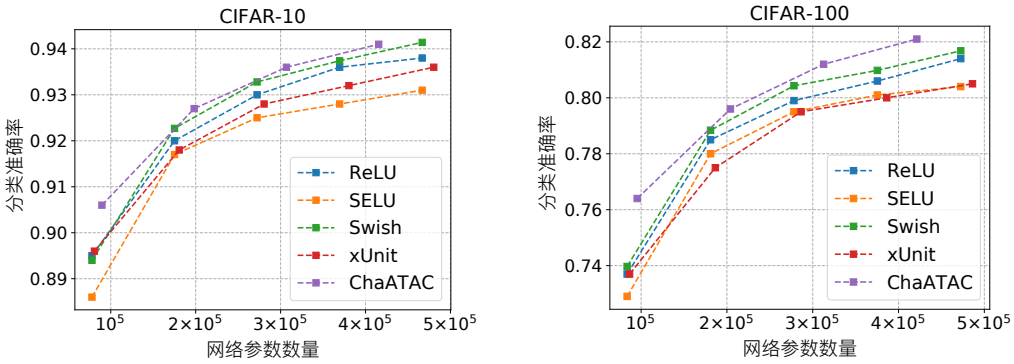
(c) DiskoBay　　(d) StopSign

图 4.9　网络性能增益与 ATAC 单元比例之间的关系图

4.4.3　方法对比与分析

图 4.10(a) 和图 4.10(b) 展示了 ChaATAC 单元在网络深度逐渐增加的情况下，与 ReLU[183]、SELU[192]、Swish[193]、xUnit[23] 等其他激活单元在 CIFAR-10 和 CIFAR-100 数据集上的性能比较。此外，该实验还对 GELU[196] 和 PReLU[197] 进行了训练和测试，但由于这两者的性能不如上述激活单元，所以没有被列入比较。从中可以看出：1）ChaATAC 单元在所有实验设置下均取得了最佳的性能，证明了其作为激活单元的有效性。此外，同样为非线性门控函数形式的 Swish 单元，其性能仅次于 ChaATAC 单元，好于其余 ReLU 类的激活单元，这表明发展非线性门控函数形式的激活单元具有良好的前景。2）Swish 单元可以看作是 ChaATAC 单元非上下文感知的标量版本，而图中 ChaATAC 单元的性能稳定地好于 Swish 单元，这表明局部的特征上下文聚

合能够有效提升激活单元的性能。3）使用 ChaATAC 单元代替 ReLU 可以提升网络的效率，即在相同性能的情况下，使用 ChaATAC 单元的网络只需要更少的层数和网络参数。例如，在图 4.10(b) 中，$b = 3$ 时的 ChaATAC-ResNet 便可以达到与 $b = 5$ 时的 ResNet 相同的分类精度，但只需要后者大约 65% 的参数。



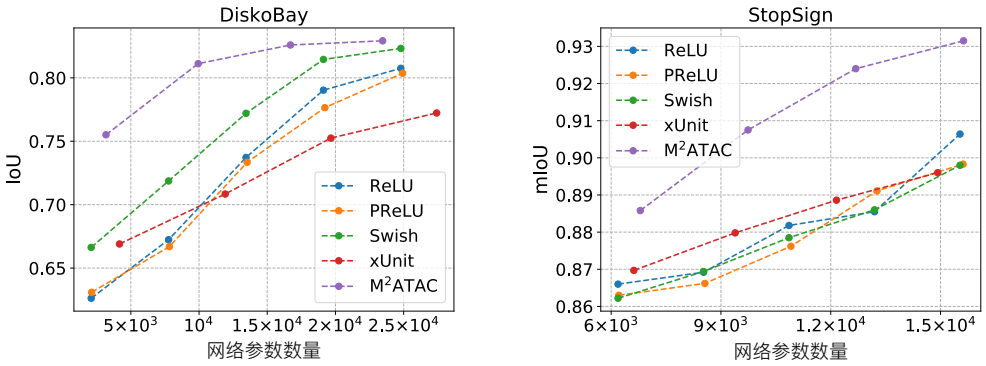
(a) CIFAR-10　　(b) CIFAR-100

图 4.10　多种激活单元在 CIFAR-10 和 CIFAR-100 数据集上的分类性能比较

图 4.11(a) 和图 4.11(b) 展示了 M^{2}ATAC 单元与 ReLU、PReLU、Swish、xUnit 等其他激活单元在 DiskoBay 和 StopSign 数据集上的对比结果。从中可以看出：1）M^{2}ATAC 单元在两个数据集的所有实验设置下均表现最好，这表明了本章所提出的 M^{2}ATAC 单元的有效性。需要注意的是，$b = 2$ 时使用 M^{2}ATAC 单元的网络性能便可以略好于 $b = 5$ 时使用 ReLU 的网络，这表明在激活单元中嵌入相应的空间特征上下文和通道特征上下文是一条构建更高效网络的可行途径。2）xUnit 在 DiskoBay 数据集上表现相对最差，而在 StopSign 数据集上，其表现仅次于 M^{2}ATAC 单元，好于包括 Swish 单元在内的其余激活单元。考虑到 StopSign 数据集尺度变化大于 DiskoBay 数据集，xUnit 在两个数据集上较大的性能差异表明，目标尺度与空间上下文聚合尺度之间的匹配程度对网络最终的性能有较大的影响。

图 4.12 展示了在 StopSign 数据集上，基础 ContextNet 第四层中 ReLU、xUnit、M^{2}ATAC 单元各自的输入特征图、激活权重图、输出特征图。其真实目标为图像顶部中间区域的圆形停车标志。从中可以看出，M^{2}ATAC 单元的特征图中背景特征的残留最少，表示其能够更好地抑制无关背景、突出目标特征。在非线性门控函数的框架下，ReLU 生成激活权重的函数是二值的指示函数[23]，对于数值在零附近的特征变化较为敏感。从图 4.12(b) 可见，在 ReLU 的激活权重图中，大量背景区域与目标区域被赋予了相同的权重。相比 xUnit，在 M^{2}ATAC 单元的激活权重图中，被激活的区域大多都围绕在真实的目标周围，这表明 M^{2}ATAC 单元的确学习到了与真实目标相关的语义信息，并且能够对目标进行选择性的激活。

图 4.13 展示了在网络逐渐加深的情况下，ChaATAC-ResNet 与 ResNet[185]、SENet[21]、GENet[172] 等网络在 CIFAR-10 和 CIFAR-100 数据集上分类准确率、参数数量、计算量的比较情况。其中



(a) DiskoBay　　(b) StopSign

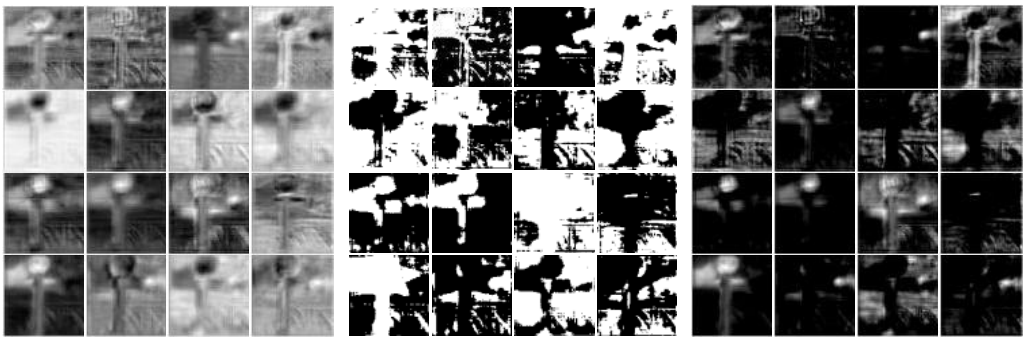
图 4.11　多种激活单元在 DiskoBay 和 StopSign 数据集上的分割性能比较

参数数量列的 M 代表百万（Million），计算量的单位为 10 亿次浮点运算（Giga Floating Point Operations, GFLOPs）。从中可以看出，ChaATAC-ResNet 在所有实验设置下都取得了最好的性能，这表明了 ChaATAC 单元逐层且同时进行特征激活和精炼的方式的有效性。此外，表 4.7 展示了给定 ResNet-50 作为宿主网络的情况下，ChaATAC-ResNet 与 SENet[21]、注意力增强（Attention Augmented, AA）的卷积网络、全注意力（Fully Attentional, FA）视觉模型[66]、GENet[172] 在 ImageNet 数据集上的比较结果。其中，ChaATAC-ResNet-50 取得了最佳的 top-1 错误率。尽管由于采用聚合局部上下文的逐点卷积，ChaATAC-ResNet-50 的 GFlops 多于 SE-ResNet-50 和 GE-θ^+-ResNet-50，但是就网络参数而言，ChaATAC-ResNet-50 则少于这两者。

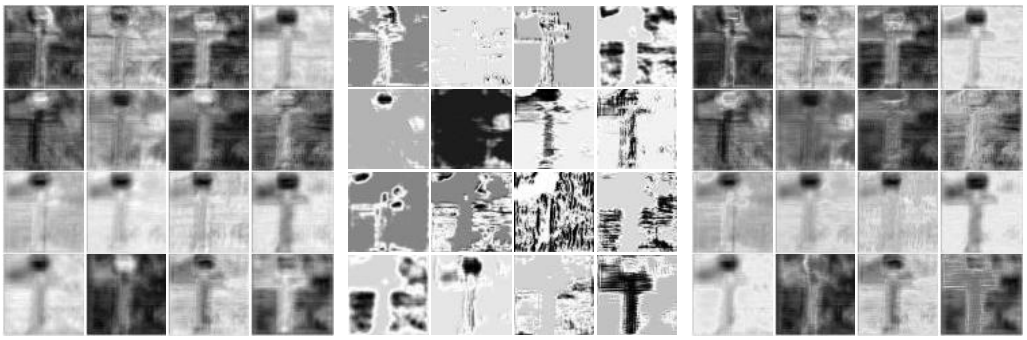
表 4.7　ChaATAC-ResNet-50 与其他深度网络在 ImageNet 数据集上的分类准确率比较

方法	GFLOPs	参数数	top-1 错误率 / %	top-5 错误率 / %
ResNet-50[204]	3.86	25.6M	23.30	6.55
SE-ResNet-50[21]	3.87	28.1M	22.12	5.99
AA-ResNet-50[95]	8.3	25.8M	22.30	6.20
FA-ResNet-50[66]	7.2	18.0M	22.40	/
GE-θ^+-ResNet-50[172]	3.87	33.7M	21.88	5.80
ChaATAC-ResNet-50	4.4	28.0M	**21.41**	6.02

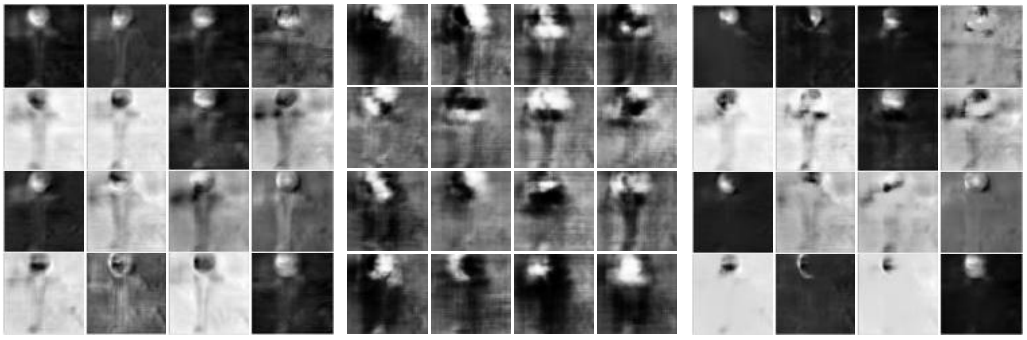
最后，ChaATAC-ResNet 在 CIFAR-10/100 数据集（当 $b = 5$ 时，事实上是 ResNet-32）和 ImageNet 数据集（ResNet-50）上的性能表现说明，使用 ATAC 单元的深度网络并不会遭受梯度消失问题。其主要原因在于网络中的批归一化层将每层神经元的输入分布重新规范化为标准正态分布，使得激活函数的输入处于非线性函数中梯度较大的区域，从而避免了梯度消失问题的产生[184]。另外，ResNet 中的残差连接让网络具有恒等映射的能力，也保证了网络性能不会随着网络深度的加深而产生退化[204]。与 ATAC 单元一样，Swish 激活单元也采用了 Sigmoid 函



(a) ReLU 输入特征图　(b) ReLU 激活权重图　(c) ReLU 输出特征图



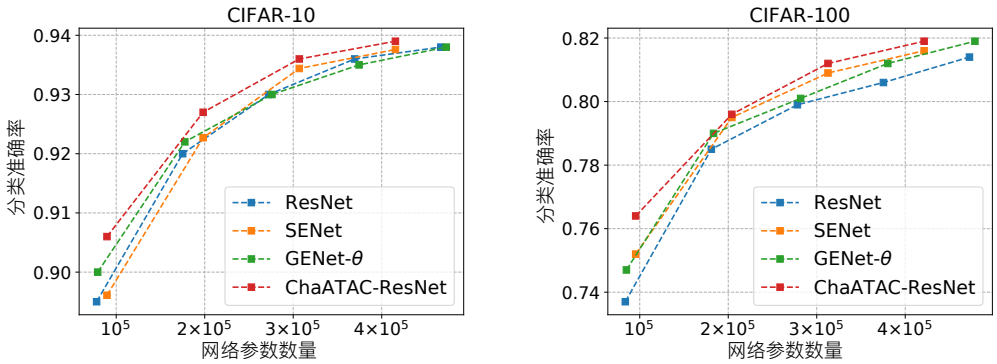
(d) xUnit 输入特征图　(e) xUnit 激活权重图　(f) xUnit 输出特征图



(g) M^{2}ATAC 输入特征图　(h) M^{2}ATAC 激活权重图　(i) M^{2}ATAC 输出特征图

图 4.12　特征图可视化结果比较

数，而使用 Swish 激活单元可以构建非常深的网络并取得良好的性能[193]。事实上，在 ATAC 单元中，使用 Sigmoid 函数的目的并非对特征进行非线性激活，而在于获得注意力权重的概率表示，这也是注意力机制、深度信念网络（Deep Belief Network, DBN）、循环神经网络（Recurrent Neural Network, RNN）、长短期记忆（Long Short-Term Memory, LSTM）网络等架构会普遍采用逐层的 Sigmoid 或 Softmax 函数的原因。



(a) CIFAR-10 数据集　　(b) CIFAR-100 数据集

图 4.13　ChaATAC-ResNet 与其他深度网络在 CIFAR-10/100 数据集上的分类准确率比较

4.5　本章小结

基于激活单元与注意力机制之间的相似性，本章设计了一类动态且具有特征上下文感知特性的注意力激活单元。除了在网络中引入非线性之外，通过聚合相应的通道、空间特征上下文信息，注意力激活单元还可以对特征进行有选择性的激活，从而抑制无关特征、强调相关特征，使得网络能够更为高效地编码高层语义。此外，通过替换网络中原有的激活单元，注意力激活单元还提供了一种构建全注意力网络的途径，可以以逐层的方式对早期卷积层在内的所有卷积层所提取的特征进行精炼。最后，本章还构建了 DiskoBay 和 StopSign 两个更大规模的数据集用于验证小目标检测算法的性能。在图像分类和小目标分割任务上的大量消融实验和对比实验验证了本章所提出的注意力激活单元的有效性。

第五章　基于注意力特征融合的图像分类与小目标分割

上一章展示了对于激活函数这一类普遍存在但又形式简单的网络结构单元，将其替换为轻量级的注意力模块可以逐层地依据特征上下文激活之前卷积层聚合的特征，从而提升网络的效率和性能。除了卷积之外，跳层连接也是许多现代网络中特征聚合的标准模块。然而目前其相关工作大多致力于在各种网络架构中引入不同形式的跳层连接，但是对于被连接特征之间的融合方式本身，仍然采用相加或拼接这种简单的方式，无法动态选择与当前任务最相关的特征，不利于后续的目标检测、语义分割等任务。

为此，本章提出了一个根据特征上下文动态分配权重的通用特征融合框架——注意力特征融合，致力于统一跨层的短跳（Short Skip）连接、长跳（Long Skip）连接和同层的 Inception 模块等多种特征融合场景。该框架构建了一个多尺度通道注意力模块（Multi-Scale Channel Attention Module, MS-CAM），旨在通过聚合不同尺度的特征上下文，满足不同尺度的物体对于不同大小的感受野的要求，从而更好地融合不同场景下语义和尺度各不相同的特征。此外，该框架还支持以迭代的方式持续优化注意力模块的输入特征，即采用另一层注意力特征融合的输出作为下一层融合权重生成模块的输入，从而进一步提升特征融合的性能。对于一个给定的网络，通过替换其中原有的特征融合模块，可以构建出基于注意力特征融合的新网络。大量消融实验和对比实验表明本章提出的框架具有良好的特征融合能力，能够显著提高不同网络在图像分类、目标分割等多个任务上的性能。出于可重复研究的考虑，本章方法的代码、训练好的模型参数和训练日志可以从项目主页上获取[1]。

本章的具体内容安排如下：第 5.1 节对于目前深度网络中特征融合的相关工作进行了分析和讨论，并且阐述了本章的研究动机与意义。第 5.2 节介绍了本章提出的多尺度通道注意力模块。在此基础上，第 5.3 节介绍了统一各个场景的注意力特征融合框架以及迭代注意力特征融合，并给出了相应的网络示例。第 5.4 节进行了详细的消融实验与对比实验，以验证注意力特征融合的通用性和有效性。第 5.5 节对本章工作的内容进行了小结。

5.1　引言

近年来，深度学习在图像分类、目标检测、语义分割等计算机视觉任务上取得了长足的进步，其中的关键之一在于各种新式的特征融合结构[104,105,185,204,205]。卷积网络从输入到输出，通常会经历若干个下采样层，通过逐步扩大感受野以获取高层语义特征。网络中的低层特征空间细节信息丰富，但是语义信息不足，对目标类别的判别能力较弱；高层特征语义信息丰富，但

[1]https://github.com/YimianDai/open-aff

是分辨率较低，不利于目标的精细定位与分割。为了解决分辨率与语义之间的矛盾，现代网络大多致力于通过长跳连接（Long Skip Connection）融合不同层的特征图以得到分辨率高且语义信息丰富的特征图用于密集预测[104,105]。例如，特征金字塔网络（Feature Pyramid Network, FPN）通过将高层特征上采样并与低层特征相加，对于每个尺度的特征图均赋予了高层语义信息[104]。U-Net 则是将编码器中的低层细节特征与解码器中的高层语义特征拼接（Concatenate）在一起，用来恢复医学影像中血管之类的精细结构[105]。InceptionNet 则是在网络同一层中拼接不同大小的卷积核输出的特征以融合不同尺度的特征[205]。此外，特征融合还能够大幅缓解深度网络中的奇异性[206]，即模型函数中的不连续或导数不存在的现象。残差网络（Residual Network, ResNet）通过短跳连接（Short Skip Connection）融合恒等映射特征和残差学习特征，有效地缓解了网络训练过程中的梯度消失和退化问题，使得训练非常深的网络成为了可能[185,204]。然而，尽管特征融合是现代神经网络架构中必不可少的组件，大多数工作主要致力于探索如何在不同的网络架构中构建新的长跳、短跳连接通路，但是对于特征融合方式本身，往往仍然采用相加或者拼接这样的简单操作。

