

基于低秩稀疏分解与注意力机制的 红外小目标检测

答辩人： 戴一冕
指导教师：吴一全教授

目录

1. 绪论
2. 基于重加权块张量模型的红外小目标检测
3. 基于双向非对称注意力调制网络的红外小目标检测
4. 基于注意力激活单元的图像分类与小目标分割
5. 基于注意力特征融合的图像分类与小目标分割
6. 基于注意力局部对比度网络的红外小目标检测
7. 总结与展望

绪论

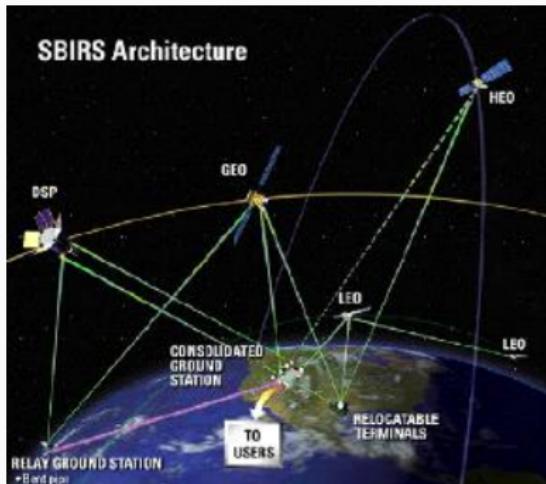
立题背景阐述

红外成像的优点：

1. 利用目标热辐射探测，可在夜间及恶劣天气下工作
2. 可探测低空目标，是雷达 & 可见光图像的重要补充
3. 被动成像，设备体积小，易搭载于各种机动平台

立题背景阐述

国防应用背景



(a) 天基红外导弹预警系统



(b) 舰载红外搜索与跟踪系统

立题背景阐述

民生应用背景



(a) 海上人员搜救



(b) 夜间停泊



(c) 夜间、恶劣天气导航



(d) 冰山块检测

近年来发展单帧检测算法的动机

- 新型飞行器的快速发展



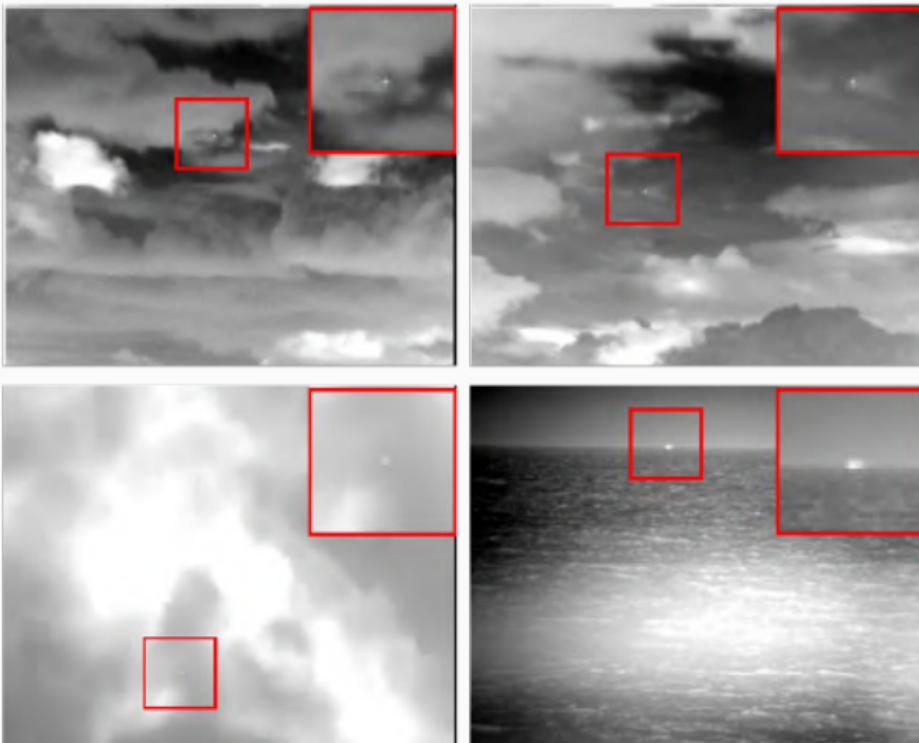
(e) 低空慢速小目标



(f) 超高音速武器

- 作为跟踪前检测算法的初始结果

典型的红外小目标图像



单帧红外小目标（Small Target）检测的难点

与通用视觉任务中的小目标（Small Object）检测存在以下差别：

1. 红外小目标的本征特征更为缺乏
 - 占图像面积：COCO [1] 的小目标 1% vs 红外小目标 0.02%
 - 更小的目标尺寸，意味着更少的形状、纹理特征
2. 红外小目标更为缺乏上下文信息
 - 通用数据集：目标与背景之间存在着较强的共生关系
 - 红外小目标：大多为非合作性目标，缺少语义上的强关联
3. 真实的红外小目标数据极为稀缺
 - 通用数据集：易收集，规模大，如 ImageNet [2]
 - 红外小目标：中短波器材受到严格管制，图像不易获取

国内外发展与研究现状

方法	技术路线	存在的问题
模型驱动	对比度检测	1. 模型判别能力不足 2. 超参数对场景敏感
	变换域分离	
	低秩稀疏分离	
数据驱动	图像块分类	1. 数据收集困难 2. 小目标特征容易被淹没
	端到端密集预测	