对于目标检测、语义分割这类任务，如何处理目标尺度的大幅度变化是其核心问题之一[3,207]。大量的研究工作表明，卷积网络本身并不具有真正的尺度不变性，因此在设计网络结构时，必须充分考虑网络预测层的感受野是否与目标尺度相匹配[208]。对于特征融合来说，相加或者拼接这样的简单操作仅能够对不同尺度的特征图进行权重固定的线性融合，无法根据特征本身进行自适应的融合[188]。近年来，注意力机制在计算机视觉的各个任务中均取得了令人瞩目的成功，其可以根据输入动态生成权重的特性赋予了网络极大的灵活性。其中，挤压-激发网络（Squeeze-and-Excitation Network, SENet）构建了一种高效、轻量级的门控机制，通过显式建模特征通道之间的相互依赖关系生成自适应的权重，实现了对特征图的动态精炼[21]。选择核网络（Selective Kernel Network, SKNet）将这种通道加权思想与 Inception 模块中的多分支网络结构相结合，实现了对于网络同一层中不同感受野大小的卷积核所提取特征的动态选择[188]。金字塔注意力网络中的全局注意力上采样（Global Attention Upsample, GAU）模块则是采用 SENet 中的通道注意力模块生成动态权重，在跨层的特征融合中实现自顶向下的特征调制[189]。然而，这些基于注意力机制的特征融合方法仍然存在着以下问题：

（1）缺少一种通用的方式来统一不同的特征融合场景：目前关于特征融合的工作，大多只针对长跳连接、短跳连接和同层融合等场景中的单个场景展开，而且主要集中于长跳连接，研究者们为其设计了各种复杂的融合方式来克服语义与分辨率之间的矛盾。然而，在短跳连接和同层融合中，特别是前者，通常还保持着相加或者拼接这样的简单方式，其特征融合的潜力尚未被充分发掘。事实上，这些不同的特征融合场景，虽然特征之间尺度差异的程度不同，但是本质上都面临着相同的挑战，即给定不同尺度的特征，如何进行融合可以使得网络性能更好。一个能够克服语义不连续性、有效融合不同尺度特征的模块，应当能够一致性地提高各个场景中

融合后特征的质量。

（2）特征上下文的聚合尺度单一：不管是 GAU 模块还是 SKNet，其所采用的注意力机制仅使用了全局特征上下文，生成的权重对大目标有很强的倾向性、容易忽视小目标。为了能够更好、更合理地生成相应特征图上的选择性权重，注意力模块应当在与待融合特征相近的尺度上聚合特征上下文。考虑到在特征融合中，不同的特征具有不同的尺度和感受野，这意味着注意力模块需要从多个尺度上聚合特征上下文。

（3）简单的特征初始融合方式：不同于 GAU、ExFuse 等模块仅利用单个特征作为融合中的指导信息来源[189,209]，SKNet 的注意力模块可以感知到全部特征的上下文，有利于生成更为合理的选择性权重。但是，这也不可避免地引入了一个特征的初始融合环节，使得最终的特征融合效果不仅取决于注意力模块的设计，还受到特征初始融合质量的限制。在特征语义的不一致性较大时，SKNet 中简单相加的特征初始融合方式无法给后续的注意力模块提供一个良好的输入，阻碍了其融合性能的进一步提升。

为了解决上述问题，本章首先构建了一个通用的注意力特征融合（Attentional Feature Fusion, AFF）框架，用以统一同层融合、长跳连接、短跳连接以及初始特征融合等多种特征融合场景。该框架采用一个多尺度的通道注意力模块，旨在通过聚合不同尺度的特征上下文匹配待融合特征中不同尺度的物体，从而更好地实现特征之间的选择性融合。最后，本章还探索了一个迭代的注意力特征融合（iterative Attentional Feature Fusion, iAFF）机制，将上一阶段注意力特征融合的结果作为下一阶段融合权重生成模块的输入，以此来提高特征初始融合的质量，从而进一步提升最终的特征融合性能。

5.2 多尺度通道注意力模块

当输入图像中的目标尺度变化时，网络中特征图上对应的目标特征的尺度也会相应变化。上一章中的多尺度空间注意力模块采用不同感受野大小的多个逐深度卷积核来聚合特征上下文，在语义分割任务上取得了比单一尺度更好的效果。但是由于无法捕获特征通道之间的依赖关系，其对于 CIFAR-100 数据集这样的图像分类任务并没有带来性能提升。事实上，在通道注意力模块中，尺度因素同样存在且影响着最终生成的权重。例如，SENet 采用全局平均池化聚合相应的全局特征上下文，此过程能够加强大目标特征，却会弱化小目标。与之相反，ChaATAC 单元使用的局部通道注意力模块去除了全局平均池化，改用逐点卷积（Point-wise Convolution）聚合局部的通道特征上下文，虽然有利于保存目标细节特征，但却缺少了全局语义信息。

为了能够让注意力模块聚合不同尺度的通道特征上下文，本节提出了一个多尺度通道注意力模块（Multi-scale Channel Attention Module, MS-CAM），具体如图 5.1 所示。其中，Global-AVGPooling 表示全局平均池化，$\oplus$ 和 $\otimes$ 分别代表附带了维数对齐和广播机制的逐元素相加和逐元素相乘，r 为通道压缩比例。给定一个特征图 $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$，其经由多尺度通道注意力模块

精炼后的特征图 $\mathbf{X}' \in \mathbb{R}^{C\times H\times W}$ 可以由下式得到：

$$\mathbf{X}' = \mathbf{M}(\mathbf{X}) \otimes \mathbf{X} = \sigma\left(\widehat{\mathbf{G}}(\mathbf{X}) \oplus \widehat{\mathbf{L}}(\mathbf{X})\right) \otimes \mathbf{X}, \tag{5.1}$$

其中 $\mathbf{M}(\mathbf{X}) \in \mathbb{R}^{C\times H\times W}$ 为多尺度通道注意力模块产生的权重，$\widehat{\mathbf{G}}(\mathbf{X}) \in \mathbb{R}^{C\times 1\times 1}$ 表示图 5.1 左侧分支所聚合的全局通道特征上下文，而 $\widehat{\mathbf{L}}(\mathbf{X}) \in \mathbb{R}^{C\times H\times W}$ 则是右侧分支所聚合的局部通道特征上下文。

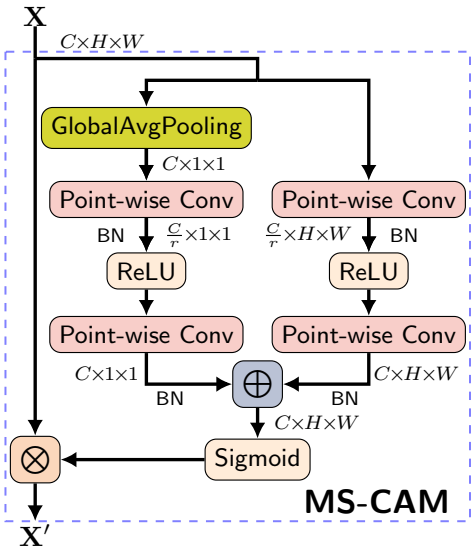


图 5.1　多尺度通道注意力模块示意图

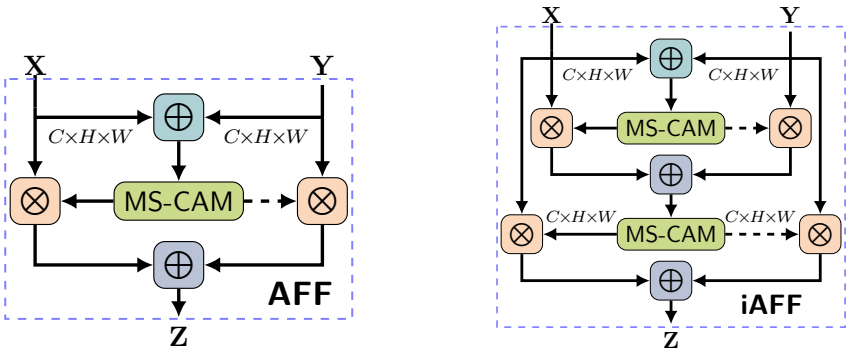
5.3　注意力特征融合

5.3.1　统一的注意力特征融合框架

对于两个给定的特征图 $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{C\times H\times W}$ 和多尺度通道注意力模块 $\mathbf{M}$，本章所提出的注意力特征融合（Attentional Feature Fusion, AFF）可以表示为

$$\mathbf{Z} = \mathbf{M}(\mathbf{X} \uplus \mathbf{Y}) \otimes \mathbf{X} + (1 - \mathbf{M}(\mathbf{X} \uplus \mathbf{Y})) \otimes \mathbf{Y} \tag{5.2}$$

式中，$\mathbf{Z} \in \mathbb{R}^{C\times H\times W}$ 表示融合后的特征，$\uplus$ 代表待融合特征 $\mathbf{X}$ 和 $\mathbf{Y}$ 的初始融合方式。多尺度注意力特征融合的架构如图 5.2(a) 所示，其中 Sigmoid 右侧的虚线表示 1 减去 Sigmoid 函数的输出权重。默认情况下，假设 $\mathbf{Y}$ 的感受野大于 $\mathbf{X}$。具体而言，在同层特征融合场景的 Inception 模块中，$\mathbf{X}$ 和 $\mathbf{Y}$ 分别为 3×3 和 5×5 卷积核输出的特征；在短跳连接场景的 ResNet 块中，$\mathbf{X}$ 是恒定映射特征，$\mathbf{Y}$ 是残差学习特征；在 FPN、U-Net 等长跳连接中，$\mathbf{X}$ 为特征金字塔低层或网络编码端的特征，$\mathbf{Y}$ 为高层或解码端的特征。为了表示方便，$\mathbf{X}$ 和 $\mathbf{Y}$ 已经经过相应的上采样和通道变换运算，使得两者具有相同的大小。



(a) AFF 模块　　(b) iAFF 模块

图 5.2　注意力特征融合模块示意图

表 5.1 总结了深度学习中较为典型的特征融合方案。其中，$\mathbf{G}$ 表示聚合全局特征上下文的注意力模块，可以是像 SENet 那样的全局通道注意力机制[21]，也可以是自注意力机制[95,96,210]。$\mathbf{X}_{:,i,j}$ 和 $\mathbf{Y}_{:,i,j}$ 表示特征图 $\mathbf{X}$ 和 $\mathbf{Y}$ 在位置 (i,j) 上的通道特征，假设特征拼接后采用逐点卷积进行通道特征融合，$\mathbf{W_A}$ 和 $\mathbf{W_B}$ 为相应的融合权重。从中可以看出，特征融合在逐渐由固定权重的线性融合方式向可以根据输入特征动态调整融合权重的非线性方式发展。权重生成模块也从只能感知部分特征的上下文信息，向着能够感知全部特征上下文的方向演进。在 SKNet 的基础上，本章方法不仅将注意力特征融合推广到短跳连接和长跳连接等更多的特征融合场景，为了更好地保存特征图中较小尺度的物体，还构建了多尺度通道注意力模块来生成融合权重。此外，相较于 SKNet 中的简单相加，本章还强调了特征初始融合方式 $\uplus$ 对于最终融合特征质量的重要性，并为此引入了更为动态的初始融合方式。

表 5.1　深度网络中不同的特征融合方案

融合权重	上下文	融合类型	公式	融合场景
固定、线性	无	相加	$\mathbf{X}+\mathbf{Y}$	短跳[185,204], 长跳[104,211,212]
		拼接	$\mathbf{W_A}\mathbf{X}_{:,i,j}+\mathbf{W_B}\mathbf{Y}_{:,i,j}$	同层[205], 长跳[105,213]
动态、非线性	部分	精炼	$\mathbf{X}+\mathbf{G}(\mathbf{Y})\otimes\mathbf{Y}$	短跳[21,22,94,172]
		调制	$\mathbf{G}(\mathbf{Y})\otimes\mathbf{X}+\mathbf{Y}$	长跳[189]
		选择	$\mathbf{G}(\mathbf{X})\otimes\mathbf{X}+(\mathbf{1}-\mathbf{G}(\mathbf{X}))\otimes\mathbf{Y}$	短跳[214]
	全部	调制	$\mathbf{G}(\mathbf{X},\mathbf{Y})\otimes\mathbf{X}+\mathbf{Y}$	长跳[210]
		选择	$\mathbf{G}(\mathbf{X}+\mathbf{Y})\otimes\mathbf{X}+(1-\mathbf{G}(\mathbf{X}+\mathbf{Y}))\otimes\mathbf{Y}$	同层[188]
			$\mathbf{M}(\mathbf{X}\uplus\mathbf{Y})\otimes\mathbf{X}+(1-\mathbf{M}(\mathbf{X}\uplus\mathbf{Y}))\otimes\mathbf{Y}$	同层, 短跳, 长跳（本章方法）

5.3.2 迭代注意力特征融合

与表 5.1 中仅感知部分特征的融合方式不同，作为全部待融合特征都对注意力模块可见的方式，SKNet 和本章方法均存在一个特征初始融合问题，即如何融合不同特征作为注意力模块的输入。SKNet 中所采用的逐元素相加并没有考虑到特征之间的差异，当网络深度较小时，特征初始融合的质量将会较大程度地影响网络最终的性能。考虑到特征初始融合仍然是一个特征融合问题，其融合质量可以由另一个注意力特征融合模块改善，即上一阶段注意力特征融合的结果作为下一阶段权重生成模块的输入。本章将这种在注意力特征融合框架下迭代优化注意力模块输入的方式称作迭代注意力特征融合（iterative Attentional Feature Fusion, iAFF），其模块结构如图 5.2(b) 所示。由此，式 (5.2) 中的特征初始融合 $\uplus$ 可以定义为

$$\mathbf{X} \uplus \mathbf{Y} = \mathbf{M}(\mathbf{X} + \mathbf{Y}) \otimes \mathbf{X} + (1 - \mathbf{M}(\mathbf{X} + \mathbf{Y})) \otimes \mathbf{Y} \tag{5.3}$$

需要注意的是，在相同的网络参数和计算量情况下，由于加大了网络的非线性程度，iAFF 模块可以在 AFF 模块的基础上进一步提升网络性能，但同时也会增加网络优化的难度。

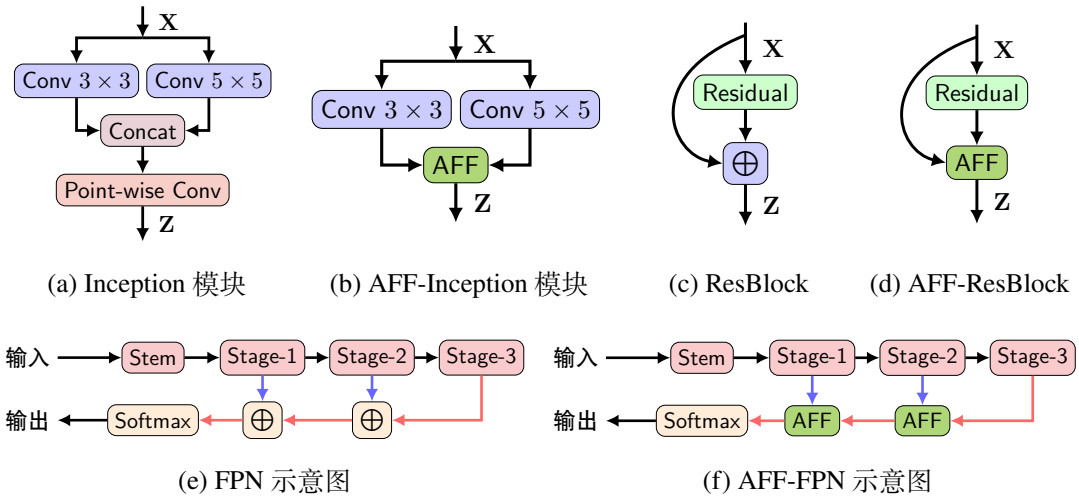
5.3.3 注意力特征融合网络示例

为了验证所提出的注意力特征融合模块，本章选择 ResNet、FPN 和 InceptionNet 作为示例网络，分别对应短跳连接、长跳连接和同层融合这三个深度网络中最常见的特征融合场景。通过将原始网络中简单的相加、拼接等操作替换为 AFF 模块，可以得到相应的 AFF-ResNet、AFF-FPN、AFF-InceptionNet。图 5.3 展示了网络模块替换前后的对比，图 5.3 (a)、(c)、(e) 分别为 Inception 模块、残差块（Residual Block, ResBlock）和 FPN 的示意图，而图 5.3 (b)、(d)、(f) 则为相应的 AFF-Inception 模块、AFF-ResBlock 和 AFF-FPN。需要注意的是，为了突出注意力模块在不同感受野大小的卷积核之间的选择性，图 5.3 (a) 中的 Inception 模块是简化后的版本。本章中 InceptionNet、AFF-InceptionNet、ResNet、AFF-ResNet 采用与表 4.3 中相同的骨干网络。FPN 和 AFF-FPN 的骨干网络如表 5.2 所示，与表 4.3 中一样，b 为骨干网络中每阶段的残差块数量。

5.4 实验分析与讨论

为了充分评估本章所提出的注意力特征融合框架和多尺度通道注意力模块，本节将进行详细消融实验和对比实验，具体探索以下问题：

（1）问题一：与在 SENet、SKNet、GAU 模块中被广泛使用的全局通道注意力机制不同，针对融合特征之间尺度、语义上的不连续性，本章构建了多尺度通道注意力模块，用于在注意力特征融合过程中更好地平衡不同尺度的目标特征。5.4.2 小节将研究在给定相同网络参数的情况下，相较于单一尺度的通道注意力模块，多尺度通道注意力模块是否能够获得更好的特征融合效果。



(a) Inception 模块　(b) AFF-Inception 模块　(c) ResBlock　(d) AFF-ResBlock

(e) FPN 示意图　(f) AFF-FPN 示意图

图 5.3　多种注意力特征融合模块和网络的示意图

表 5.2　AFF-FPN 的骨干网络架构

网络阶段	输出特征图大小	ResNet-20-V2
Stem	120 × 120	$\begin{bmatrix} 3\times3 \text{ conv}, 8 \\ 3\times3 \text{ conv}, 8 \\ 3\times3 \text{ conv}, 16 \\ 3\times3 \text{ MaxPool} \end{bmatrix}$
Stage-1	120 × 120	$\begin{bmatrix} 3\times3 \text{ conv}, 16 \\ 3\times3 \text{ conv}, 16 \end{bmatrix} \times b$
Stage-2	60 × 60	$\begin{bmatrix} 3\times3 \text{ conv}, 32 \\ 3\times3 \text{ conv}, 32 \end{bmatrix} \times b$
Stage-3	30 × 30	$\begin{bmatrix} 3\times3 \text{ conv}, 64 \\ 3\times3 \text{ conv}, 64 \end{bmatrix} \times b$

（2）问题二：表 5.1 总结了目前各类深度网络中较为典型的特征融合方案，这些方案在是否动态分配融合权重、注意力模块是否能够感知到全部特征、注意力模块是用来调制或者精炼单个特征图还是为多个特征分配相应的融合权重等策略上有着诸多差异。5.4.2 小节还将研究在给定相同的网络参数和多尺度通道注意力模块的情况下，相较其他特征融合方案，本章方法所采用的动态、感知全部特征、为多个特征分配融合权重的策略是否能够获得更好的特征融合效果。

（3）问题三：通过将原有的相加或者拼接操作替换为本章所提出的多尺度注意力特征融合模块，5.4.3 小节针对同层融合、短跳连接、长跳连接等特征融合场景构建了 AFF-Inception、AFF-ResNet、AFF-FPN 等新的示例网络。5.4.3 小节将研究相较于使用其他融合方式或者其他注意力机制的网络，这些示例网络能否取得更好的性能。

5.4.1 实验设置

为了验证本章所构建的注意力特征融合框架的通用性，本节采用 CIFAR-10、CIFAR-100、ImageNet 等图像分类数据集验证代表短跳连接场景的残差网络（ResNet）和代表同层融合场景的 InceptionNet，而在 StopSign 数据集上对代表长跳连接场景的特征金字塔网络（FPN）进行小目标语义分割任务的性能评估。

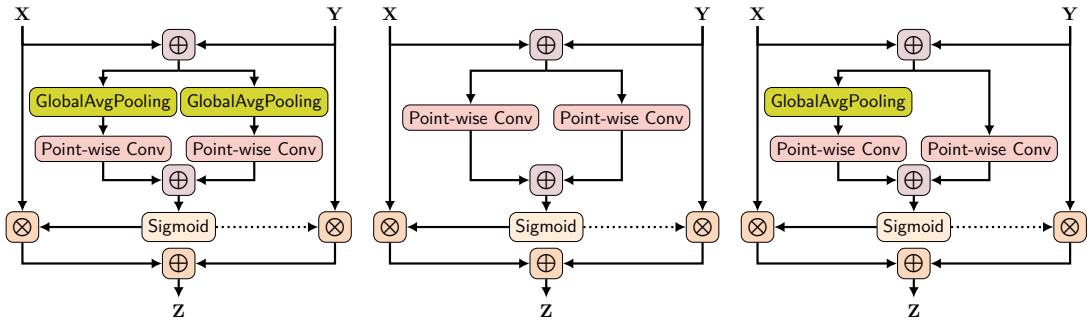
对于图像分类实验，分别选用 ResNet-20 和 ResNet-50 作为 CIFAR-10/100 数据集和 ImageNet 数据集上具体的宿主网络，具体网络配置如表 4.3 所示。InceptionNet 采用与 ResNet-20 相同的网络结构配置，只是将原有的 3×3 卷积替换为 3×3 卷积和 5×5 卷积的并行结构。为了节省网络参数，5×5 卷积由膨胀因子为 2 的 3×3 卷积实现。CIFAR-10/100 上的图像分类实验采用交叉熵作为损失函数、Kaiming 方法[197] 初始化网络、Nesterov 加速梯度（Nesterov Accelerated Gradient, NAG）方法来优化网络，权重衰减率为 0.0001，批大小（Batchsize）为 128，一共训练 300 轮，初始学习率为 0.1 且在第 200 轮和 250 轮后各除以 10，注意力特征融合模块中的通道压缩比例为 4。对于 ImageNet 数据集上的实验，出于保持参数量相当的目的，本节仅在 ResNet-50 中后两个阶段的 9 个瓶颈残差块中使用注意力特征融合。考虑到每个瓶颈残差块中最后的卷积会将通道数膨胀 4 倍，注意力特征融合模块中的通道压缩比例为 16。网络学习率为 0.075 并按照余弦曲线衰减，一共训练 160 轮，其余的优化参数设置与 CIFAR-10/100 数据集相同。

对于小目标语义分割实验，网络的总体架构和骨干网络分别如图 5.3(f) 和表 5.2所示，采用平均交并比（mean Intersection over Union, mIoU）作为评价指标、Soft-IoU[190] 作为损失函数、AdaGrad 作为优化方法[191]，学习率为 0.1，权重衰减率为 0.0001，训练环节的批大小为 32，测试环节的批大小为 23，一共训练 200 轮，学习率为 0.05。读入图像大小为 512×512，训练过程中数据增广环节采用大小为 480×480 的随机裁剪和随机镜像翻转。

5.4.2 消融实验与分析

对于问题一，图 5.5 描述了三种通道特征上下文的聚合方式，即全局 + 全局、局部 + 局部、全局 + 局部，用于验证聚合多尺度的通道特征上下文对于注意力特征融合的重要性。三者具有相同的网络参数数量，差别只在于通道特征上下文聚合的尺度不同。表 5.3 给出了这三者在 InceptionNet（同层融合）、ResNet（短跳连接）、FPN（长跳连接）上的实验结果。从中可以看出，在所有实验设置上，聚合多尺度上下文（全局 + 局部）的网络均优于仅使用单一尺度特征上下文的网络，这表明多尺度的通道特征上下文聚合能够有效提升注意力特征融合的质量。

对于问题二，在给定相同的网络参数和多尺度通道注意力模块的情况下，图 5.5 给出了四种特征融合方式的模块结构示意图。其中，前缀 Ab 表示这些结构是专门针对消融（Ablation）实验设计的，不同于这些融合方式原有的模块结构，MS 表示这些模块都采用了本章所构建的



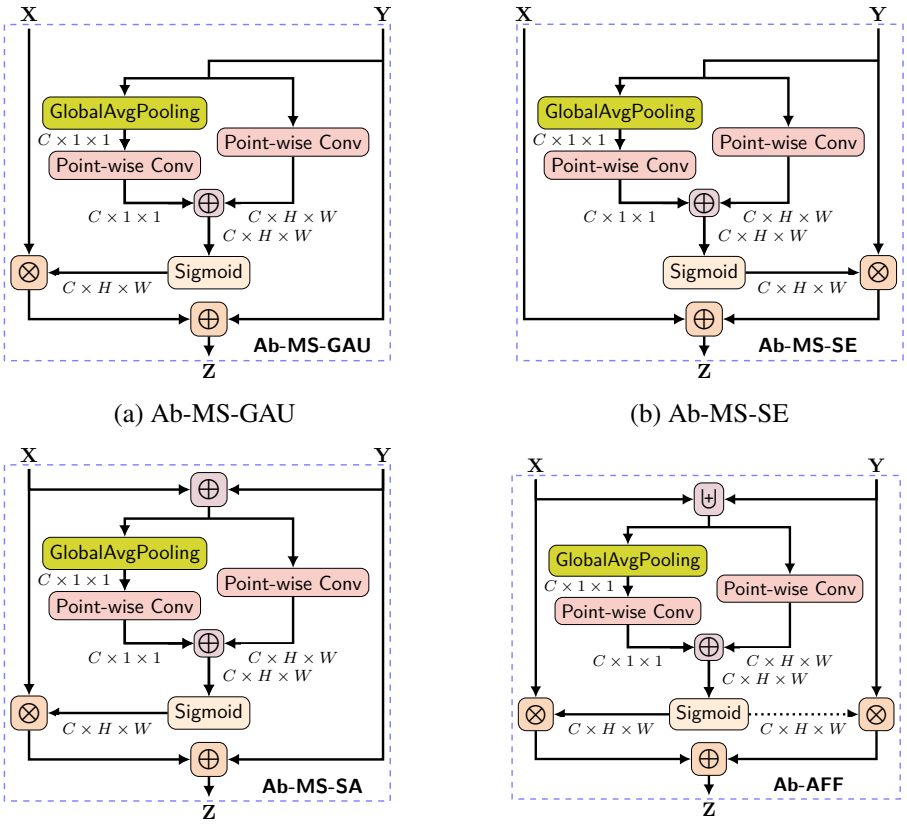
(a) 全局 + 全局　　(b) 局部 + 局部　　(c) 全局 + 局部

图 5.4　特征上下文聚合尺度实验所采用的模块结构图

表 5.3　不同特征上下文聚合尺度的结果比较

聚合尺度	InceptionNet（同层）				ResNet（短跳）				FPN（长跳）			
	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
全局 + 全局	0.735	0.766	0.775	0.789	0.754	0.796	0.811	0.821	0.911	0.923	0.936	0.939
局部 + 局部	0.746	0.771	0.785	0.787	0.754	0.794	0.808	0.814	0.895	0.919	0.921	0.924
全局 + 局部	**0.756**	**0.784**	**0.794**	**0.801**	**0.763**	**0.804**	**0.816**	**0.826**	**0.924**	**0.935**	**0.939**	**0.944**

多尺度（Multi-scale，MS）通道注意力模块。表 5.4 展示了在同层融合的 InceptionNet、短跳连接的 ResNet、长跳连接的 FPN 这三种特征融合场景下，上述融合模型的对比结果。从中可以看到，1）相比于属于线性特征融合的相加和拼接，基于注意力机制的非线性融合方式可以提供更好的特征融合效果。2）在这些非线性融合方式之间，Ab-MA-GAU 模块和 Ab-MS-SA 模块都是对低层特征进行自上而下调制，两者的差别在于 Ab-MA-GAU 模块的权重只依赖于高层特征 $\mathbf{Y}$，而 Ab-MS-SA 模块的输入则是待融合特征之和 $\mathbf{X}+\mathbf{Y}$。在大多数实验设置中，Ab-MS-SA 模块的性能均稍好于 Ab-MA-GAU 模块，这表明相较于仅能感知部分特征的融合方式，能够感知全部特征的注意力融合方式具有更好的潜力和表现。3）对比 Ab-MS-SA 模块与 Ab-AFF 模块，这两者均以 $\mathbf{X}+\mathbf{Y}$ 作为注意力模块的输入，差别在于在 Ab-MS-SA 模块中，$\mathbf{Y}$ 的权重恒定为 1，生成的权重只用于调制 $\mathbf{X}$。而在 Ab-AFF 模块中，$\mathbf{X}$ 和 $\mathbf{Y}$ 两者的权重之和为 1，共同被调制。在大多数实验设置中，Ab-AFF 模块的性能均稍好于 Ab-MS-SA 模块，这表明以加权平均的方式在特征之间做 ‘‘软’’ 选择的特征融合方式效果好于只调制单个特征的方式。4）相比上述较为微小的改进，在给定相同网络参数的情况下，iAFF 模块在几乎所有实验设置下均显著好于 Ab-AFF 模块以及其他对比方案，这表明在注意力特征融合框架内，除了设计更加先进的注意力模块之外，改善特征初始融合质量也可以有效提升最终的特征融合效果。但是需要注意的是，iAFF 模块在优化上存在一定困难，比如 $b=4$ 时的 iAFF-ResNet 其性能会有退化现象。



(a) Ab-MS-GAU (b) Ab-MS-SE

(c) Ab-MS-SA (d) Ab-AFF

图 5.5 消融实验所采用的特征融合模块示意图

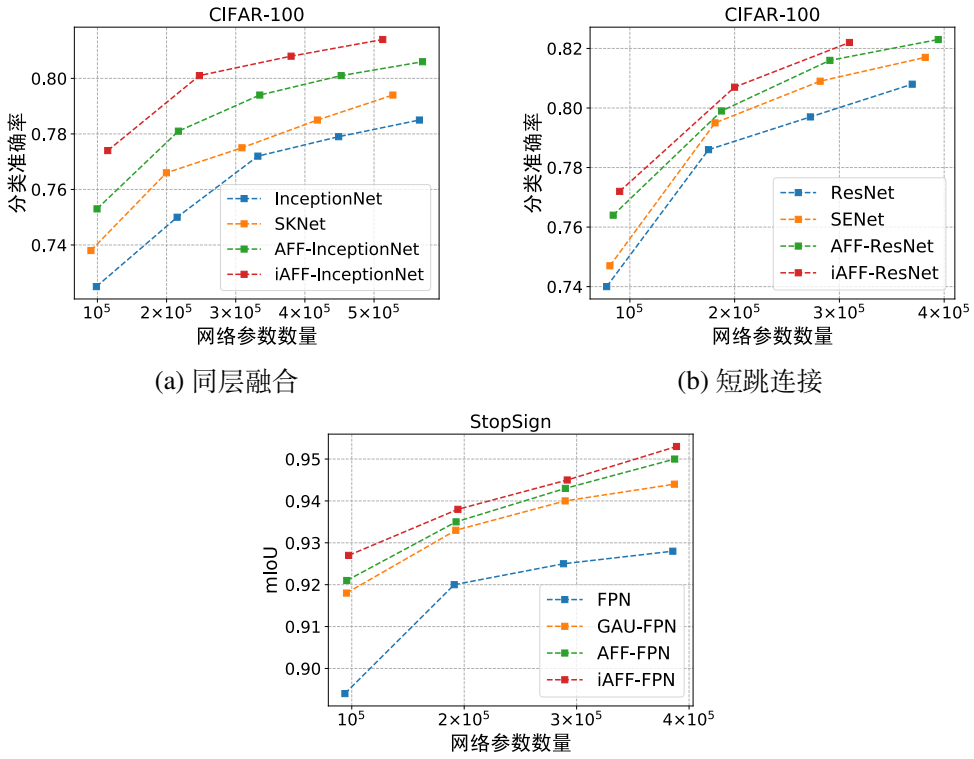
表 5.4 不同特征融合场景下多种融合策略的消融实验结果比较

方法	上下文	类型	InceptionNet（同层）				ResNet（短跳）				FPN（长跳）			
			$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
相加	无	\	0.720	0.753	0.771	0.782	0.740	0.786	0.797	0.808	0.895	0.920	0.925	0.928
拼接	无	\	0.725	0.749	0.772	0.779	0.742	0.782	0.793	0.798	0.897	0.909	0.925	0.939
Ab-MS-GAU	部分	调制	0.751	0.774	0.788	0.795	0.766	0.803	0.815	0.819	0.917	0.926	0.937	0.941
Ab-MS-SE	部分	精炼	0.752	0.780	0.790	0.798	0.765	0.799	0.814	0.820	0.915	0.929	0.940	0.940
Ab-MS-SA	全部	调制	0.756	0.779	0.790	0.798	0.761	0.801	0.814	0.822	0.920	0.932	0.938	0.941
Ab-AFF	全部	选择	*0.756*	*0.784*	*0.794*	*0.801*	*0.763*	*0.804*	*0.816*	**0.826**	*0.924*	*0.935*	*0.939*	*0.944*
iAFF	全部	选择	**0.774**	**0.801**	**0.808**	**0.814**	**0.772**	**0.807**	**0.822**	/	**0.927**	**0.938**	**0.945**	**0.953**

5.4.3 方法对比与分析

图 5.6 展示了在网络深度逐渐增加的情况下，本章所构建的注意力特征融合网络与其他深度网络的性能比较。从中可以看到，1）相较于 InceptionNet、ResNet、FPN，将这些网络中原本简单

的特征融合操作替换为本章提出的注意力特征融合模块得到的 AFF-InceptionNet、AFF-ResNet、AFF-FPN 性能更好。当网络深度相同时，采用 AFF 模块虽然会增加一定的参数，但其所带来的性能提升使得网络整体更加高效。比如，在同层融合场景中，$b = 2$ 时的 AFF-InceptionNet 可以在 CIFAR-100 数据集上取得与 $b = 5$ 时的 InceptionNet 接近的性能，而此时 AFF-InceptionNet 的参数量只有 InceptionNet 的 38%；在短跳连接场景中，$b = 4$ 时的 AFF-ResNet 其分类准确率好于 $b = 5$ 时的 ResNet；在长跳连接场景中，$b = 2$ 时的 AFF-FPN 所取得的 mIoU 指标好于 $b = 5$ 时的 FPN，而前者的参数量只有后者的 50%。2）同样的，与 SKNet、SENet、GAU-FPN 相比较，在所有实验设置下，AFF-InceptionNet、AFF-ResNet、AFF-FPN 都相应地取得了更好的效果。考虑到每个特征融合场景中，这些网络之间的区别主要在于 SKNet、SENet、GAU-FPN 等对比网络均使用了单一尺度的全局通道注意力模块，而本章所构建的注意力特征融合网络使用了多尺度通道注意力模块，这也再次表明了聚合多尺度特征上下文可以有效提升注意力模块的性能。3）对比 AFF-InceptionNet、AFF-ResNet、AFF-FPN 与 iAFF-InceptionNet、iAFF-ResNet、iAFF-FPN，可以看到在优化没有问题的情况下，在网络中使用 iAFF 模块可以进一步提升相应的预测性能。这表明在设计更加精巧、先进的注意力模块之外，改善作为注意力模块输入的特征初始融合质量也是一条提升特征融合效果的可行途径。



(a) 同层融合　　(b) 短跳连接

(c) 长跳连接

图 5.6　本章所构建的注意力特征融合网络与其他深度网络的性能比较

为了探索所提出的 MS-CAM 模块对于物体定位和小物体识别的影响，采用梯度权重类激活映射[215]（Gradient-weighted Class Activation Mapping, Grad-CAM) 方法对 ResNet-50，SENet-50 和 AFF-ResNet-50 在部分 ImageNet 数据集图像上进行了网络可视化比较，具体结果如图 5.7 所示。图中显示了预测类的热力图，越接近红色的区域表示该区域对分类贡献越大，越偏蓝色则表示贡献越小，预测错误的情形用红色叉号表示。类别名称和模型给出的类别概率则在热力图底部。从图 5.7 的上半部可以看出，基准网络 ResNet-50 的定位能力相对较差，在许多情况下其高热度区域与真实目标区域存在较大幅度的错位，而 SENet-50 虽然能够定位到真实的物体，但是其注意分配的区域过大，包含了许多背景成分，这是由于 SENet-50 仅利用全局的通道注意力，因此其对于全局范围具有很强的倾向性。与之相反，本章提出的 AFF-ResNet-50 所关注的区域与标记物体所处的区域高度重合，而且类别预测的概率也最高，这表明 AFF 模块可以帮助网络很好地学习如何准确定位物体，并且通过抑制与物体无关的背景干扰，可以提高网络的分类准确率。这些性能提升的原因在于除了全局尺度的通道上下文之外，AFF 所采用的 MS-CAM 模块还汇聚了局部尺度的通道上下文，这有助于网络抑制背景杂波干扰，将注意力投向需要关注的目标特征，这也有利于小物体识别。从图 5.7 的下半部可以清楚地看到，AFF-ResNet-50 可以正确预测小尺度的物体，而 ResNet-50 在大多数情况下则会失败。

最后，为了验证本章所提出的 AFF/iAFF 模块在更深的网络上的性能，以 ResNet 作为基准网络，将其自 Stage-3 之后的残差块（ResBlock）替换为采用 AFF 或 iAFF 模块融合恒等映射特征与残差学习特征的 AFF-ResBlock 和 iAFF-ResBlock，构建了 AFF-ResNet 和 iAFF-ResNet。此外为了验证 AFF/iAFF 模块的通用性，还选取了 ResNeXt[216] 作为另一个宿主网络，构造了相应的 AFF-ResNeXt 和 iAFF-ResNeXt。在 CIFAR-100 和 ImageNet 这两个数据集上，选取了多种方法进行对比，包括注意力增强的宽残差网络[95]（Attention Augmented Wide ResNet）、注意力增强的残差网络（Attention Augmented ResNet）、挤压-激发网络[21]（Squeeze-and-Excitation Networks, SENet）、选择核网络[188]（Selective Kernel Networks, SKNet）、金字塔网络[217]（PyramidNet）、神经结构转换网络[218]（Neural Architecture Transfer, NAT）、基于自动增广[219]（AutoAugment）与 ShakeDrop 正则化[220] 的金字塔网络[219]（AutoAugment+PyramidNet+ShakeDrop）、残差网络[204]（ResNet）、高效通道注意力网络[90]（Efficient Channel Attention Net）、聚集-激发网络[172]（Gather-Excite Networks, GENet）、局部重要性池化残差网络[221]（Local Importance Pooling ResNet）。实验结果如**??** 和表 5.5 所示，其中 M 表示百万（Million）。从中可以看出，在网络最浅、参数量最小的情况下，本章基于 AFF/iAFF 模块所构造的网络取得了最佳的分类性能，表明了注意力特征融合模块中所采用的多尺度通道注意力模块特征精炼的有效性。