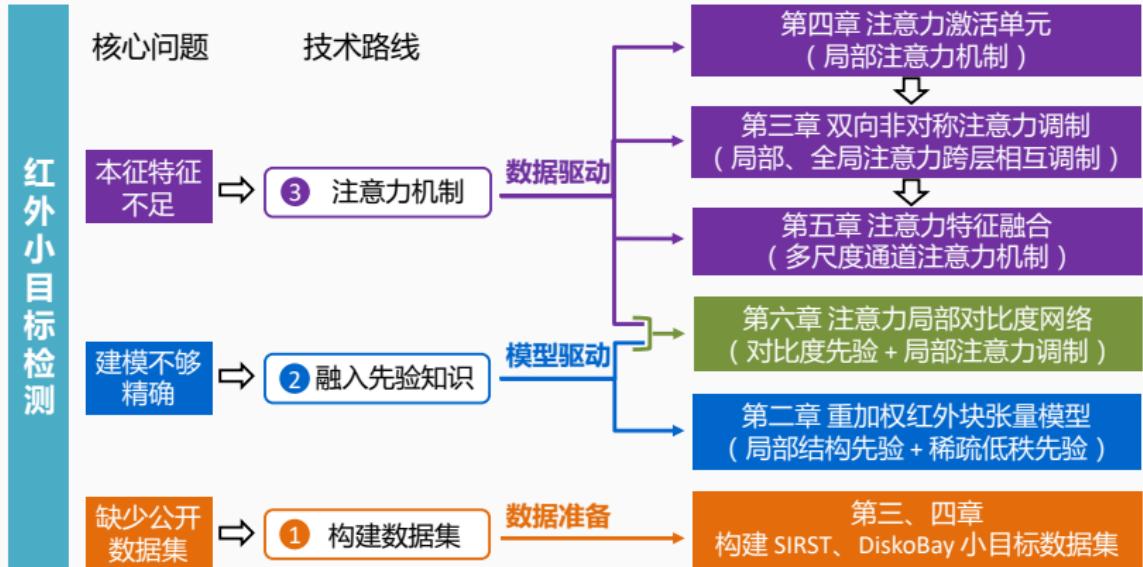
立题意义

1. 理论和工具通常是针对一般性的通用任务设计的
2. 通用任务的图像特点与实际的小目标图像有着巨大的鸿沟
3. 针对性地利用和改进这些工具、提高红外小目标检测的性能

核心问题

1. 构建开放的红外小目标检测数据集和基准测试平台
2. 针对红外小目标特点，构建检测性能更好的模型或网络
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
4. 融合嵌入领域知识的传统模型和数据驱动的深度网络

全文结构安排



基于重加权块张量模型的 红外小目标检测

- 相关成果（被引用 162 次）：

[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)

[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)

[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)

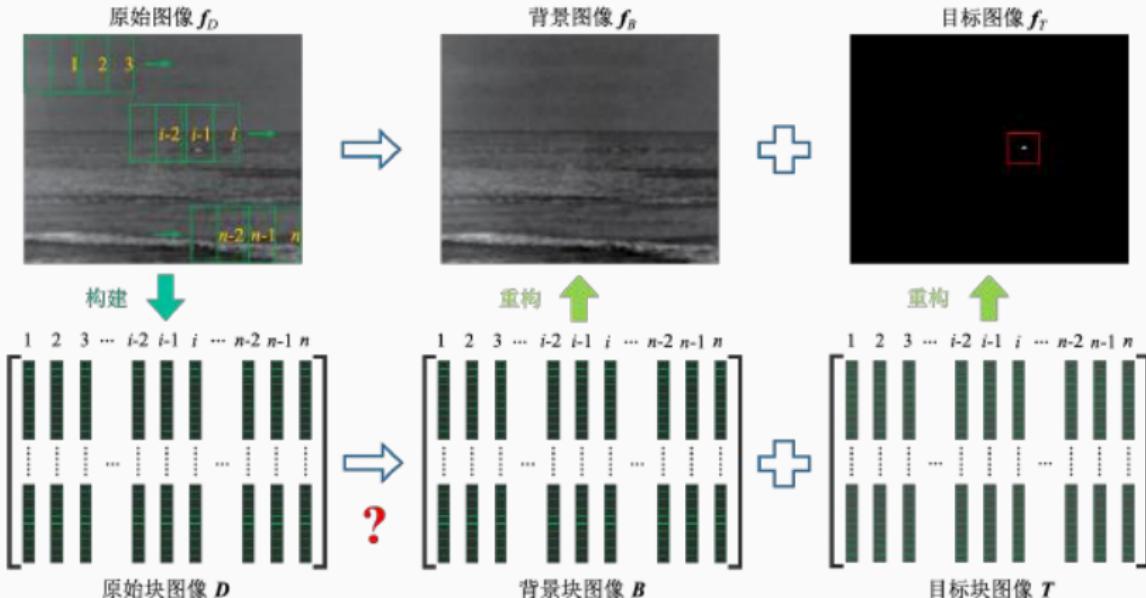
- 代码、数据集（16 Star, 9 Fork）：

<https://github.com/YimianDai/dentist>

本章背景

红外块图像 (Infrared Patch-Image, IPI) 模型 [3]:

思路：将检测问题转化为低秩背景与稀疏前景的分离问题



本章背景

- 背景分量采用低秩约束刻画
- 目标分量采用稀疏约束刻画

$$\begin{aligned} \mathbf{f}_D &= \mathbf{f}_B + \mathbf{f}_T + \mathbf{f}_N & \longrightarrow & \mathbf{D} = \mathbf{B} + \mathbf{T} + \mathbf{N} \\ &&&\swarrow \quad \searrow \\ && \boxed{\text{rank}(\mathbf{B}) \leq r \quad \& \quad \|\mathbf{T}\|_0 \leq k} & \end{aligned}$$

凸松弛

$$\left\{ \begin{array}{ll} \min_{\mathbf{B}, \mathbf{T}} \|\mathbf{B}\|_* + \lambda \|\mathbf{T}\|_0 & \text{s.t. } \mathbf{D} = \mathbf{B} + \mathbf{T} \\ \min_{\mathbf{B}, \mathbf{T}} \|\mathbf{B}\|_* + \lambda \|\mathbf{T}\|_1 & \text{s.t. } \|\mathbf{D} - \mathbf{B} - \mathbf{T}\|_{\text{F}} \leq \delta \end{array} \right.$$

先前工作存在的问题：

- 直接原因：刻画目标的稀疏约束无法区分真实目标与同样相对稀疏的背景干扰物
- 根本原因：在单帧检测的要求下，密集采样的图像块远非严格对齐，目标并不具有全局唯一的稀疏性
- 导致结果：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

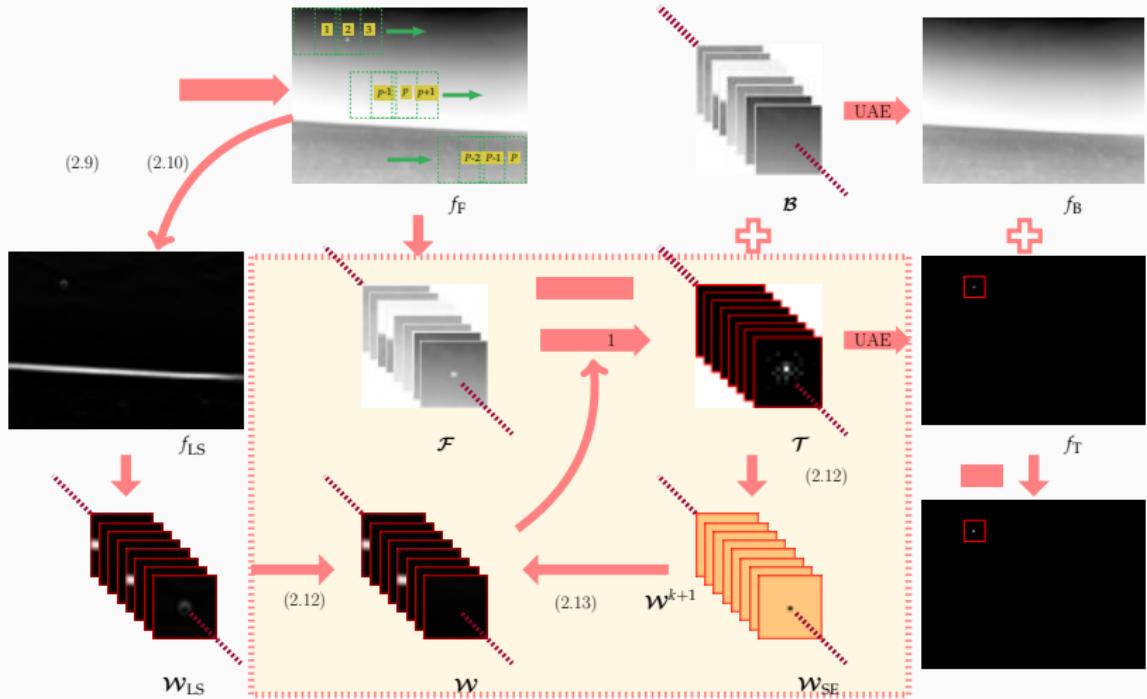
本章观察：

- 非目标残留大多为全局稀有的强边缘成分
 - 缺少同类型块，实质上的稀疏成分
 - 非局部自相似先验下稀疏的强边缘成分 => 局部的边缘分析

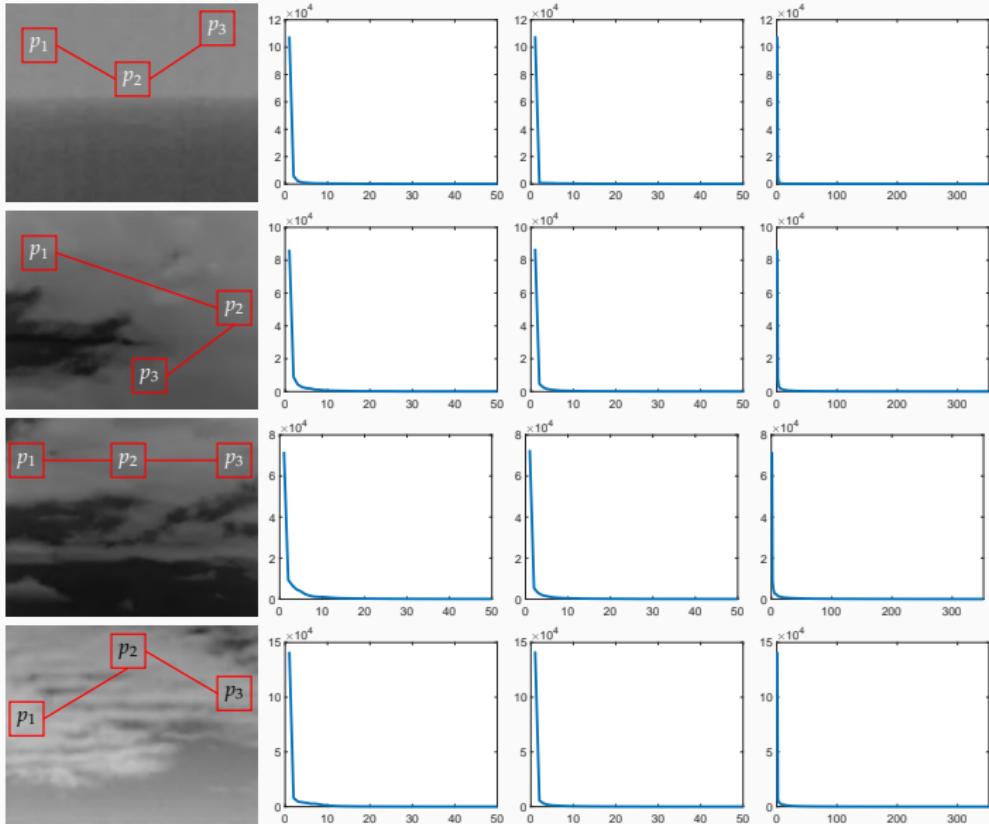
本章方法的动机：

1. 针对背景强边缘残留问题
 - 构建反映边缘强度的权重，自适应调节迭代的收缩阈值
2. 针对迭代时间长的问题
 - 根据红外小目标检测的实际情况，重新设计了算法终止条件
 - $\log \det$ 非凸正则导出的稀疏性增强权重，减少迭代次数

方法 – 框架



方法 – 背景的低秩建模



方法 – 基础模型

$$\min_{\mathcal{B}, \mathcal{T}} \text{rank}(\mathcal{B}) + \lambda \|\mathcal{T}\|_0, \text{ s.t. } \mathcal{B} + \mathcal{T} = \mathcal{F}. \quad (1)$$

↓ 松弛

$$\min_{\mathcal{B}, \mathcal{T}} \sum_{i=1}^3 \|\mathcal{B}_{(i)}\|_* + \lambda \|\mathcal{T}\|_1, \text{ s.t. } \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\text{F}} \leq \delta. \quad (2)$$