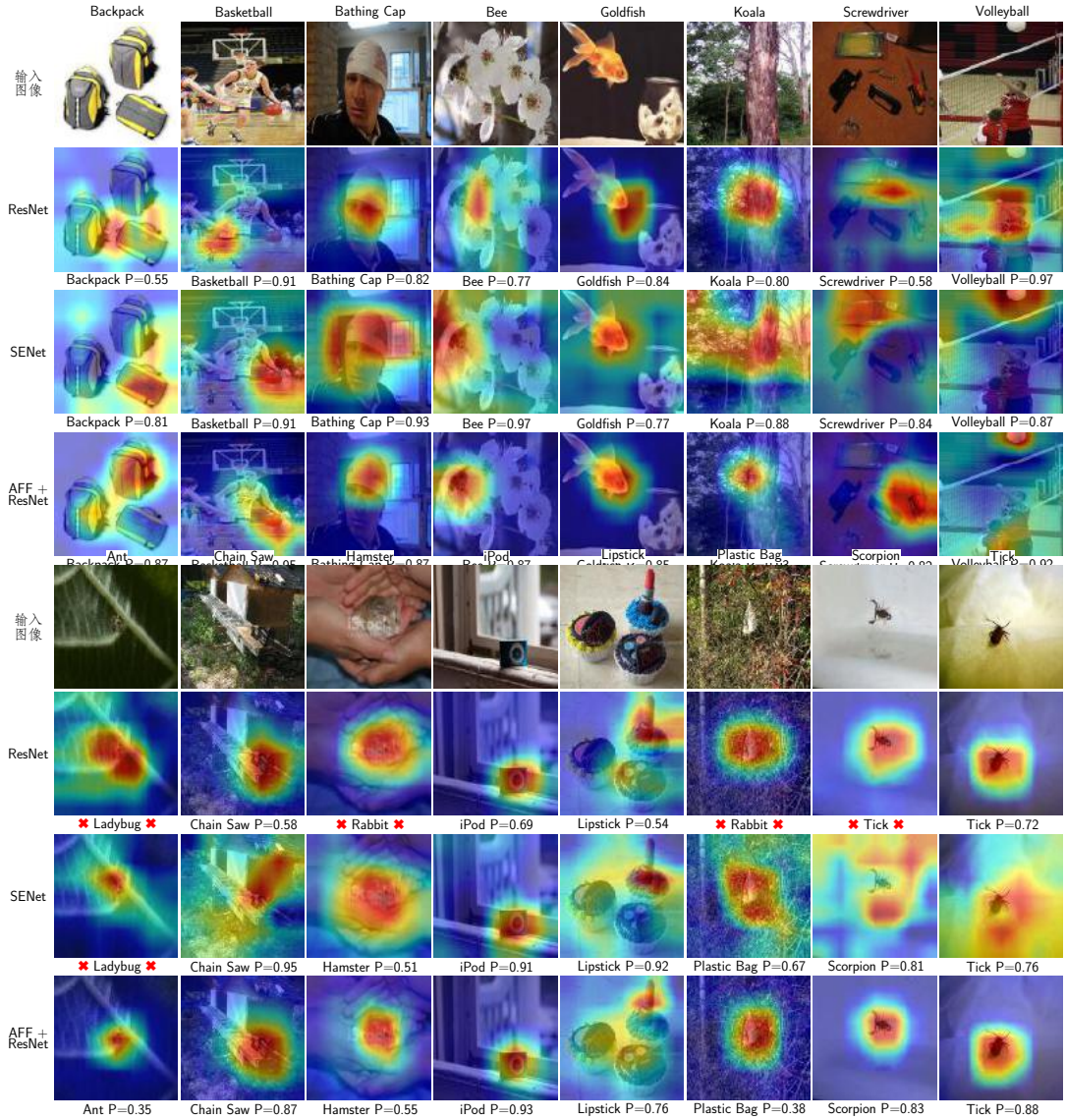


图 5.7　基于 Grad-CAM 的网络可视化比较

表 5.5　与其他先进方法在 ImageNet 数据集上的性能对比

网络架构	top-1 错误率 / %	参数数量
ResNet-101 [204]	23.2	42.5 M
Efficient-Channel-Attention-Net-101 [90]	21.4	42.5 M
Attention-Augmented-ResNet-101 [95]	21.3	45.4 M
SENet-101 [21]	20.9	49.4 M
Gather-Excite-θ^+-ResNet-101 [172]	20.7	58.4 M
Local-Importance-Pooling-ResNet-101 [221]	20.7	42.9 M
AFF-ResNet-50（本章方法）	***20.9***	***30.3 M***
AFF-ResNeXt-50-32x4d（本章方法）	***20.8***	***29.9 M***
iAFF-ResNet-50（本章方法）	***20.2***	***35.1 M***
iAFF-ResNeXt-50-32x4d（本章方法）	***19.8***	***34.7 M***

5.5　本章小结

特征融合是各种深度网络中必不可少的环节，一个更高质量的、即插即用的特征融合方式可以普遍性地提高各种网络的性能。为此，本章提出了一个根据特征上下文动态分配权重的通用特征融合框架——注意力特征融合，并将其运用于短跳连接、长跳连接和 Inception 模块等特征融合场景。考虑到待融合特征之间可能存在较大的语义和尺度不连续性，该框架通过聚合初始融合特征上不同尺度的特征上下文生成相应的融合权重，避免融合过程过于倾向大目标而弱化小目标。此外，本章还展示了在注意力特征融合框架中特征初始融合的重要性，通过另一层注意力特征融合可以改善注意力模块的输入，生成更高质量的融合权重，从而更好地实现特征融合。在图像分类和小目标分割任务上的大量消融实验和对比实验证明了本章所构建的注意力特征融合框架的有效性。实验结果表明，与单纯地提升网络深度相比，采用更为先进的特征融合方案是一种更加高效、经济的提升网络性能的方式。

第六章　基于注意力局部对比度网络的红外小目标检测

对于小目标检测，前面的章节各自在模型驱动方法和数据驱动方法上进行了探索，两类方法各自有着优缺点。模型驱动方法依赖领域知识建立数学模型，通过相应的算法步骤或者求解具体的优化问题获得小目标的显著性图，但是模型的不精确性、特征的判别性不足、超参数对图像场景变化敏感等问题使其难以应对复杂多变的真实场景。数据驱动方法则是将深度神经网络作为黑箱，以端到端的方式从标记数据学习输入图像与检测结果之间的非线性映射关系。这类方法需要大量高质量的标记数据，然而红外小目标数据集较小的规模限制了其性能的提升。

为了进一步提升红外小目标检测的性能，本章提出了一个注意力局部对比度网络（Attentional Local Contrast Network, ALCNet）来同时利用模型驱动方法中的先验知识与数据驱动方法所依赖的标注数据。ALCNet 将传统模型驱动方法中的局部对比度度量方法模块化，作为特定的非线性特征变换层嵌入深度网络中，并对同一特征图进行不同尺度的局部对比度度量以获得与目标尺度相匹配的最佳度量。随后，ALCNet 采用一个自底向上的局部通道注意力调制（Bottom-up Local Attentional Modulation, BLAM）模块进行跨层的多尺度特征融合以获得最终的局部对比度特征。详细的消融实验和对比实验结果表明，在深度网络中嵌入传统模型驱动方法中的局部对比度先验可以大幅提升红外小目标检测的效果。出于可重复研究的考虑，本章的代码和训练好的模型参数可以从项目主页上获取[1]。

本章的具体内容安排如下：第 6.1 节详细阐述了为何对于红外小目标检测，模型驱动方法和数据驱动方法可以彼此缓解各自的不足，进而阐释了本章将两者融合到同一框架中的研究动机与意义；第 6.2 节详细描述了本章所提出的 ALCNet；第 6.3 节对 ALCNet 进行了消融实验以分别验证嵌入局部对比度先验、同层和跨层的对比度特征融合、BLAM 模块各自对于红外小目标检测的有效性以及模块设计上的合理性，同时还对比了其他方法以验证 ALCNet 的检测性能。第 6.4 节对本章工作的内容进行了小结。

6.1　引言

近年来，红外小目标检测领域取得了长足的发展，各类效果更好的新方法层出不穷，但是其中的模型驱动方法和数据驱动方法大体上各自沿着不同的方向发展，交集并不多。本质上，模型驱动方法是将红外小目标检测看作是一个与图像去噪、图像恢复类似的低层视觉问题。不管是低秩稀疏分解，还是局部对比度度量，这些方法大多致力于依据先验知识将小目标建模成红外图像中的离群点（Outlier），通过求解特定的优化问题或者算法步骤计算稀疏度、显著性、对

[1]https://github.com/YimianDai/open-alcnet

比度等可以反映其离群程度的度量，从而将小目标分割出来。由于真实场景的红外小目标图像收集难度较大，这类方法无需大量被标注数据的优点使其得到了更多研究者的关注。但是模型驱动方法的缺陷在于很难精确建模，其通常是对真实数据模式的某种简化，难以应对复杂多变的现实场景，面临着检测性能和鲁棒性的双重挑战。与之相反的数据驱动方法，特别是基于深度学习的方法，则是将红外小目标检测看作是一个相对高层的视觉感知问题。在不依赖数学模型和领域知识的情况下，这些方法将深度网络作为黑箱以端到端的方式自动学习具有语义判别性的特征以实现输入图像到检测结果之间的映射，取得了显著的性能提升。然而，数据驱动方法的性能在很大程度上依赖于大量高质量的标记数据，而红外小目标图像的稀缺性严重限制了数据集的规模，阻碍了其性能的进一步提升。其次，由于红外小目标本征特征的稀缺以及特征语义与分辨率之间的矛盾，在一定的网络层数之上，继续加深网络对于检测性能的提升效果微乎其微。

事实上，对于红外小目标检测问题，模型驱动方法和数据驱动方法可以在很大程度上缓解彼此的不足。以局部对比度度量方法为例，这些方法建模不精确的根源在于其采用的均值、最大值、熵等特征过于简单，不具有足够的语义判别性来区分真实目标和背景干扰。相比之下，深度网络能够以端到端的方式从标记数据中自动学习如何抽取具有足够语义判别能力的特征。此外，由于在整个训练数据集上进行了多轮的优化，数据驱动方法并不像模型驱动方法那样对于超参数敏感，其具有足够的鲁棒性以应对复杂多变的真实场景。对于深度网络来说，红外小目标本征特征缺乏以及数据集规模较小所引发的问题很大程度上源于其纯粹依赖于标记数据来学习目标外观的特征表示。与之相反的是，红外小目标检测中的模型驱动方法往往并不直接建模目标外观本身，而是依赖于目标与周围邻域或者整体背景之间的某些性质差异，包括显著性、稀疏度、局部对比度等。

基于上述观察，本章提出了一个新的红外小目标检测模型——注意力局部对比度网络（Attentional Local Contrast Network, ALCNet），旨在将数据驱动的深度卷积网络和模型驱动的局部对比度度量方法统一到同一个框架下，从而能够同时利用标记数据和领域知识来提升网络的检测性能。在特征金字塔网络的基础上，ALCNet 将传统方法中的图像块大小与滤波器大小解耦，采用膨胀因子和有效感受野作为代替，使得原本单阶段的多尺度局部对比度度量可以被进一步拆分为两个阶段，依次在网络的同层特征上和跨层特征之间进行局部对比度的多尺度度量与融合。具体而言，在第一阶段，ALCNet 采用并行的多分支架构，其中每个分支共享相同的局部对比度度量模块，但是各自具有不同膨胀因子，以此对网络同一层的特征进行多尺度度量。在第二阶段，ALCNet 构建了一个基于局部通道注意力机制的自底向上调制通路用于特征金字塔中多个局部对比度特征之间的跨层融合，将较小尺度的目标细节信息嵌入高层语义特征中，从而更好地保存和突出红外小目标。此外，为了加快网络的训练和推理速度，本章还设计了一种基于特征图循环移位的加速方案，用以快速计算第一阶段的局部对比度度量。

尽管融合了传统的局部对比度度量机制，ALCNet 仍然是一个端到端的网络，无需任何预训练、预处理和后处理。从深度卷积网络的角度看，ALCNet 中的局部对比度度量模块可以视作为一个具有一定可解释性的非线性变换层，依据相应的物理机理捕获局部特征与其区域上下文之间的交互关系，将输出特征从以目标外观为中心的表示转换为对目标局部对比度的表示，缓解了红外小目标本征特征不足的问题。此外，在卷积神经网络中，有效感受野实际远小于理论感受野，而且随着网络的加深，有效感受野的增长速率也在逐渐降低。局部对比度度量则显式地打破了有效感受野的限制，在保持特征图大小一样的情况下，对相对长程的特征上下文进行了编码，有助于解决红外小目标检测中特征分辨率与语义层次之间的矛盾。从局部对比度度量的角度看，ALCNet 其实是将传统方法中过于简单的均值、最大值等特征替换为由网络以端到端的方式从标注数据中学习到的具有高度语义判别性的特征，提高了对于真实目标与背景干扰物之间的判别能力，从而大幅降低了网络检测的虚警率。

6.2 注意力局部对比度网络

6.2.1 从图像块对比度到特征图对比度

在 SIRST 数据集附带的红外小目标检测性能排行榜中，多尺度块对比度度量（Multi-scale Patch-based Contrast Measure, MPCM）方法[10] 在被测试的多种局部对比度度量方法中表现最好。因此，本章选取 MPCM 作为 ALCNet 中局部对比度模块的设计参考。对于给定的尺度 l，MPCM 方法采用大小为 $l \times l$ 的图像块的均值作为特征，计算当前中心图像块与其互不重叠的背景邻域图像块之间的差异，算法 2 给出了其计算流程。其中 B_k 代表中心块 T 的第 k 个邻域块，如图 6.1(a) 所示。为了快速计算中心块与邻域块之间的均值之差，MPCM 采用八个预先定义的滤波器模板对均值图像进行滤波。对于给定大小为 $H \times W$ 的红外图像以及块大小 l，这需要大约 $8(3l)^2HW$ 次乘法和 $8(3l)^2HW$ 次加法，当 l 数值较大时，这会产生巨大的计算开销。

算法 2: 灰度图像的块对比度度量方法[10]

输入: 红外灰度图像 $f \in \mathbb{R}^{H\times W}$, 给定尺度 l

输出: 尺度 l 下的块对比度 $C^l \in \mathbb{R}^{H\times W}$

计算尺度 l 下的均值图像 m_T^l;

for $i = 1 : H$ **do**

  for $j = 1 : W$ **do**

    for $k = 1 : 4$ **do**

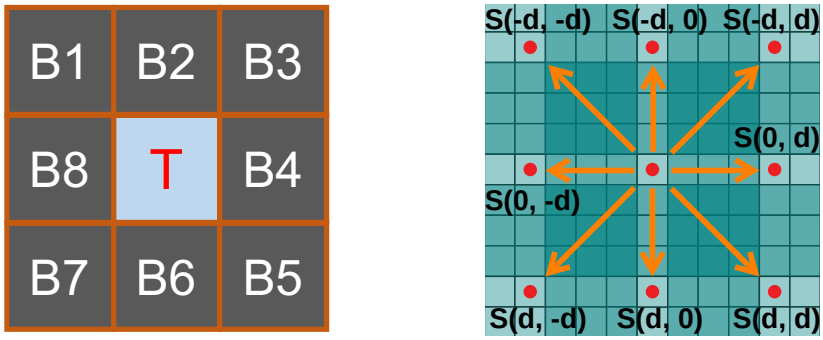
      $\tilde{d}_k^l(i,j) = \left[m_T^l(i,j) - m_{B_k}^l(i,j)\right] \cdot \left[m_T^l(i,j) - m_{B_{k+4}}^l(i,j)\right]$;

    end

    $C_{(i,j)}^l = \min_{k=1,2,3,4} \tilde{d}_k^l(i,j)$;

  end

end



(a) 图像块对比度　　(b) 特征对比度

图 6.1　图像块对比度度量与特征对比度度量的中心和邻域结构示意图

与 MPCM 需要依赖图像块来定义人工特征不同，卷积神经网络能够从标记数据中自动学习特征，因此对于在深度网络特征图上进行的局部对比度度量而言，不再需要像 MPCM 那样划分互不重叠的块，可以采用更为灵活的膨胀因子 d 来表示中心特征点与邻域特征点之间的距离（如图 6.1(b) 所示）。给定一个待度量的特征图 $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$，其位置 (c,i,j) 上的方向性局部对比度 $\mathbf{D}^{(x,y)}_{[c,i,j]}$ 可以表示为

$$\mathbf{D}^{(x,y)}_{[c,i,j]} = \left(\mathbf{F}_{[c,i,j]} - \mathbf{F}_{[c,i-x,j-y]}\right) \cdot \left(\mathbf{F}_{[c,i,j]} - \mathbf{F}_{[c,i+x,j+y]}\right), \tag{6.1}$$

其中 $(x,y) \in \Omega = \{(-d,-d),(-d,0),(-d,d),(0,-d)\}$ 是二维的方向索引。相应的，给定膨胀因子 d，位置 (c,i,j) 上的局部对比度 $\mathbf{C}^{d}_{[c,i,j]}$ 可以表示为

$$\mathbf{C}^{d}_{[c,i,j]} = \min_{(x,y)\in\Omega} \left\{\mathbf{D}^{(x,y)}_{[c,i,j]}\right\}. \tag{6.2}$$

为了快速计算特征图上的局部对比度，本章引入一个额外的假设，即红外图像的边缘互相之间是相似且平滑过渡的。基于该假设，可以通过特征图循环移位的方式来获取对应的邻域特征图，从而将每个中心特征点与其邻域特征点的度量转换为特征图与其邻域特征图之间的度量，即可以把标量形式的式 (6.1) 转化为特征图的整体运算：

$$\mathbf{D}^{(x,y)} = \left(\mathbf{F} - \mathbf{S}^{(x,y)}\right) \odot \left(\mathbf{F} - \mathbf{S}^{(-x,-y)}\right), \tag{6.3}$$

式中，$\odot$ 表示逐元素点积运算，$\mathbf{S}^{(x,y)}$ 为沿着 (x,y) 方向循环移位后得到的邻域特征图。图 6.2 展示了邻域八个方向上循环移位前后的示意图，其中箭头末尾为原始的特征图 $\mathbf{F}$，箭头端点表示移位后的特征图 $\mathbf{S}^{(x,y)}$。借助该技巧，对于每个大小为 $H\times W$ 的特征图切片，其局部差异的计算量只需 $8HW$ 次减法，大幅低于 MPCM 中所采用的滤波技巧的计算量。在 SIRST 数据集所附带的工具包中，相同条件下 MPCM 的 MATALB 实现，在采用了该技巧后计算速度可以提升大约 15%，从平均 2.67 帧每秒到 3.07 帧每秒。相应的，给定膨胀因子 d，特征图 $\mathbf{F}$ 上的局部对

比度（Dilated Local Contrast, DLC）度量 LC($\mathbf{F}, d$) 可以表示为

$$\mathrm{DLC}(\mathbf{F}, d) = \min_{(x,y)\in\Omega} \left\{ \mathbf{D}^{(x,y)} \right\}. \tag{6.4}$$