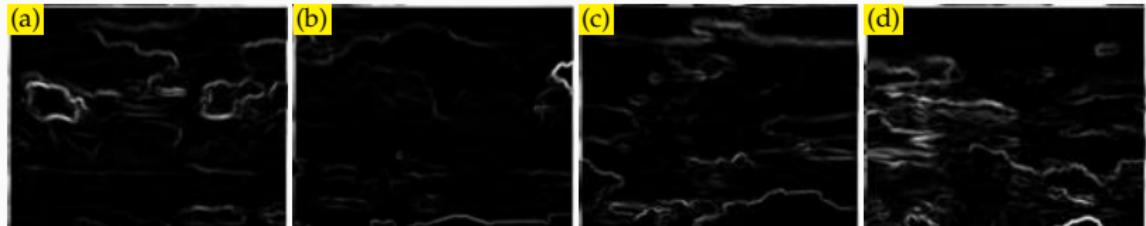
方法 – 利用边缘强度加权

出于计算简单的考虑，本章选择了结构张量：

$$J_\alpha(\nabla u_\sigma) = G_\alpha * (\nabla \mathbf{u}_\sigma \otimes \nabla \mathbf{u}_\sigma) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \quad (3)$$

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

$$\mathcal{W}_{\text{LS}} = \exp \left(h \cdot \frac{(\mathcal{L}_1 - \mathcal{L}_2) - d_{\min}}{d_{\max} - d_{\min}} \right), \quad (5)$$



方法 – 加权后的模型

$$\min_{\mathcal{B}, \mathcal{T}} \sum_{i=1}^3 \|\mathcal{B}_{(i)}\|_* + \lambda \|\mathcal{W}_{\text{LS}} \odot \mathcal{T}\|_1, \text{ s.t. } \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\text{F}} \leq \delta, \quad (6)$$

方法 – 对于终止条件的再思考

传统的终止条件: $\frac{\|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_F}{\|\mathcal{F}\|_F} \leq 10^{-7}$

- 针对受污染信号和数据的精确恢复而设计
- 需要许多轮（几十、上百）迭代才终止，**非常耗时**

对于终止条件的再思考：

- 红外小目标的分割任务是否需要那么高的恢复精度？
- 红外小目标在目标分量中较显著，后续的迭代是否必要？

本章的观察：

- 可以重新设计终止条件，减少不必要的迭代

终止条件的再设计

$$\|\mathcal{T}^{k+1}\|_0 = \|\mathcal{T}^k\|_0$$

- 理想情况: $\|\mathcal{T}^k\|_0$ 越来越小, 最后仅剩真实目标
- 真实情况: $\|\mathcal{T}^k\|_0$ 越来越大, 较多轮次后才不再变动

问题:

- 怎么让 $\|\mathcal{T}^k\|_0$ 按照我们期望的状态变化 ?
- 即怎么让分离出的稀疏分量在迭代过程中越来越稀疏 ?

稀疏性增强权重

log det 正则的经典用法

- 稀疏表示: Reweighted ℓ_1 Minimization (Candès)
- 低秩约束: Weighted Nuclear Norm Minimization (WNNM)

本章提出的 log det 正则的新用法

- 对稀疏低秩分解中的稀疏项施加, 而非经典的低秩项
- 用于确保 $\|\mathcal{T}^k\|_0$ 的快速下降, 减少迭代次数

$$\mathcal{W}_{\text{SE}}^{k+1}(i, j, p) = \frac{1}{|\mathcal{T}^k(i, j, p)| + \epsilon}, \quad (7)$$

最终模型的确立

局部结构权重与稀疏性增强权重的融合：

$$\mathcal{W}^k = \mathcal{W}_{\text{LS}} \odot \mathcal{W}_{\text{SE}}^k. \quad (8)$$

Reweighted Infrared Patch-Tensor Model 的增广拉格朗日方程：

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^N \|\mathcal{B}_{i,(i)}\|_* + \lambda \|\mathcal{W} \odot \mathcal{T}\|_1 + \\ & \sum_{i=1}^N \frac{1}{2\mu} \|\mathcal{B}_i + \mathcal{T} - \mathcal{F}\|^2 - \langle \mathcal{Y}_i, \mathcal{B}_i + \mathcal{T} - \mathcal{F} \rangle, \end{aligned} \quad (9)$$

为什么要着眼于设计权重项？

是权重，也是阈值：

$$\mathcal{B}_i^{k+1} = \text{fold}_i \left(\mathcal{D}_\mu \left[\left(\mathcal{F} + \mu \mathcal{Y}_i^k - \mathcal{E}^k \right)_{(i)} \right] \right) \quad (10)$$

$$\mathcal{T}^{k+1} = \mathcal{S}_{\frac{\mu\lambda}{N}} \mathcal{W}^k \left[\frac{1}{N} \sum_{i=1}^N \left(\mathcal{F} + \mu \mathcal{Y}_i^k - \mathcal{B}_i^{k+1} \right) \right] \quad (11)$$

动态自适应的权重即动态自适应的自适应阈值。

优化求解 – 交替方向乘子法 (ADMM)

算法 1: RIPT 模型优化算法

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

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

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

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

 ▷ 固定其他项, 更新 \mathcal{B}_i

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

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

end

 ▷ 固定其他项, 更新 \mathcal{T}

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

 ▷ 固定其他项, 更新 \mathbf{Y}_i

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

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

end

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

 ▷ 更新 μ : $\mu^{k+1} := \mu^k / \rho$

 ▷ 更新 k : $k = k + 1$

end

整体实验的设计思路：

1. 实验设定
 - 1.1 对比方法选取及参数设定
 - 1.2 定量评价指标
2. 参数分析
3. 消融实验
 - 3.1 局部结构权重的有效性验证
 - 3.2 稀疏性增强权重的有效性验证
4. 与其他方法对比
 - 4.1 定量指标对比
 - 4.2 主观视觉效果对比
 - 4.3 计算时间对比

12 种对比方法选取及参数设定

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	$N = 1, 3, \dots, 9$
5	WLDM	$L = 4, m = 2, n = 2$
6	TDLMS	支撑区域大小: 5×5 , 步长: $\mu = 5 \times 10^{-8}$
7	IPI	块大小: 50×50 , 滑动步长: 10, $\lambda = \frac{L}{\sqrt{\min(I, P)}}$, $L = 3, \varepsilon = 10^{-7}$
8	PRPCA	块大小: 50×50 , 滑动步长: 10, $\lambda = \frac{L}{\sqrt{\min(I, P)}}$, $L = 3, \varepsilon = 10^{-7}$
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, $\lambda = \frac{L}{\sqrt{\min(I, P)}}$, $L = 2, r = 5 \times 10^{-3}$
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, $\lambda = \frac{L}{\sqrt{\min(I, J, P)}}$, $L = 1, h = 10, \epsilon = 0.01, \varepsilon = 10^{-7}$

1. 背景抑制性能：

- 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)