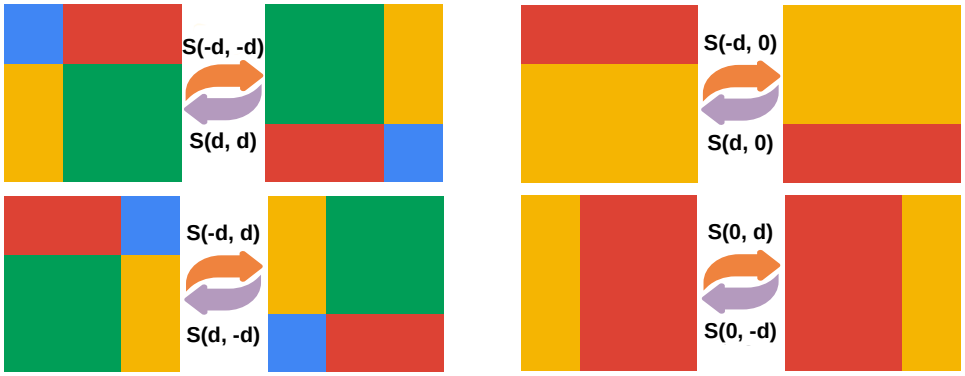


图 6.2　循环移位方案示意图

6.2.2　两阶段的多尺度局部对比度特征融合

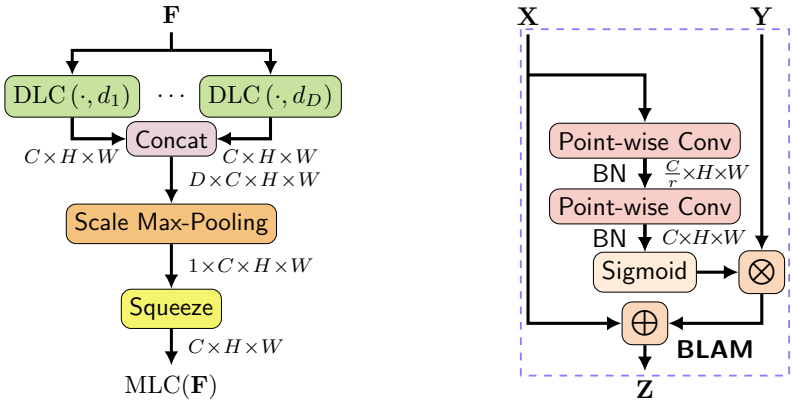
局部对比度度量的基础假设是小目标大小与图像块大小相近。然而，不同场景下小目标尺度变化较大，单一尺度下的度量无法有效捕获未知目标真实的对比度，因此融合多尺度的度量结果对于红外小目标检测至关重要。在基于局部对比度度量的传统方法中，常用的方式是通过变动图像块大小在原灰度图像上实现不同尺度的度量。在卷积神经网络的特征图上，可以借助对膨胀因子 d 选取不同的值，在特征图上实现不同尺度的局部对比度度量。给定一组膨胀因子 $\{d_1, d_2, \cdots, d_D\}$，基于特征图 $\mathbf{F}$ 的多尺度局部对比度（Multi-scale Local Contrast, MLC）特征定义如下：

$$\mathrm{MLC}(\mathbf{F}) = \mathrm{Squeeze}\left(\mathrm{SMP}\left(\mathrm{Concat}\left(\mathrm{DLC}\left(\mathbf{F}, d_1\right), \cdots, \mathrm{DLC}\left(\mathbf{F}, d_D\right)\right)\right)\right) \tag{6.5}$$

式中，SMP 表示尺度最大池化（Scale Max-Pooling, SMP），该运算将拼接后不同尺度的局部对比度特征沿着尺度维度进行最大池化。Squeeze 运算将特征张量中长度为一的维度剔除，以保证 $\mathrm{MLC}(\mathbf{F}) \in \mathbb{R}^{C\times H\times W}$ 与特征图 $\mathbf{F}$ 大小相同。MLC 模块的流程图如图 6.3(a) 所示。

与在二维的原始灰度图像上进行局部对比度度量的传统检测方法不同，ALCNet 中的局部对比度度量在三维的特征图上进行，还可以利用卷积神经网络本征的多尺度特征金字塔进行第二阶段的局部对比度特征融合。通过迭代的方式逐渐将分辨率低但富含语义信息的高层特征与反映图像细节结构的低层特征相融合，可以得到经过第二阶段跨层融合后的多尺度局部对比度特征：

$$\mathrm{M^2LC}(f) = \uplus\left(\mathrm{MLC}(\mathbf{F}^{(1)}), \uplus\left(\cdots, \uplus\left(\mathrm{MLC}(\mathbf{F}^{(L-1)}), \mathrm{MLC}(\mathbf{F}^{(L)})\right)\right)\right), \tag{6.6}$$



(a) MLC 模块　　(b) BLAM 模块

图 6.3　同层和跨层的多尺度局部对比度特征融合模块

式中，$\mathrm{M}^2\mathrm{LC}(f) \in \mathbb{R}^{C\times H\times W}$ 为最终经过两阶段多尺度融合后的局部对比度特征图，f 表示给定的待检测红外图像。$\uplus$ 表示跨层特征融合的方式，常用的方式是 FPN 所采用的逐元素相加[104]。

考虑到局部对比度模块实质上已经显式打破有效感受野的限制，编码了较为长程的特征上下文之间的交互关系，跨层的注意力调制不再像纯数据驱动的 ABAMNet 那样需要特征图的全局统计量作为自顶向下调制的指导信息。因此，本章采用自底向上的局部注意力调制（Bottom-up Local Attentional Modulation, BLAM）模块作为跨层的多尺度局部对比度特征的融合方案，具体如图 6.3(b) 所示。给定低层特征图 $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ 和相对更高层的特征图 $\mathbf{Y} \in \mathbb{R}^{C\times H\times W}$，BLAM 模块所产生的融合特征 $\mathbf{Z} \in \mathbb{R}^{C\times H\times W}$ 可以表示为：

$$\mathbf{Z} = \mathbf{X} + \mathbf{L}(\mathbf{X}) \otimes \mathbf{Y} \tag{6.7}$$

在本章方法中，将 $\mathbf{X}$ 和 $\mathbf{Y}$ 替换为各自经由 MLC 模块的输出 $\mathrm{MLC}(\mathbf{X})$ 和 $\mathrm{MLC}(\mathbf{Y})$，第二阶段的多尺度局部对比度特征融合结果可以表示为：

$$\mathbf{Z}' = \mathrm{MLC}(\mathbf{X}) \uplus \mathrm{MLC}(\mathbf{Y}) = \mathrm{MLC}(\mathbf{X}) + \mathbf{L}\left(\mathrm{MLC}(\mathbf{X})\right) \otimes \mathrm{MLC}(\mathbf{Y}) \tag{6.8}$$

基于 MLC 模块与 BLAM 模块，ALCNet 的整体架构如图 6.4 所示，其中骨干网络仍然采用第三章表 3.1 所示的 ResNet-20[204]。

6.3　实验分析与讨论

为了验证 ALCNet 在网络结构设计上的合理性和有效性，本节将进行详细的消融实验，具体探索以下问题：

（1）问题一：从卷积神经网络的角度看，ALCNet 相当于将局部对比度度量作为一个特殊的非线性特征变换模块移植到网络中。相较于纯数据驱动的网络，在给定相同网络参数的情况下，

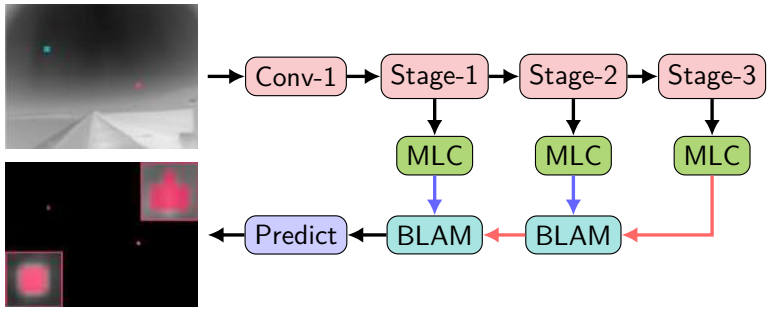


图 6.4　ALCNet 网络架构示意图

局部对比度度量模块是否能够提升网络的检测效果?

（2）问题二：从局部对比度度量的角度看，ALCNet 其实是一个两阶段的多尺度度量模型，其中同层多尺度融合、跨层多尺度融合各自的重要性如何?

（3）问题三：通常，深度卷积网络中的跨层特征融合多是以自顶向下的方式将高层特征中的语义信息集成到低层特征中，但是 ALCNet 则是使用了一条反向的自底向上的调制通路。在给定相同网络参数量的情况下，相较于其他跨层融合方式，自底向上的局部注意力调制是否能够获得更好的红外小目标检测性能?

最后，本节还将所提出的 ALCNet 与现有的多种红外小目标检测方法进行了对比以验证其性能。实验在 SIRST 数据集上进行，消融实验与对比实验的方法选取、参数设置、定量评价指标与第三章相同。

6.3.1　消融实验与分析

为了更好地理解 ALCNet，本小节通过去除或者替换 ALCNet 中的某些部分来构建一些用于消融实验的网络，以单独验证具体各个模块的作用。依据同层对比度度量和跨层对比度融合方式上的差异，表 6.1罗列了消融实验中用到的网络架构。其中，自底向上全局注意力调制（Bottom-up Global Attentional Modulation, BGAM）模块和自顶向下局部注意力调制（Top-down Local Attentional Modulation, TLAM）模块分别如图 6.5(a) 和图 6.5(b) 所示。

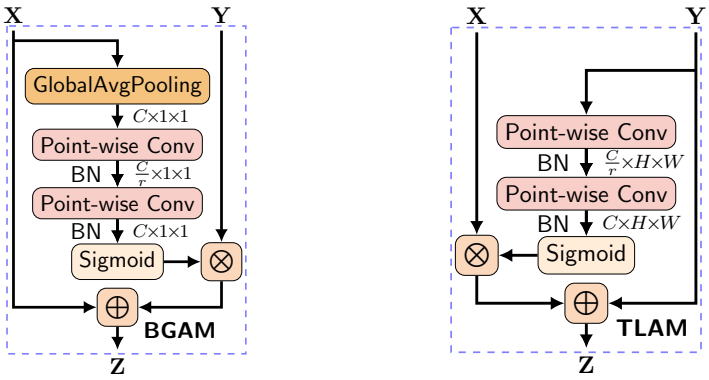
6.3.1.1　编码局部对比度的重要性

对于问题一，首先比较 FPN 和 DLC-FPN，这两者的差别只在于是否对特征金字塔每层的输出特征进行单尺度的局部对比度编码。图 6.6(a) 和图 6.6(b) 展示了两者在 IoU 和 nIoU 指标上的性能比较，其中 DLC-FPN 的膨胀因子 d 为 13。从中可以看到，在各个网络深度设置下，DLC-FPN 的小目标检测性能始终以较大的幅度好于 FPN，这表明在网络中融入局部对比度先验能够较为显著地提升红外小目标的检测效果。此外，不管是 IoU 还是 nIoU，$b = 3$ 时的 DLC-FPN 可以取得与 $b = 4$ 时的 FPN 大体相当的性能，这也表明编码局部对比度先验可以使得网络更加

高效，更符合平台计算能力和容量受限的真实需求场景。

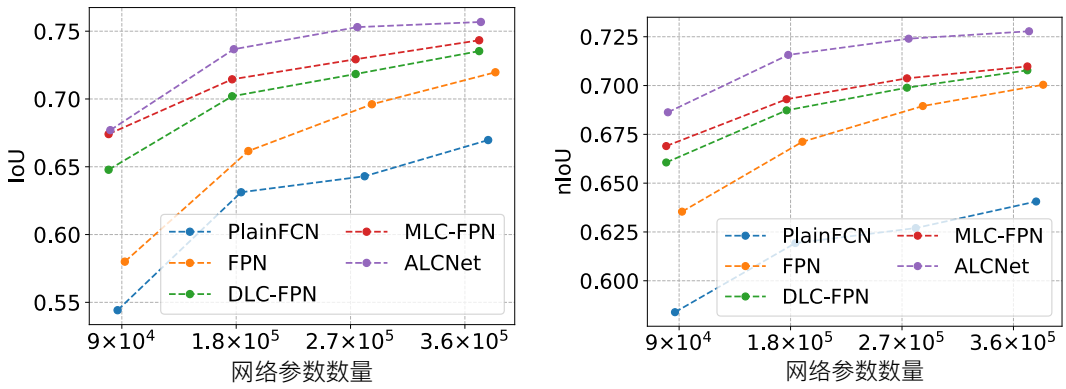
表 6.1　消融实验所构建网络的同层和跨层对比度度量方案

同层对比度度量	跨层对比度融合	消融网络架构
无	无	PlainFCN
无	相加	FPN
单尺度（DLC）	相加	DLC-FPN
多尺度（MLC）	相加	MLC-FPN
多尺度（MLC）	尺度最大池化（SMP）	SMP-FPN
多尺度（MLC）	自顶向下局部注意力调制（TLAM）	TLAM-FPN
多尺度（MLC）	自底向上全局注意力调制（BGAM）	BGA-FPN
多尺度（MLC）	自底向上局部注意力调制（BLAM）	ALCNet



(a) BGAM 模块　　(b) TLAM 模块

图 6.5　消融实验所需的跨层局部对比度特征融合方案

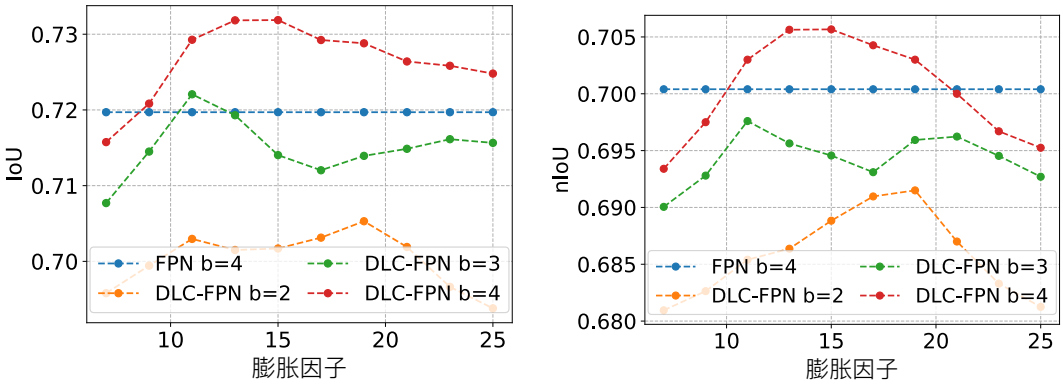


(a) IoU 比较　　(b) nIoU 比较

图 6.6　消融实验中各网络架构的 IoU 和 nIoU 性能比较

6.3.1.2 多尺度局部对比度特征融合的重要性

对于问题二，首先研究在同层特征图上融合多个尺度的局部对比度特征的重要性。图 6.7 展示了 DLC-FPN 在不同网络深度和不同膨胀因子下的 IoU 和 nIoU 比较。对比 $b = 4$ 时的 FPN 和 DLC-FPN，可以看到在大多数情况下，DLC-FPN 的性能均优于 FPN。但是当膨胀因子 d 过小或过大时，DLC-FPN 的性能反而会更差，尤其是在 nIoU 指标上表现得更为明显。$b = 3$ 和 $b = 2$ 时的 IoU 和 nIoU 曲线并不只有一个峰值，这表明即使选择最佳的膨胀因子，单一尺度的度量也很难精确抓取所有红外小目标的局部对比度信息。为了解决这个问题，ALCNet 在同一特征图上采用具有不同膨胀因子的并行多分支结构来提取多尺度的局部对比度度量。在本实验中，MLC-FPN 的膨胀因子为 13 和 17。对比图 6.6(a) 和图 6.6(b) 上的 DLC-FPN 和 MLC-FPN 曲线，可以看到 MLC-FPN 的性能始终好于 DLC-FPN，特别是在网络相对较浅的时候。这表明，在网络其他结构保持相同的情况下，多尺度的局部对比度度量模块能够在特征图上更好地编码红外小目标。



(a) IoU 比较　　(b) nIoU 比较

图 6.7　不同网络深度和不同膨胀因子下 DLC-FPN 的 IoU 和 nIoU 比较

此外，图 6.7 同样还展示了进一步融合特征金字塔中跨层的局部对比度特征的重要性。对比全卷积网络（Fully Convolutional Networks, FCN）和 PlainFCN 两者的曲线可以看到，即使没有局部对比度度量模块，在纯粹的数据驱动方法中，跨层的特征融合对于红外小目标检测也是非常重要。对比 MLC-FPN 和 ALCNet 的曲线可以看到，ALCNet 的性能始终显著高于 MLC-FPN。特别是在 nIoU 指标中，$b = 2$ 时的 ALCNet 可以取得比 $b = 4$ 时的 MLC-FPN 更好的性能，而只需后者大约 50% 的参数量。这表明，对于红外小目标检测来说，相较于增加网络深度，提升检测性能更高效的方式是设计更符合红外小目标特点的跨层特征融合方式。不同于通用的目标检测任务主要依赖于目标本征特征，检测红外小目标更多依赖的是目标与背景上下文之间的关系，因此多尺度的特征融合非常关键。

6.3.1.3 跨层特征融合方案的重要性

对于问题三，表 6.2 展示了本章所采用的 BLAM 模块以及其他跨层特征融合方案在 IoU 和 nIoU 指标上的比较，从中可以看出：1）对比 FPN 和 Max-FPN，可以看到采用相加方式融合的 FPN 在 IoU 指标上表现更好，而采用尺度最大值池化的 Max-FPN 则是在 nIoU 指标上更好，但是这两者的性能总体都不如采用通道注意力调制的其余跨层融合方案。2）对比 BGA-FPN 与 ALCNet，两者均采用自底向上的注意力调制，差别只在于前者使用全局通道注意力机制，后者使用了局部通道注意力机制。ALCNet 的效果在所有实验设定下都较为显著地好于 BGA-FPN，这表明对于红外小目标来说，更应该采用局部细节信息作为特征调制的信息参考。3）在同样采用局部通道注意力机制的情况下，TLA-FPN 以自顶向下的方式将高层语义嵌入低层特征中，而 ALCNet 则是反向地使用低层特征作为参考来调制高层特征。从表 6.2 中可以看到，随着网络深度的加深，自底向上调制的 ALCNet 的 IoU 和 nIoU 指标逐渐好于自顶向下调制的 TLA-FPN。两者对比的结果表明，在相同参数与计算量的情况下，使用自底向上的调制方式更有利于小目标检测。因此，综合来看，对于红外小目标检测，本章所采用的自底向上局部通道注意力调制方案性能相对最好。