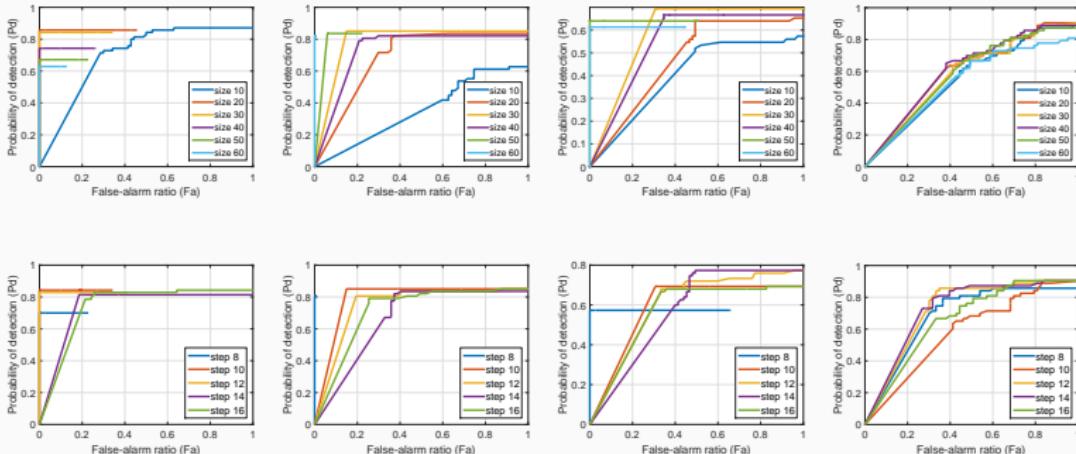
2. 检测性能：

- 检测率
- 虚警率
- 接收机工作特性 (Receiver Operating Characteristic, ROC) 曲线

参数分析

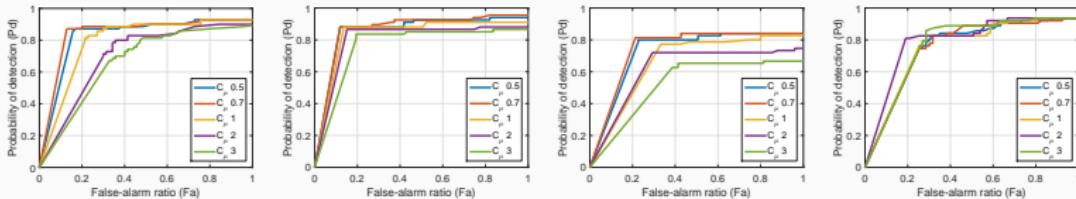
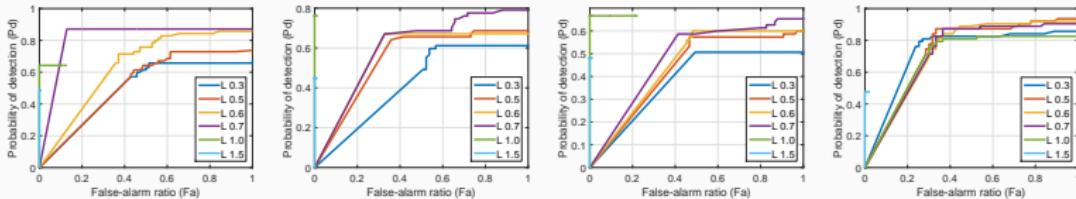
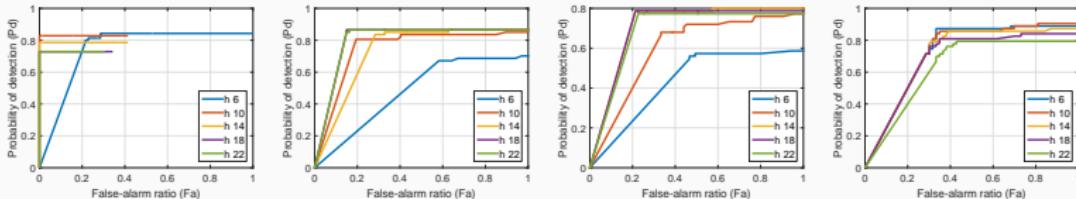
参数分析的动机：

1. 寻找最佳的超参数
2. 检验超参数取值对于模型性能的影响

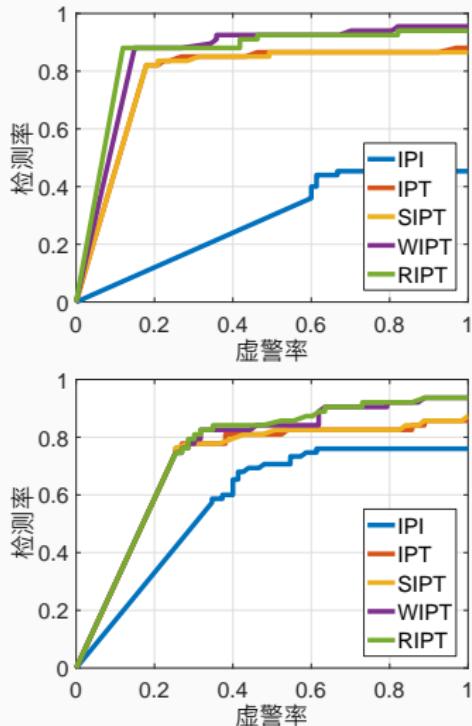
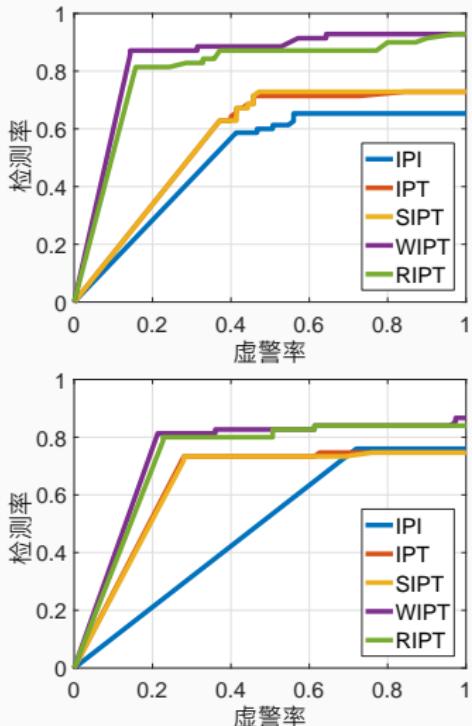


参数分析

(接上一页)

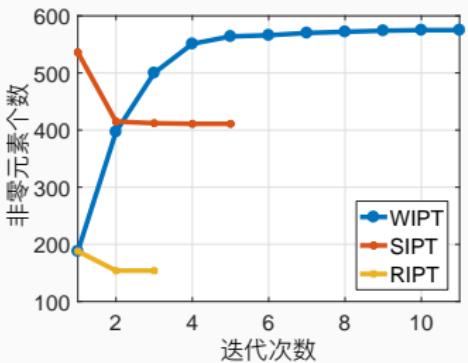
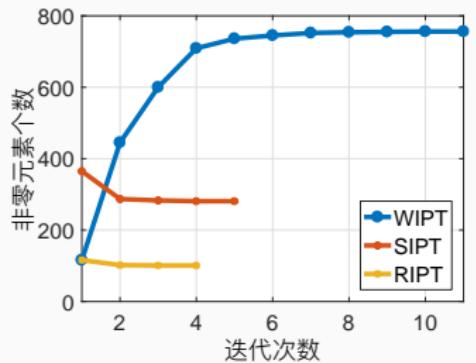
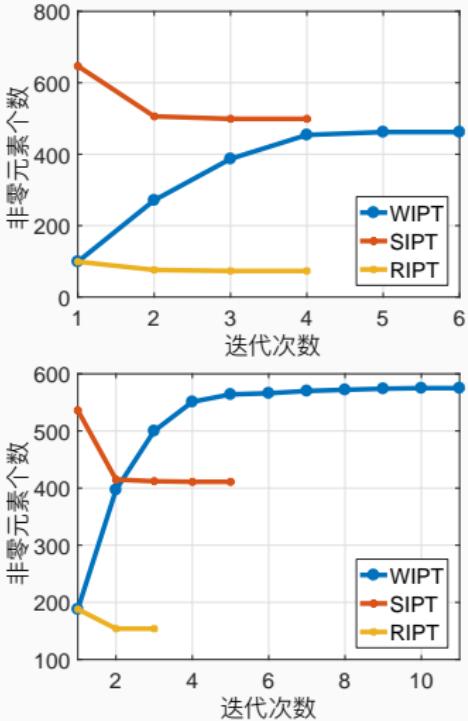
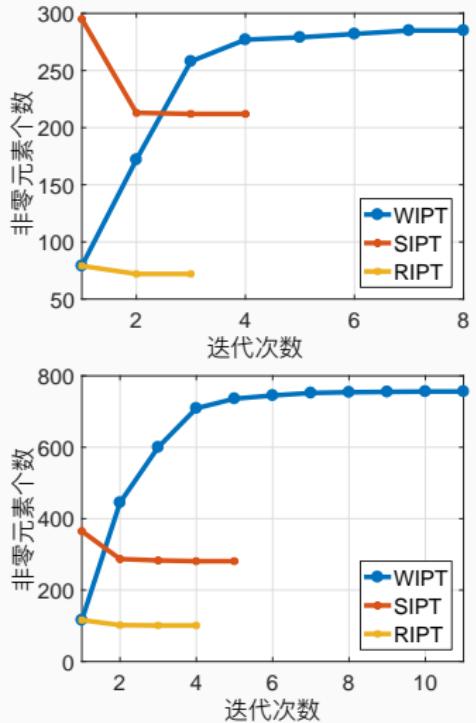


局部结构权重的有效性验证（对最终检测性能的影响）



1. WIPT 效果好于 IPT，表明引入局部结构权重可以改善检测性能
2. WIPT 与 RIPT 效果相当，表明稀疏度增强权重不会影响检测效果

稀疏性增强权重的有效性验证 (对于目标图像稀疏性的影响分析)



稀疏性增强权重可以显著减少新终止条件下的算法迭代次数

定量指标对比 – 背景抑制性能比较

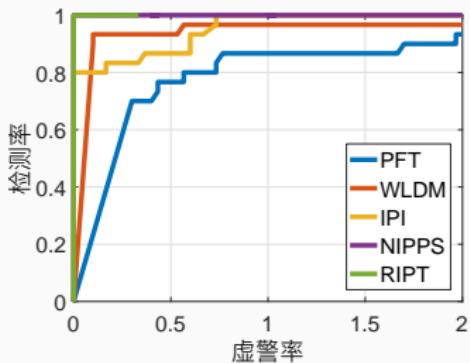
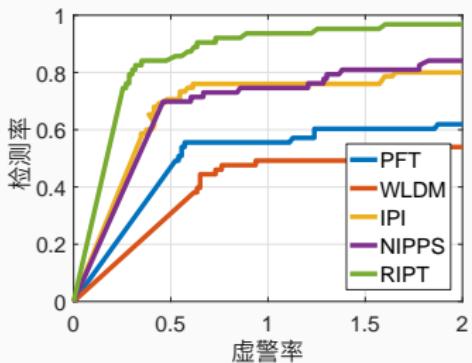
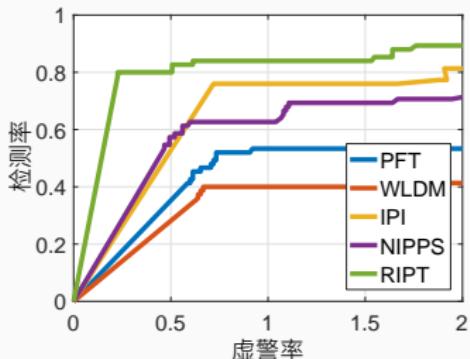
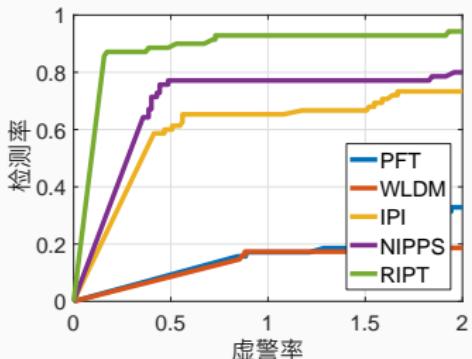
Table 1: 序列 1 - 4 代表性图像上的定量评价指标比较