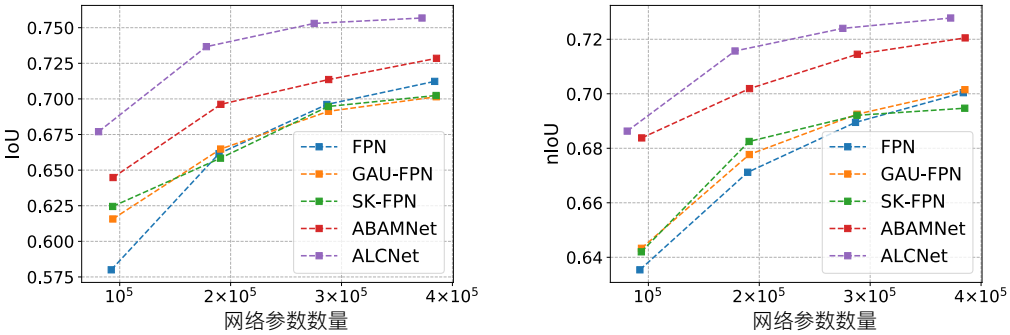
表 6.2　不同跨层特征融合方案的 IoU 和 nIoU 比较

上下文尺度	调制方向	公式	网络	IoU				nIoU			
				$b=1$	$b=2$	$b=3$	$b=4$	$b=1$	$b=2$	$b=3$	$b=4$
无	无	$\mathbf{X}+\mathbf{Y}$	FPN	0.674	0.713	0.729	0.744	0.669	0.691	0.702	0.710
		$\max(\mathbf{X},\mathbf{Y})$	Max-FPN	0.665	0.713	0.722	0.734	0.674	0.698	0.706	0.712
全局	自底向上	$\mathbf{X}+\mathbf{G}(\mathbf{X})\otimes\mathbf{Y}$	BGA-FPN	0.676	0.714	0.731	0.736	0.679	0.698	0.704	0.711
局部	自顶向下	$\mathbf{L}(\mathbf{X})\otimes\mathbf{X}+\mathbf{Y}$	TLA-FPN	**0.688**	0.729	0.750	0.753	**0.688**	0.708	0.722	0.718
	自底向上	$\mathbf{X}+\mathbf{L}(\mathbf{X})\otimes\mathbf{Y}$	ALCNet	0.677	**0.737**	**0.753**	**0.757**	0.686	**0.716**	**0.724**	**0.728**

6.3.2 方法对比与分析

本小节将 ALCNet 与其他多种模型驱动的红外小目标检测方法和其他深度网络模型进行了对比，包括稳态多子空间学习（Stable Multi-Subspace Learning, SMSL）方法[155]、基于小面核与随机游走（Facet Kernel and Random Walker, FKRW）的方法[159]、多尺度块对比度度量（Multi-scale Patch-based Contrast Measurement, MPCM）方法[10]，红外块图像（Infrared Patch-Image, IPI）模型方法[12]、基于奇异值部分和的非负 IPI 模型方法[148]（Non-Negative IPI Model via Partial Sum Minimization of Singular Values, NIPPS）、重加权红外块张量（Reweighted Infrared Patch-Tensor, RIPT）模型方法[13]、特征金字塔网络[104]（Feature Pyramid Network，FPN）、基于选择核的特征

金字塔网络[188]（Selective Kernel Feature Pyramid Network, SK-FPN）、基于全局注意力上采样的特征金字塔网络[189]（Global Attention Upsample Feature Pyramid Network, GAU-FPN）以及第三章所提出的双向非对称注意力调制网络（Asymmetric Bidirectional Attentional Modulation Network, ABAMNet）。图 6.8 展示了在网络深度逐渐增加的情况下，ALCNet 与 FPN、GAU-FPN、SK-FPN、ABAMNet 在 IoU 和 nIoU 上的比较。从中可以看到，在各个深度网络的参数量相近的情况下，ALCNet 的红外小目标检测性能显著好于纯数据驱动的 ABAMNet 以及其他深度网络。考虑到 ABAMNet 同样使用了注意力调制模块，而 ACLNet 表现更佳，特别是在 IoU 指标上，$b = 2$ 时的 ALCNet 好于 $b = 4$ 时的其余网络，这表明在深度网络中嵌入局部对比度先验对于检测红外小目标非常重要，是一种比增加网络深度更为高效的途径。表 6.3 展示了上述深度网络与 SMSL、FKRW、MPCM、IPI、NIPPS、RIPT 等方法在 SIRST 数据集上的 IoU 和 nIoU 比较。其中，ALCNet 取得了最好的表现。特别是相较于本章对比度度量参考的 MPCM，IoU 指标提升了 100% 多，nIoU 提升了 62%，这显示了将局部对比度度量方法网络化的巨大潜力。



(a) IoU 比较　　(b) nIoU 比较

图 6.8　ALCNet 与其他深度网络在 SIRST 数据集上的性能比较

表 6.3　ALCNet 与其他十种方法的定量评价指标比较

方法	SMSL	FKRW	MPCM	IPI	NIPPS	RIPT	FPN	SK-FPN	GAU-FPN	ABAMNet	ALCNet
IoU	0.081	0.268	0.357	0.466	0.473	0.146	0.720	0.702	0.701	0.731	**0.757**
nIoU	0.279	0.339	0.445	0.607	0.602	0.245	0.700	0.695	0.701	0.721	**0.728**

最后，图 6.9 比较了 ALCNet 与其他五种方法的 ROC 曲线。趋于稳定后，ALCNet 与 ABAMNet 的性能接近，但是 ALCNet 有着比 ABAMNet 更陡峭的斜率，在虚警率小于 0.006% 时，ALCNet 可以获得更高的检测率。相较于 GAU-FPN，不管是虚警率还是检测率，ALCNet 始终都表现更好，也好于 MPCM、RIPT、NIPPS 这些模型驱动的方法。综合 IoU、nIoU、ROC 等指标上的表现，本章方法相较于对比方法能够稳定地取得更好的红外小目标检测效果。

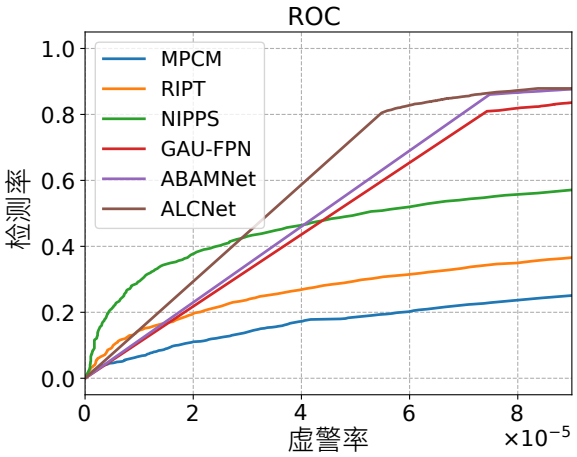


图 6.9　ALCNet 与其他方法的 ROC 比较

6.4　本章小结

针对红外小目标检测任务，本章通过在特征金字塔网络中嵌入多尺度的局部对比度度量模块，实现了模型驱动方法与数据驱动方法的融合与统一，使得所提出的 ALCNet 能够同时利用标记数据和领域知识来提升检测性能。首先将传统基于图像块的对比度度量改造为基于特征图和膨胀因子的对比度度量，并在此基础上，利用特征图的循环移位构造了一种在卷积网络中快速计算局部对比度的方法。ALCNet 采用同层和跨层两阶段的方式融合多尺度的局部对比度特征。在同一特征图上，使用不同膨胀因子得到的局部对比度特征图先是通过尺度维度上的最大池化实现特征融合，随后自底向上的局部通道注意力调制模块会将这些不同层的特征图依次融合从而得到最终的多尺度对比度特征。详细的消融实验和对比实验结果表明，在深度网络中嵌入红外小目标的局部对比度先验可以大幅提升检测效果。此外，相较于提升网络深度，依据小目标的特点设计针对性的多尺度特征融合方案能够更为高效地提升检测效果。

第七章　总结与展望

7.1　工作总结

在本征特征极度稀缺的情况下，如何有效区分真实目标与背景干扰是红外小目标检测的核心问题。由于缺少标准数据集、场景复杂多变、背景杂波干扰大等原因，红外小目标检测的性能仍有较大的进步空间。本文沿着从模型驱动到数据驱动再到模型驱动的深度学习的发展脉络，针对该问题开展了研究和探索。具体而言，从建立更符合红外小目标特点的低秩稀疏模型开始，构建了具有高质量标注的小目标数据集，并探索了新型的注意力机制及其在深度网络中更多样的应用，最后实现了深度神经网络与传统红外小目标检测方法的融合。全文的具体贡献如下：

- 在第二章中，针对刻画目标的稀疏约束无法区分稀疏的真实目标与同样相对稀疏的背景干扰物这一问题，构建了重加权红外块张量模型。为了更好地保留图像块的空间相关性，该模型以三维张量形式堆叠被采样的图像块，利用张量鲁棒低秩恢复模型来分离红外小目标与图像背景。此外，通过设计反映图像边缘强度的局部结构权重，在迭代过程中，稀疏度相似的小目标与强边缘残留会被分配不同的收缩阈值，从而实现对目标和背景干扰有选择性的抑制。考虑到小目标检测的实际情况，以目标块张量的稀疏度为依据设计了新的终止条件，并且采用基于元素重加权的稀疏性增强权重来减少模型所需的迭代次数。相比于其他低秩稀疏分解方法，该模型提高了对复杂云杂波干扰的抑制能力，同时也大幅降低了背景目标分离所需的时间。
- 在第三章中，针对模型驱动方法缺少语义判别能力、对场景变化不够鲁棒的问题，构建了一个人工像素级标注的单帧红外小目标检测基准数据集，旨在通过深度网络以端到端的方式从标记数据中自动学习红外小目标的语义特征表示。考虑到红外小目标本征特征稀缺、容易被背景特征淹没的特点，设计了一个双向非对称的注意力调制网络，分别利用全局通道注意力模块和局部通道注意力模块实现高层语义特征和低层细节特征之间的彼此双向调制，从而能够更为有效地保存红外弱小目标，避免其被临近的背景杂波特征所淹没。相比于采用其他特征调制模块的网络，该网络大幅提升了红外小目标检测性能。
- 在第四章中，针对目前的激活单元无法根据上下文信息对特征进行选择性激活这一不足，提出了注意力激活单元，在非线性门控函数的框架下，实现了对激活单元和注意力机制的统一。注意力激活单元不仅能够完成在网络中引入非线性这一基本功能，还可以依据聚合的通道和空间特征上下文，实现对于特征图的动态精炼。通过逐层替换网络中原有的激活函数，还可以构建出基于注意力激活单元的全注意力网络，在低层网络中及早抑制无关特

征、强调相关特征，从而实现对高层语义更为有效的编码。此外，出于在更大规模的数据集上验证小目标检测算法的目的，本章还利用可见光遥感图像，构建了与红外小目标图像具有相似特点的弱小冰山检测数据集。给定相同的宿主网络，相比于其他激活单元，注意力激活单元可以大幅提升各类网络在多个计算机视觉任务上的性能。

- 在第五章中，针对深度网络中相加、拼接等线性特征融合方式无法自适应调整融合权重的问题，给出了一个适用于短跳连接、长跳连接、Inception 模块等多种场景的动态特征融合框架——注意力特征融合。为了克服特征之间语义和尺度的不连续性，构建了一个多尺度通道注意力模块，通过聚合不同尺度的特征上下文匹配不同尺度的物体，有效避免了融合过程过于强调大目标、弱化小目标的现象。在此基础上，迭代注意力特征融合采用另一层注意力特征融合改善初始注意力模块的输入，能够进一步生成更高质量的融合权重。相比于增加网络深度，作为一个即插即用的模块，注意力特征融合可以用来替换各类基准网络中原有的特征融合操作，赋予网络动态选择特征融合权重的能力，从而更为高效地提高网络在各个视觉任务上的性能。

- 在第六章中，针对模型驱动方法难以精确建模而数据驱动方法需要大量标记数据的问题，提出了注意力局部对比度网络，旨在通过同时利用标记数据和领域知识来提升检测性能。为了克服红外小目标本征特征稀缺的问题，将传统依据先验知识建模的局部对比度度量方法模块化为网络中具有特定物理意义的非线性特征变换层，并将其嵌入特征金字塔中，用以捕获目标特征与相邻背景上下文之间的交互关系。该网络采用特征图上的循环移位技巧实现了同层特征上的多尺度局部对比度度量，不仅大幅减少了冗余计算量，同时也使得其仍能够以端到端的方式被训练和测试。对于跨层的对比度特征融合，采用自底向上的局部通道注意力调制模块以保存高层特征中的弱小目标。从模型驱动方法的角度看，注意力局部对比度网络其实是将传统模型中过于简单的均值、最大值等特征替换为网络从标注数据中学到的语义特征，从而大幅提高了模型对于真实目标与背景干扰物的判别能力。

7.2 未来展望

长久以来，由于缺少本征特征，单帧的红外小目标检测常常被作为一个低层视觉问题看待，通常采用滤波、局部对比度度量、矩阵或张量的低秩稀疏分解等建模方式来实现背景抑制和目标增强。由于真实场景中的图像场景高度复杂，单帧图像也不具有序列图像的时空冗余性，单帧图像中的红外小目标实际上很难满足全局唯一的稀疏性、显著性假设，这些传统的模型驱动方法面临着难以精确建模、特征判别性不足、超参数对场景变化敏感等挑战。本文通过构建高质量标注的红外小目标基准数据集，将小目标检测问题看作是一个带有一定低层视觉特点的高层语义任务，在深度学习的框架内，利用注意力机制以及领域知识以特定方式聚合目标的特征

上下文信息，用于克服小目标本征特征不足的问题。然而，相对于人类视觉系统中复杂的注意力机制，在小目标检测这个大方向上，本文工作仅仅是一个微小的开始。结合现实应用场景，红外小目标检测仍有以下几个问题值得进一步探索：

（1）**小样本问题：**红外小目标数据收集难度较大，特别是对于单帧红外小目标检测而言，为了避免机器学习模型通过机械记忆特定场景中的目标或者图像背景，在场景和目标变化不大的情况下，从一个含有红外小目标的图像序列中通常仅能选取一幅单帧图像用于构建数据集。本论文构建的 SIRST 数据集选取了 400 余个红外序列中的代表性图像，但是相比于同类的可见光数据集规模仍然较小。因此，单帧红外小目标检测问题还是一个典型的小样本问题，这极大地限制了深度学习方法在该任务上的表现，阻碍了红外小目标检测性能的进一步提升。然而，在数据获取更为开放、容易的可见光波段遥感影像中，存在诸如本文构建的 DiskoBay 之类的数据集，其目标特性与红外小目标高度相似，在这些更大规模的小目标数据集上学习到的特征，对于红外小目标检测任务也应当具有一定程度的泛化性。因此，充分利用好迁移学习，通过对其他成像波段下的源领域数据的学习降低深度神经网络对于红外小目标数据的样本数量要求，有望进一步提高红外小目标检测的性能。

（2）**低秩稀疏分解模型的深度展开：**论文第六章通过将传统模型驱动方法中的局部对比度度量方法模块化，作为特定的非线性特征变换层嵌入深度网络，取得了比其他章节的纯数据驱动方法更好的性能，显示了模型驱动的深度学习在红外小目标检测中的巨大潜力。在 SIRST 数据集上的评估中，各类基于低秩稀疏分解模型的方法能够普遍取得比局部对比度度量方法更好的性能，将这些模型进行深度展开有望取得比第六章方法更好的性能，但是由于存在奇异值分解运算，这些方法很难直接被网络化。因此，设计无需奇异值分解的低秩稀疏分解模型并将其进行深度展开对于进一步提升红外小目标检测性能具有重要意义。

（3）**场景上下文推理：**图像的场景上下文包含了丰富的语义信息，特别是在目标较小、本征特征不明显的情况下，充分利用好图像场景中的上下文信息能够大幅提高深度神经网络在各类任务上的预测准确率。本文利用注意力机制通过对全局特征或者局部特征的上下文编码，实现了深度网络各模块中特征的自适应调制。然而，这种方式忽视了目标彼此之间以及目标与整体场景之间的关系，也就是图像场景中的上下文关联信息。在未来的工作中，可以考虑在深度网络中嵌入场景上下文的推理模块，综合利用目标视觉特征、场景上下文、目标相互关系等信息来联合解决复杂背景下的小目标检测问题。

（4）**网络模型轻量化：**本文主要致力于在复杂多变且干扰严重的图像背景下提高单帧红外小目标检测算法的准确率和鲁棒性。虽然各章节展示了通过设计新的终止条件、合理运用注意力机制可以在保持模型性能的前提下，减少相应的计算时间或者网络所需的参数数量，但距离实际工程中实时检测的性能要求仍有较大距离。为了能够在资源有限的嵌入式设备端快速检测红外小目标，需要通过参数剪枝、轻量级滤波设计等技巧对本文中的模型进行压缩与加速。

参考文献

[1] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[C]. Proceedings of 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 2014, 740–755.

[2] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211–252.

[3] Singh B, Davis L S. An Analysis of Scale Invariance in Object Detection - SNIP[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 3578–3587.

[4] Wei Zhang, Mingyu Cong, Liping Wang. Algorithms for Optical Weak Small Targets Detection and Tracking: Review[C]. Proceedings of International Conference on Neural Networks and Signal Processing, Nanjing, China, 2003, 643–647.

[5] Shrivastava A, Gupta A. Contextual Priming and Feedback for Faster R-CNN[C]. Proceedings of 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016, 330–348.

[6] Hu P, Ramanan D. Finding Tiny Faces[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 1522–1530.

[7] Kuznetsova A, Rom H, Alldrin N, et al. The Open Images Dataset V4[J]. International Journal of Computer Vision, 2020, 128(7): 1956–1981.

[8] Rivest J, Fortin R. Detection of Dim Targets in Digital Infrared Imagery by Morphological Image Processing[J]. Optical Engineering, 1996, 35(7): 1886–1893.

[9] Deshpande S D, Er M H, Venkateswarlu R, et al. Max-Mean and Max-Median Filters for Detection of Small Targets[C]. Proceedings of SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, Denver, CO, United States, 1999, 74–83.

[10] Wei Y, You X, Li H. Multiscale Patch-Based Contrast Measure for Small Infrared Target Detection[J]. Pattern Recognition, 2016, 58: 216–226.

[11] Chen C L P, Li H, Wei Y, et al. A Local Contrast Method for Small Infrared Target Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1): 574–581.

[12] Gao C, Meng D, Yang Y, et al. Infrared Patch-Image Model for Small Target Detection in a Single Image[J]. IEEE Transactions on Image Processing, 2013, 22(12): 4996–5009.

[13] Dai Y, Wu Y. Reweighted Infrared Patch-Tensor Model with Both Nonlocal and Local Priors for Single-Frame Small Target Detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3752–3767.

[14] Wang H, Zhou L, Wang L. Miss Detection vs. False Alarm: Adversarial Learning for Small Object Segmentation in Infrared Images[C]. Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, 8508–8517.

[15] 胡艳梅, 张明, 徐展, 等. 客体工作记忆对注意的导向作用: 抑制动机的影响 [J]. 心理学报, 2013, 45(2): 127–138.

[16] Mnih V, Heess N, Graves A, et al. Recurrent Models of Visual Attention[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2014, Montreal, Quebec, Canada, 2014, 2204–2212.

[17] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]. Proceedings of 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, 1–15.

[18] Bengio Y, Courville A C, Vincent P. Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798–1828.

[19] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2017, Long Beach, CA, USA, 2017, 5998–6008.

[21] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 7132–7141.

[22] Woo S, Park J, Lee J, et al. CBAM: Convolutional Block Attention Module[C]. Proceedings of 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018, 3–19.

[23] Kligvasser I, Shaham T R, Michaeli T. xUnit: Learning a Spatial Activation Function for Efficient Image Restoration[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 2433–2442.

[24] Elad M. Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing[M]. 2010.

[25] 练秋生, 石保顺, 陈书贞. 字典学习模型, 算法及其应用研究进展 [J]. 自动化学报, 2015, 41(2): 240–260.

[26] Donoho D L. Compressed Sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289–1306.

[27] 李树涛, 魏丹. 压缩传感综述 [J]. 自动化学报, 2009, 35(11): 1369–1377.

[28] 任越美, 张艳宁, 李映, 等. 压缩感知及其图像处理应用研究进展与展望 [J]. 自动化学报, 2014, 40(8): 1563–1575.

[29] Candès E J, Li X, Ma Y, et al. Robust Principal Component Analysis?[J]. Journal of the ACM, 2011, 58(3): 11:1–11:37.

[30] Zhou X, Yang C, Yu W. Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 597–610.

[31] Cao W, Wang Y, Sun J, et al. Total Variation Regularized Tensor RPCA for Background Subtraction From Compressive Measurements[J]. IEEE Transactions on Image Processing, 2016, 25(9): 4075–4090.

[32] Ji H, Huang S, Shen Z, et al. Robust Video Restoration by Joint Sparse and Low Rank Matrix Approximation[J]. SIAM Journal on Imaging Sciences, 2011, 4(4): 1122–1142.

[33] Fazel M. Matrix Rank Minimization with Applications[D]. Stanford University, 2002.

[34] Forster J, Schmitt N, Simon H U, et al. Estimating the Optimal Margins of Embeddings in Euclidean Half Spaces[J]. Machine Learning, 2003, 51(3): 263–281.

[35] Cai T T, Zhou W X. Matrix Completion via Max-Norm Constrained Optimization[J]. Electronic Journal of Statistics, 2016, 10(1): 1493–1525.

[36] 王斯琪, 冯象初, 张瑞, 等. 基于最大范数的低秩稀疏分解模型 [J]. 电子与信息学报, 2015, 37(11): 2601–2607.

[37] 杨永鹏, 杨真真, 李建林. 基于广义非凸鲁棒主成分分析的视频前背景分离 [J]. 仪器仪表学报, 2020, 41(1): 250–258.

[38] Mohan K, Fazel M. Iterative Reweighted Algorithms for Matrix Rank Minimization[J]. Journal of Machine Learning Research, 2012, 13: 3441–3473.

[39] Nie F, Huang H, Ding C H Q. Low-Rank Matrix Recovery via Efficient Schatten p-Norm Minimization[C]. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Ontario, Canada, 2012, 655–661.

[40] Hu Y, Zhang D, Ye J, et al. Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2117–2130.

[41] Gu S, Xie Q, Meng D, et al. Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision[J]. International Journal of Computer Vision, 2017, 121(2): 183–208.

[42] Xie Y, Gu S, Liu Y, et al. Weighted Schatten p-Norm Minimization for Image Denoising and Background Subtraction[J]. IEEE Transactions on Image Processing, 2016, 25(10): 4842–4857.

[43] 张倩颖, 谢晓振. 加权 Schatten 范数低秩表示的高光谱图像恢复 [J]. 光学 精密工程, 2019, 27(2): 421–432.

[44] 蒋明峰, 陆亮, 吴龙, 等. 基于加权 Schatten p 范数最小化的磁共振图像重构方法研究 [J]. 电子学报, 2019, 47(4): 784–790.

[45] Liu G, Lin Z, Yan S, et al. Robust Recovery of Subspace Structures by Low-Rank Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171–184.

[46] Liu G, Yan S. Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction[C]. Proceedings of 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011, 1615–1622.

[47] 柳欣, 钟必能, 张茂胜, 等. 基于张量低秩恢复和块稀疏表示的运动显著性目标提取 [J]. 计算机辅助设计与图形学学报, 2014, 26(10): 1753–1763.

[48] Håstad J. Tensor Rank Is NP-Complete[J]. Journal of Algorithms, 1990, 11(4): 644–654.

[49] Carroll J D, Chang J J. Analysis of Individual Differences in Multidimensional Scaling via An N-Way Generalization of "Eckart-Young" Decomposition[J]. Psychometrika, 1970, 35(3): 283–319.

[50] Tucker L R. Some Mathematical Notes on Three-mode Factor Analysis[J]. Psychometrika, 1966, 31(3): 279–311.

[51] Kilmer M E, Martin C D. Factorization Strategies for Third-Order Tensors[J]. Linear Algebra and its Applications, 2011, 435(3): 641–658.

[52] Lu C, Feng J, Chen Y, et al. Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 5249–5257.

[53] Goldfarb D, Qin Z T. Robust Low-Rank Tensor Recovery: Models and Algorithms[J]. SIAM Journal on Matrix Analysis and Applications, 2014, 35(1): 225–253.

[54] Tomioka R, Suzuki T. Convex Tensor Decomposition via Structured Schatten Norm Regularization[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2013, Lake Tahoe, Nevada, United States, 2013, 1331–1339.

[55] Kong H, Xie X, Lin Z. t-Schatten-p Norm for Low-Rank Tensor Recovery[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(6): 1405–1419.

[56] Chang Y, Yan L, Zhao X L, et al. Weighted Low-Rank Tensor Recovery for Hyperspectral Image Restoration[J]. IEEE Transactions on Cybernetics, 2020, 50(9): 1–15.

[57] Deng L, Li G, Han S, et al. Model Compression and Hardware Acceleration for Neural Networks: A Comprehensive Survey[J]. Proceedings of the IEEE, 2020, 108(4): 485–532.

[58] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]. Proceedings of the ACM International Conference on Multimedia (MM), Orlando, FL, USA, 2014, 675–678.

[59] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. Proceedings of 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, 1–14.

[60] Oseledets I V. Tensor-Train Decomposition[J]. SIAM Journal on Scientific Computing, 2011, 33(5): 2295–2317.

[61] Novikov A, Podoprikhin D, Osokin A, et al. Tensorizing Neural Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2015, Montreal, Quebec, Canada, 2015, 442–450.

[62] Kim Y, Park E, Yoo S, et al. Compression of Deep Convolutional Neural Networks for Fast and Low Power Mobile Applications[C]. Proceedings of 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2016, 1–11.

[63] Luo W, Li Y, Urtasun R, et al. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2016, Barcelona, Spain, 2016, 4898–4906.

[64] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 2818–2826.

[65] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 1800–1807.

[66] Parmar N, Ramachandran P, Vaswani A, et al. Stand-Alone Self-Attention in Vision Models[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2019, Vancouver, BC, Canada, 2019, 68–80.

[67] Sandler M, Howard A G, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 4510–4520.

[68] Wilson R A, Keil F. The MIT Encyclopedia of the Cognitive Sciences (MITECS)[M]. May, 1999.

[69] Treisman A M, Gelade G. A Feature-Integration Theory of Attention[J]. Cognitive Psychology, 1980, 12(1): 97–136.

[70] Wolfe J M, Cave K R, Franzel S L. Guided Search: An Alternative to the Feature Integration Model for Visual Search[J]. Journal of Experimental Psychology: Human Perception and Performance, 1989, 15(3): 419–433.

[71] Rodieck R W, Stone J. Analysis of Receptive Fields of Cat Retinal Ganglion Cells[J]. Journal of Neurophysiology, 1965, 28(5): 833–849.

[72] Koch C, Ullman S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry[J]. Human Neurobiology, 1985, 4(4): 219–227.

[73] 张豹, 黄赛. 工作记忆表征对视觉注意的引导机制 [J]. 心理科学进展, 2013, 21(9): 1578–1584.

[74] Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254–1259.

[75] Hamker F H. The Emergence of Attention by Population-Based Inference and Its Role in Distributed Processing and Cognitive Control of Vision[J]. Computer Vision and Image Understanding, 2005, 100(1): 64–106.

[76] Cerf M, Harel J, Einhaeuser W, et al. Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2007, Vancouver, British Columbia, Canada, 2007, 241–248.

[77] Harel J, Koch C, Perona P. Graph-Based Visual Saliency[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2006, Vancouver, British Columbia, Canada, 2006, 545–552.

[78] Borji A, Itti L. Exploiting Local and Global Patch Rarities for Saliency Detection[C]. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 2012, 478–485.

[79] Itti L, Koch C. Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems[C]. Proceedings of Human Vision and Electronic Imaging, San Jose, CA, United States, 1999, 473–482.

[80] Judd T, Ehinger K A, Durand F, et al. Learning to Predict Where Humans Look[C]. Proceedings of 2009 IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009, 2106–2113.

[81] Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach[C]. Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, Minnesota, USA, 2007, 1–8.

[82] Guo C, Ma Q, Zhang L. Spatio-Temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform[C]. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, USA, 2008, 1–8.

[83] Hou X, Harel J, Koch C. Image Signature: Highlighting Sparse Salient Regions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194–201.

[84] Liu T, Sun J, Zheng N, et al. Learning to Detect A Salient Object[C]. Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, Minnesota, USA, 2007, 1–8.

[85] Jiang M, Huang S, Duan J, et al. SALICON: Saliency in Context[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015,

1072–1080.

[86] Kümmerer M, Wallis T S A, Bethge M. Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics[C]. Proceedings of 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018, 798–814.

[87] Xie S, Tu Z. Holistically-Nested Edge Detection[C]. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, 1395–1403.

[88] Hou Q, Cheng M, Hu X, et al. Deeply Supervised Salient Object Detection with Short Connections[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(4): 815–828.

[89] Wang W, Shen J, Dong X, et al. Salient Object Detection Driven by Fixation Prediction[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 1711–1720.

[90] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, 11534–11542.

[91] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial Transformer Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2015, Montreal, Quebec, Canada, 2015, 2017–2025.

[92] Dai J, Qi H, Xiong Y, et al. Deformable Convolutional Networks[C]. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, 764–773.

[93] Gao H, Zhu X, Lin S, et al. Deformable Kernels: Adapting Effective Receptive Fields for Object Deformation[C]. Proceedings of 8th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 2020, 1–12.

[94] Park J, Woo S, Lee J, et al. BAM: Bottleneck Attention Module[C]. Proceedings of British Machine Vision Conference (BMVC) 2018, Newcastle, UK, 2018, 1–14.

[95] Bello I, Zoph B, Vaswani A, et al. Attention Augmented Convolutional Networks[C]. Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, 3286–3295.

[96] Fu J, Liu J, Tian H, et al. Dual Attention Network for Scene Segmentation[C]. Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 3146–3154.

[97] Carion N, Massa F, Synnaeve G, et al. End-to-End Object Detection with Transformers[C]. Proceedings of European Conference on Computer Vision (ECCV), 2020, 1–19.

[98] Wang X, Girshick R B, Gupta A, et al. Non-Local Neural Networks[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 7794–7803.

[99] Zhu Z, Xu M, Bai S, et al. Asymmetric Non-Local Neural Networks for Semantic Segmentation[C]. Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, 593–602.

[100] Cao Y, Xu J, Lin S, et al. GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond[C]. Proceedings of 2019 IEEE International Conference on Computer Vision Workshops, Seoul, Korea (South), 2019, 1971–1980.

[101] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection, 2019.

[102] Zhang S, Zhu X, Lei Z, et al. S3FD: Single Shot Scale-Invariant Face Detector[C]. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy,

[103] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. Proceedings of 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016, 21–37.

[104] Lin T, Dollár P, Girshick R B, et al. Feature Pyramid Networks for Object Detection[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 936–944.

[105] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015, 234–241.

[106] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499–1503.

[107] Li J, Liang X, Wei Y, et al. Perceptual Generative Adversarial Networks for Small Object Detection[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 1951–1959.

[108] Torralba A, Murphy K P, Freeman W T. Using the Forest to See the Trees: Exploiting Context for Visual Object Detection and Localization[J]. Communications of the ACM, 2010, 53(3): 107–114.

[109] 董维科, 张建奇, 邵晓鹏, 等. 检测红外弱小目标的对比滤波时域廓线算法 [J]. 西安电子科技大学学报, 2014, 41(1): 13–17.

[110] Salmond D J, Birch H. A Particle Filter for Track-Before-Detect[C]. Proceedings of American Control Conference, Arlington, VA, USA, 2001, 3755–3760.

[111] 李翠芸, 姬红兵. 新遗传粒子滤波的红外弱小目标跟踪与检测 [J]. 西安电子科技大学学报, 2009, 36(4): 619–623.

[112] 罗寰, 王芳, 陈中起, 等. 基于对称差分和光流估计的红外弱小目标检测 [J]. 光学学报, 2010, 30(6): 1715.

[113] 曹琦, 王德江, 张齐, 等. 红外点目标检测中的能量累积 [J]. 光学 精密工程, 2010, 18(3): 741–747.

[114] 徐剑峰, 吴一全, 周建江. 基于时域背景预测检测红外图像序列中的小目标 [J]. 中国图象图形学报, 2007, 12(9): 1598–1603.

[115] Reed I S, Gagliardi R M, Stotts L B. Optical Moving Target Detection with 3-D Matched Filtering[J]. IEEE Transactions on Aerospace and Electronic Systems, 1988, 24(4): 327–336.

[116] 朱金标, 李建勋. 匹配滤波器优化设计及在红外弱小点目标检测中的应用 [J]. 光学学报, 2009, 29(8): 2128–2133.

[117] Arce G, McLoughlin M. Theoretical Analysis of the Max/Median Filter[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1987, 35(1): 60–69.

[118] Tomasi C, Manduchi R. Bilateral Filtering for Gray and Color Images[C]. Proceedings of the Sixth IEEE International Conference on Computer Vision (ICCV), Bombay, India, 1998, 839–846.

[119] Bae T W, Sohng K I. Small Target Detection Using Bilateral Filter Based on Edge Component[J].

Journal of Infrared, Millimeter, and Terahertz Waves, 2010, 31(6): 735–743.

[120] Zhao Y, Pan H, Du C, et al. Bilateral Two-Dimensional Least Mean Square Filter for Infrared Small Target Detection[J]. Infrared Physics & Technology, 2014, 65: 17–23.

[121] 吴一全, 吴文怡. 基于变邻域变步长 LMS 背景预测检测红外小目标 [J]. 宇航学报, 2009, 30(2): 735–739.

[122] 曹瑛, 李志永, 卢晓鹏, 等. 基于自适应邻域双边滤波的点目标检测预处理算法 [J]. 电子与信息学报, 2008, 30(8): 1909–1912.

[123] 靳永亮, 王延杰, 刘艳滢, 等. 红外弱小目标的分割预检测 [J]. 光学 精密工程, 2012, 20(1): 171–178.

[124] Bai X, Zhou F. Analysis of New Top-Hat Transformation and the Application for Infrared Dim Small Target Detection[J]. Pattern Recognition, 2010, 43(6): 2145–2156.

[125] Bae T W, Zhang F, Kweon I S. Edge Directional 2D LMS Filter for Infrared Small Target Detection[J]. Infrared Physics & Technology, 2012, 55(1): 137–145.

[126] Kim S. Double Layered-Background Removal Filter for Detecting Small Infrared Targets in Heterogenous Backgrounds[J]. Journal of Infrared, Millimeter, and Terahertz Waves, 2011, 32(1): 79–101.

[127] Zhou A, Xie W, Pei J. Background Modeling in the Fourier Domain for Maritime Infrared Target Detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(8): 2634–2649.

[128] Deng H, Sun X, Zhou X. A Multiscale Fuzzy Metric for Detecting Small Infrared Targets Against Chaotic Cloudy/Sea-Sky Backgrounds[J]. IEEE Transactions on Cybernetics, 2019, 49(5): 1694–1707.

[129] Liu D, Cao L, Li Z, et al. Infrared Small Target Detection Based on Flux Density and Direction Diversity in Gradient Vector Field[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(7): 2528–2554.

[130] 吴一全, 尹丹艳, 纪守新. 基于双树复数小波和 SVR 的红外小目标检测 [J]. 仪器仪表学报, 2010, 31(8): 1834–1839.

[131] 吴一全, 纪守新, 占必超. 基于无下采样 Contourlet 变换和独立分量分析的红外弱小目标检测 [J]. 光学学报, 2011, 31(5): 82–89.

[132] 张翔, 张建奇, 秦翰林, 等. 基于对偶树复小波变换的红外弱小目标背景抑制 [J]. 光子学报, 2010, 39(9): 1672–1677.

[133] Wang X, Lv G, Xu L. Infrared Dim Target Detection Based on Visual Attention[J]. Infrared Physics & Technology, 2012, 55(6): 513–521.

[134] Kim S. Min-Local-LoG Filter for Detecting Small Targets in Cluttered Background[J]. Electronics Letters, 2011, 47(2): 105–106.

[135] Kim S, Yang Y, Lee J, et al. Small Target Detection Utilizing Robust Methods of the Human Visual System for IRST[J]. Journal of Infrared, Millimeter, and Terahertz Waves, 2009, 30(9): 994–1011.

[136] Shao X, Fan H, Lu G, et al. An Improved Infrared Dim and Small Target Detection Algorithm Based on the Contrast Mechanism of Human Visual System[J]. Infrared Physics & Technology, 2012, 55(5): 403–408.

[137] Han J, Ma Y, Huang J, et al. An Infrared Small Target Detecting Algorithm Based on Human

Visual System[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(3): 452–456.

[138] Qi S, Ma J, Tao C, et al. A Robust Directional Saliency-Based Method for Infrared Small-Target Detection Under Various Complex Backgrounds[J]. IEEE Geoscience and Remote Sensing Letters, 2013, 10(3): 495–499.

[139] Achanta R, Hemami S, Estrada F, et al. Frequency-Tuned Salient Region Detection[C]. Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, Florida, USA, 2009, 1597–1604.

[140] 胡暾, 赵佳佳, 曹原, 等. 基于显著性及主成分分析的红外小目标检测 [J]. 红外与毫米波学报, 2010, 29(4): 303–306.

[141] 易翔, 王炳健. 基于多特征的快速红外弱小目标检测算法 [J]. 光子学报, 2017, 46(6): 0610002.

[142] Deng H, Sun X, Liu M, et al. Small Infrared Target Detection Based on Weighted Local Difference Measure[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(7): 4204–4214.

[143] Deng H, Sun X, Liu M, et al. Entropy-Based Window Selection for Detecting Dim and Small Infrared Targets[J]. Pattern Recognition, 2017, 61: 66–77.

[144] Bai X, Bi Y. Derivative Entropy-Based Contrast Measure for Infrared Small-Target Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(4): 2452–2466.

[145] Cao X, Rong C, Bai X. Infrared Small Target Detection Based on Derivative Dissimilarity Measure[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(8): 3101–3116.

[146] Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183–202.

[147] Beck A, Teboulle M. Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems[J]. IEEE Transactions on Image Processing, 2009, 18(11): 2419–2434.

[148] Dai Y, Wu Y, Song Y, et al. Non-Negative Infrared Patch-Image Model: Robust Target-Background Separation via Partial Sum Minimization of Singular Values[J]. Infrared Physics & Technology, 2017, 81: 182–194.

[149] 张丛丛, 王欢, 楼竞. 基于加权核范数最小化的红外弱小目标检测 [J]. 华中科技大学学报: 自然科学版, 2017, 45(10): 31–37.

[150] Zhang T, Wu H, Liu Y, et al. Infrared Small Target Detection Based on Non-Convex Optimization with Lp-Norm Constraint[J]. Remote Sensing, 2019, 11(5): 559.

[151] Zhou F, Wu Y, Dai Y, et al. Detection of Small Target Using Schatten 1/2 Quasi-Norm Regularization with Reweighted Sparse Enhancement in Complex Infrared Scenes[J]. Remote Sensing, 2019, 11(17): 2058.

[152] Zhang L, Peng L, Zhang T, et al. Infrared Small Target Detection via Non-Convex Rank Approximation Minimization Joint l2,1 Norm[J]. Remote Sensing, 2018, 10(11): 1821.

[153] Zhu H, Liu S, Deng L, et al. Infrared Small Target Detection via Low-Rank Tensor Completion with Top-Hat Regularization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(2): 1004–1016.

[154] Zhang L, Peng Z. Infrared Small Target Detection Based on Partial Sum of the Tensor Nuclear Norm[J]. Remote Sensing, 2019, 11(4): 382.

[155] Wang X, Peng Z, Kong D, et al. Infrared Dim and Small Target Detection Based on Stable Multisubspace Learning in Heterogeneous Scene[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(10): 5481–5493.

[156] Boyd S P, Parikh N, Chu E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1–122.

[157] Dai Y, Wu Y, Song Y. Infrared Small Target and Background Separation via Column-Wise Weighted Robust Principal Component Analysis[J]. Infrared Physics & Technology, 2016, 77: 421–430.

[158] Wang X, Peng Z, Kong D, et al. Infrared Dim Target Detection Based on Total Variation Regularization and Principal Component Pursuit[J]. Image and Vision Computing, 2017, 63: 1–9.

[159] Qin Y, Bruzzone L, Gao C, et al. Infrared Small Target Detection Based on Facet Kernel and Random Walker[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9): 7104–7118.

[160] Yang C, Liu H, Liao S, et al. Small Target Detection in Infrared Video Sequence Using Robust Dictionary Learning[J]. Infrared Physics & Technology, 2015, 68: 1–9.

[161] He Y, Li M, Zhang J, et al. Small Infrared Target Detection Based on Low-Rank and Sparse Representation[J]. Infrared Physics & Technology, 2015, 68: 98–109.

[162] Wang X, Shen S, Ning C, et al. A Sparse Representation-Based Method for Infrared Dim Target Detection under Sea–sky Background[J]. Infrared Physics & Technology, 2015, 71: 347–355.

[163] Lu Y, Huang S, Zhao W. Sparse Representation Based Infrared Small Target Detection via an Online-Learned Double Sparse Background Dictionary[J]. Infrared Physics & Technology, 2019, 99: 14–27.

[164] Miller C W, Edelberg J A, Wilson M L, et al. Application of Rich Feature Descriptors to Small Target Detection in Wide-Area Persistent ISR Systems[C]. Proceedings of SPIE Defense + Security, Baltimore, Maryland, United States, 2014, 909202.

[165] Kim S. Analysis of Small Infrared Target Features and Learning-Based False Detection Removal for Infrared Search and Track[J]. Pattern Analysis and Applications, 2014, 17(4): 883–900.

[166] Bi Y, Bai X, Jin T, et al. Multiple Feature Analysis for Infrared Small Target Detection[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1333–1337.

[167] 吴双忱, 左峥嵘. 基于深度卷积神经网络的红外小目标检测 [J]. 红外与毫米波学报, 2019, 38(3): 371–380.

[168] 刘俊明, 孟卫华. 融合全卷积神经网络和视觉显著性的红外小目标检测 [J]. 光子学报, 2020, 49(7): 0710003.

[169] Zhang Z, Ganesh A, Liang X, et al. TILT: Transform Invariant Low-Rank Textures[J]. International Journal of Computer Vision, 2012, 99(1): 1–24.

[170] Everingham M, Gool L V, Williams C K I, et al. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303–338.

[171] Wang F, Jiang M, Qian C, et al. Residual Attention Network for Image Classification[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 6450–6458.

[172] Hu J, Shen L, Albanie S, et al. Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2018, Montréal, Canada, 2018, 9423–9433.

[173] Zhang H, Dana K J, Shi J, et al. Context Encoding for Semantic Segmentation[C]. Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, 7151–7160.

[174] Xu Z, Sun J. Model-Driven Deep-Learning[J]. National Science Review, 2018, 5(1): 22–24.

[175] Yang Y, Sun J, Li H, et al. ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(3): 521–538.

[176] Li Y, Tofighi M, Geng J, et al. Efficient and Interpretable Deep Blind Image Deblurring via Algorithm Unrolling[J]. IEEE Transactions on Computational Imaging, 2020, 6: 666–681.

[177] Weickert J. Coherence-Enhancing Diffusion Filtering[J]. International Journal of Computer Vision, 1999, 31(2): 111–127.

[178] Wu Z, Wang Q, Jin J, et al. Structure Tensor Total Variation-Regularized Weighted Nuclear Norm Minimization for Hyperspectral Image Mixed Denoising[J]. Signal Processing, 2017, 131: 202–219.

[179] Salmon J, Strozecki Y. Patch Reprojections for Non-Local Methods[J]. Signal Processing, 2012, 92(2): 477–489.

[180] Hadhoud M, Thomas D. The Two-Dimensional Adaptive LMS (TDLMS) Algorithm[J]. IEEE Transactions on Circuits and Systems, 1988, 35(5): 485–494.

[181] Wang C, Qin S. Adaptive Detection Method of Infrared Small Target Based on Target-Background Separation via Robust Principal Component Analysis[J]. Infrared Physics & Technology, 2015, 69: 123 – 135.

[182] Lin M, Chen Q, Yan S. Network In Network[C]. Proceedings of 2nd International Conference on Learning Representations (ICLR), Banff, AB, Canada, 2014, 1–10.

[183] Nair V, Hinton G E. Rectified Linear Units Improve Restricted Boltzmann Machines[C]. Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 2010, 807–814.

[184] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]. Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 2015, 448–456.

[185] He K, Zhang X, Ren S, et al. Identity Mappings in Deep Residual Networks[C]. Proceedings of 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016, 630–645.

[186] Chen Y, Xin Y. An Efficient Infrared Small Target Detection Method Based on Visual Contrast Mechanism[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(7): 962–966.

[187] Zhao M, Cheng L, Yang X, et al. TBC-Net: A real-time detector for infrared small target detection using semantic constraint[J]. arXiv preprint arXiv:2001.05852, 2019..

[188] Li X, Wang W, Hu X, et al. Selective Kernel Networks[C]. Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 510–519.

[189] Li H, Xiong P, An J, et al. Pyramid Attention Network for Semantic Segmentation[C]. Proceedings

of British Machine Vision Conference (BMVC) 2018, Newcastle, UK, 2018, 1–13.

[190] Rahman M A, Wang Y. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation[C]. Proceedings of 12th International Symposium on Visual Computing (ISVC), Las Vegas, NV, USA, 2016, 234–244.

[191] Duchi J C, Hazan E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121–2159.

[192] Klambauer G, Unterthiner T, Mayr A, et al. Self-Normalizing Neural Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2017, Long Beach, CA, USA, 2017, 971–980.

[193] Ramachandran P, Zoph B, Le Q V. Searching for Activation Functions[C]. Proceedings of 6th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 2018, 1–13.

[194] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2012, Lake Tahoe, Nevada, United States, 2012, 1106–1114.

[195] Maas A L, Hannun A Y, Ng A Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models[C]. Proceedings of ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013, 1–6.

[196] Hendrycks D, Gimpel K. Gaussian Error Linear Units (GELUs)[C]. Proceedings of ArXiv Preprint, 2016, 1–9.

[197] He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C]. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, 1026–1034.

[198] Yang Y, Sun J, Li H, et al. Deep ADMM-Net for Compressive Sensing MRI[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2016, Barcelona, Spain, 2016, 10–18.

[199] Agostinelli F, Hoffman M D, Sadowski P J, et al. Learning Activation Functions to Improve Deep Neural Networks[C]. Proceedings of 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, 1–9.

[200] Chen L, Zhang H, Xiao J, et al. SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 6298–6306.

[201] Sitzmann V, Martel J N P, Bergman A W, et al. Implicit Neural Representations with Periodic Activation Functions[C]. Proceedings of ArXiv Preprint, 2020, 1–11.

[202] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]. Proceedings of 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2016, 1–10.

[203] Drusch M, Del Bello U, Carlier S, et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services[J]. Remote Sensing of Environment, 2012, 120: 25–36.

[204] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 770–778.

[205] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, 1–9.

[206] Orhan E, Pitkow X. Skip Connections Eliminate Singularities[C]. Proceedings of 6th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 2018, 1–11.

[207] Singh B, Najibi M, Davis L S. SNIPER: Efficient Multi-Scale Training[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2018, Montréal, Canada, 2018, 9333–9343.

[208] Li Y, Chen Y, Wang N, et al. Scale-Aware Trident Networks for Object Detection[C]. Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, 6053–6062.

[209] Zhang Z, Zhang X, Peng C, et al. ExFuse: Enhancing Feature Fusion for Semantic Segmentation[C]. Proceedings of 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018, 273–288.

[210] Yuan W, Wang S, Li X, et al. A Skip Attention Mechanism for Monaural Singing Voice Separation[J]. IEEE Signal Processing Letters, 2019, 26(10): 1481–1485.

[211] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, 3431–3440.

[212] Shrivastava A, Sukthankar R, Malik J, et al. Beyond Skip Connections: Top-Down Modulation for Object Detection[C]. Proceedings of ArXiv Preprint, 2016, 1–11.

[213] Huang G, Liu Z, Maaten L, et al. Densely Connected Convolutional Networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 2261–2269.

[214] Srivastava R K, Greff K, Schmidhuber J. Training Very Deep Networks[C]. Proceedings of Annual Conference on Neural Information Processing Systems (NeurIPS) 2015, Montreal, Quebec, Canada, 2015, 2377–2385.

[215] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization[J]. International Journal of Computer Vision, 2020, 128(2): 336–359.

[216] Xie S, Girshick R B, Dollár P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, 5987–5995.

[217] Han D, Kim J, Kim J. Deep Pyramidal Residual Networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 6307–6315.

[218] Lu Z, Sreekumar G, Goodman E, et al. Neural Architecture Transfer[J]. arXiv e-prints, 2020. arXiv:2005.05859.

[219] Cubuk E D, Zoph B, Mane D, et al. AutoAugment: Learning Augmentation Strategies From Data[C]. Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 113–123.

[220] Yamada Y, Iwamura M, Akiba T, et al. ShakeDrop Regularization for Deep Residual Learning[J]. IEEE Access, 2019, 7: 186126–186136.

[221] Gao Z, Wang L, Wu G. LIP: Local Importance-Based Pooling[C]. Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, 3354–3363.

致　　谢

行文至此，我的求学生涯也到了该谢幕的时候。回首硕博七年多的南航时光，既漫长又短暂，学习成长之路不易，由衷地感谢一路走来帮助我、陪伴我的每一位良师益友，与你们的相识是我人生中弥足珍贵的记忆。

首先，衷心感谢我的恩师吴一全教授。百忙之中，吴老师仍然挤出时间悉心指导了本论文，不管是论文的选题，还是内容结构安排，乃至于语言的组织和撰写，吴老师都投入了大量的时间和精力。吴老师学识渊博、勤奋认真，是将我带入图像处理领域的引导者，能够有幸成为吴老师的学生，是我这一生中莫大的荣幸。在学术上，吴老师思维创新、洞察敏锐、治学态度严谨细致、工作态度精益求精且积极进取，不仅给了我许多启发和指导，也是我终身学习的榜样。在生活上，热心的吴老师也给予了我极大的关怀，通过分享人生经验和阅历，为我指引前进的道路。犹记得在我人生最低潮和迷茫的时候，每天下午吴老师会陪我一起去操场跑八公里，帮助我走出低谷。每每想到这些画面，我不经懊悔自己曾经的不理解、不懂事和任性。此外，还要特别感谢吴老师的耐心和宽容，感谢他容忍了我博士期间长达三年多都没有发表文章。正是这份宽容，让我有了构建和标注 SIRST 和 DiskoBay 这两个小目标数据集的时间，使得本文从模型驱动方法向检测性能更好的数据驱动方法转变成为可能。吴老师的言传身教是我在南航期间最大的收获，在此衷心感谢吴老师给予我的悉心指导和谆谆教诲。

同时，特别感谢明斯特大学的 Fabian Gieseke 教授、亚利桑那大学的 Kobus Barnard 教授和南京航空航天大学的陈松灿教授。Gieseke 教授是带我进入深度学习领域的引路人，深度参与了多篇论文的撰写、修改和投稿环节，付出了相当多的心血和精力，疫情期间仍旧积极帮我收集用于 ImageNet 数据集实验的 GPU 资源。在 Barnard 教授的指导下，我系统学习了概率图模型以及立体视觉中的视觉定位与三维结构恢复等理论，拓展了对于计算机视觉的认知和理解。同时，也是 Barnard 教授让我下定决心制作 SIRST 和 DiskoBay 数据集，并且将思考问题的角度从小目标检测拓展为计算机视觉中的普适问题。陈老师广阔深邃的学术视野、高超的学术造诣以及对于科学问题的执着追求，一直是我鞭策自己的动力。同时也是陈老师的前后两次的点拨，让我开始关注低秩稀疏分解和注意力机制，构成了本文技术路线的主要脉络。在此衷心感谢三位教授的热情帮助和鼓励。还要再次感谢陈松灿教授以及南京大学的吴建鑫教授，两位老师作为预评审专家提出了高屋建瓴又细致详实的修改意见，帮助我更好地完善了立项背景、研究动机、技术路线等关键内容的阐述，提高了本文的质量。

其次，还要感谢我的师兄宋昱博士，趣味相投、博学多闻、善于思辨的师兄是我人生路上不可多得的亦师亦友，不仅在学习上一直都帮我答疑解惑，在生活上也是形影不离的好朋友。感

谢周飞、刘忠林等师弟以及哥本哈根大学的 Stefan Oehmcke 博士，作为论文的主要合作者，他们不仅在学术上提供了富有启发性的讨论，还在生活上给予了我非常多的帮助。感谢李海杰、史骏鹏、孟天亮、倪康、钱宇雷、李仲年等同窗好友，从他们身上我受益良多，为我的求学生涯增添了许多色彩。特别是周飞师弟，四处奔波帮助我办理各种手续、流程，费时费力，非常辛苦。对于毕业、答辩流程，钱宇雷博士给予了我耐心又细致的提醒和指导。感谢西北工业大学的文载道副教授、广东第二师范学院的张菲菲讲师、西安交通大学的杨燕博士生、清华大学的杨成竹博士生等好友在我博士求学期间对我的热心帮助，不管是学术上还是生活上，认识你们都是我人生的幸事。

感谢含辛茹苦养育我的父母，一直坚定地鼓励我、支持我，给了我坚持下去的精神动力。也感谢柏林夏里特医学院的博士生安娜，对我多年来的陪伴以及给予的无微不至的关心与照顾。

最后，再次由衷地感谢所有曾经关心过、帮助过我的人，谢谢你们！

在学期间的研究成果及学术论文情况

攻读博士学位期间发表（录用）论文情况

[1] **Yimian Dai**, Yiquan Wu. Reweighted Infrared Patch-Tensor Model with Both Nonlocal and Local Priors for Single-Frame Small Target Detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3752-3767.
（SCI，中科院二区，IF: 3.827，代码：https://github.com/YimianDai/DENTIST）

[2] **Yimian Dai**, Yiquan Wu, Yu Song, Jun Guo. Non-Negative Infrared Patch-Image Model: Robust Target-Background Separation via Partial Sum Minimization of Singular Values[J]. Infrared Physics & Technology, 2017, 81: 182-194.
（SCI，中科院二区，IF: 2.379，代码：https://github.com/YimianDai/DENTIST）

[3] **Yimian Dai**, Yiquan Wu and Yu Song. Infrared Small Target and Background Separation via Column-Wise Weighted Robust Principal Component Analysis[J]. Infrared Physics & Technology, 2016, 77: 421-430.
（SCI，中科院二区，IF: 2.379，代码：https://github.com/YimianDai/DENTIST）

[4] **Yimian Dai**, Stefan Oehmcke, Fabian Gieseke, Yiquan Wu, Kobus Barnard. Attention as Activation[C]. 25th International Conference on Pattern Recognition (ICPR), 2020.
（已录用，Oral Presentation，代码：https://github.com/YimianDai/open-atac）

[5] **Yimian Dai**, Yiquan Wu, Fei Zhou, Kobus Barnard. Attentional Local Contrast Networks for Small Infrared Target Detection[J]. IEEE Transactions on Geoscience and Remote Sensing.
（已录用，SCI，中科院一区，Top 期刊，代码：https://github.com/YimianDai/open-alcnet）

[6] **Yimian Dai**, Yiquan Wu, Fei Zhou, Kobus Barnard. Asymmetric Contextual Modulation for Infrared Small Target Detection[C]. IEEE Winter Conference on Applications of Computer Vision, WACV 2021.
（已录用，代码：https://github.com/YimianDai/open-acm）

[7] **Yimian Dai**, Fabian Gieseke, Stefan Oehmcke, Yiquan Wu, Kobus Barnard. Attentional Feature Fusion[C]. IEEE Winter Conference on Applications of Computer Vision, WACV 2021.
（已录用，代码：https://github.com/YimianDai/open-aff）

[8] Yiquan Wu, **Yimian Dai**, Jiansheng Wu. A Fast Non-local Means Algorithm Based on Krawtchouk Moments[J]. Transactions of Tianjin University, 2015, 21(2): 104-112.（EI，导师一作）

[9] Yiquan Wu, **Yimian Dai**, Jiansheng Wu. An Improved Preprocessed Yaroslavsky Filter Based on

Shearlet Features[J]. Journal of Beijing Institute of Technology, 2016, 25 (1): 135-144.（EI，导师一作）

[10] Yu Song, Yiquan Wu, **Yimian Dai**. A New Active Contour Remote Sensing River Image Segmentation Algorithm Inspired from the Cross Entropy[J]. Digital Signal Processing, 2016, 48: 322-332.（SCI，中科院二区，IF: 2.871）

[11] Jun Guo, Yiquan Wu, **Yimian Dai**. Small Target Detection Based on Reweighted Infrared Patch-Image Model[J]. IET Image Processing, 2017, 12(1): 70-79.（SCI，中科院三区，IF: 1.995）

[12] Fei Zhou, Yiquan Wu, **Yimian Dai**, Peng Wang. Detection of Small Target Using Schatten 1/2 Quasi-Norm Regularization with Reweighted Sparse Enhancement in Complex Infrared Scenes[J]. Remote Sensing, 2019, 11(17): 2058.（SCI，中科院二区，IF: 4.509）

[13] Fei Zhou, Yiquan Wu, **Yimian Dai**, Peng Wang, Kang Ni. Graph-Regularized Laplace Approximation for Detecting Small Infrared Target Against Complex Backgrounds[J]. IEEE Access, 2019, 7: 85354-85371.（SCI，中科院二区，IF: 3.745）

[14] Fei Zhou, Yiquan Wu, **Yimian Dai**, Peng Wang, Kang Ni. Robust Infrared Small Target Detection via Jointly Sparse Constraint of l1/2-Metric and Dual-Graph Regularization[J]. Remote Sensing, 2020, 12 (12): 1963.（SCI，中科院二区，IF: 4.509）

攻读博士学位期间参加科研项目情况

[1] 国家自然科学基金，基于张量低秩约束和稀疏表示的红外小目标检测方法研究，批准号：61573183。2016 年 01 月至 2019 年 12 月，已结题，主要参与人。

[2] 西安光机所中科院光谱成像技术重点实验室开放基金，基于高光谱遥感图像的小目标检测方法研究，批准号：LSIT201401。2014 年 07 月至 2016 年 06 月，已结题，主要参与人。

[3] 江苏省大数据分析技术重点实验室开放基金：大数据量多源遥感图像融合技术研究，批准号：KXK1403。2015 年 01 月至 2016 年 12 月，已结题，主要参与人。

[4] 模式识别国家重点实验室开放基金：遥感图像中典型目标提取与场景分类方法研究，批准号：201900029。2019 年 01 月至 2020 年 12 月，在研，主要参与人。