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。

1. 低秩稀疏分解方法总体抑制背景更为彻底，RIPT 模型最佳
2. 针对传统滤波方法设计的评价指标并不适合新的技术路线

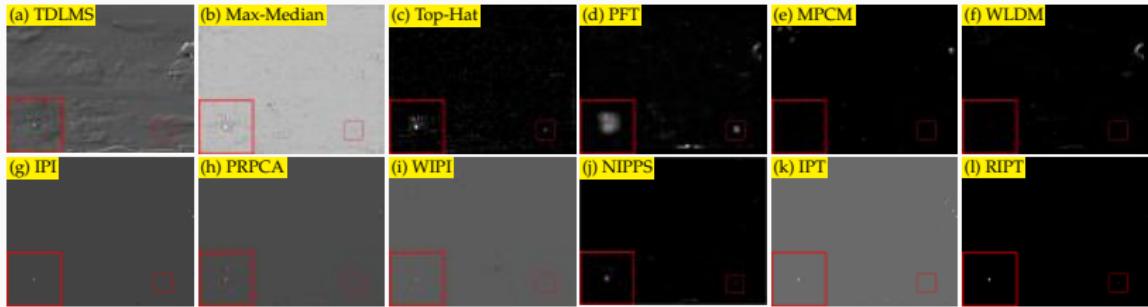
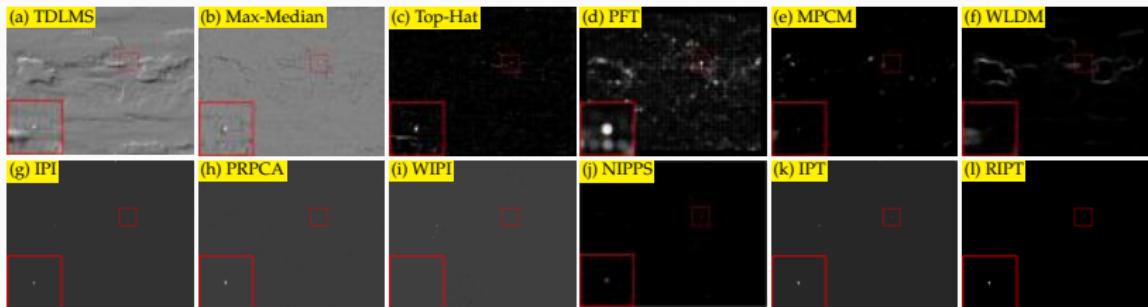
定量指标对比 – 检测性能（ROC 曲线）比较



RIPT 模型可以在相同的虚警率下实现更高的检测率

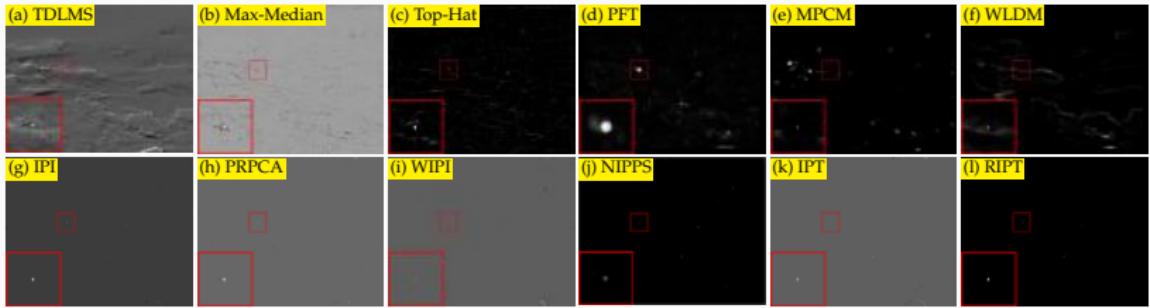
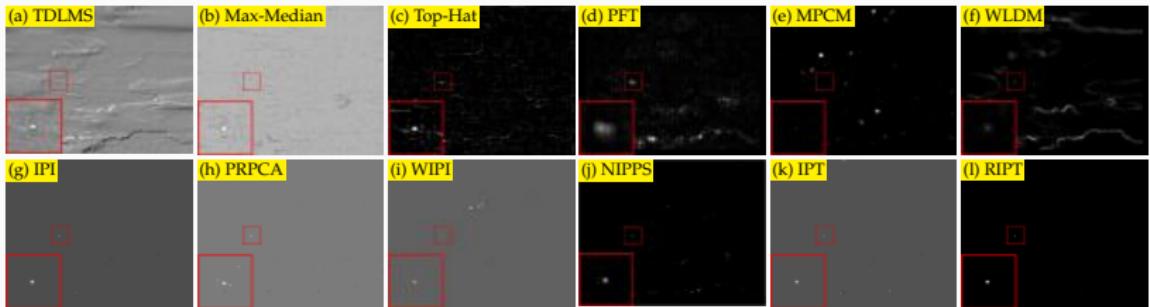
主观视觉效果对比

与其他 11 种方法对比，本章的 RIPT 模型主观视觉对比最佳
(需放大后查看)



主观视觉效果对比

(接上一页)



计算时间对比

Table 2: 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	$\mathcal{O}(L^2 MN)$	$\mathcal{O}(MN \log MN)$	$\mathcal{O}(L^3 MN)$	$\mathcal{O}(L^3 MN)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$
时间/秒	0.162	0.025	0.083	6.059	16.998	52.995	15.515	8.598	6.932	3.169

本章所提出的 RIPT 模型在所有低秩稀疏分解方法中用时最少，仅为 IPI 模型的 18.6%。

本章小结

1. 利用图像的局部结构信息，作为以非局部自相似性为基础的低秩稀疏分解方法的补充，抑制稀疏的强边缘残留
2. 从重构与分割任务的差异出发，通过设计新的终止条件以及引入稀疏性促进权重，降低了算法运行时间

基于双向非对称注意力调制网络的 红外小目标检测

- 相关成果：

[1] **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)

- 所贡献的红外弱小目标单帧数据集 (11 Star) :

<https://github.com/YimianDai/sirst>

- 代码、训练好的模型 (10 Star) :

<https://github.com/YimianDai/open-acm>

以 RIPT 模型为代表的模型驱动方法的不足：

1. 模型判别能力不足，易混淆真实目标与其他高频背景干扰物
 - 所采用的特征过于简单、语义判别能力不足
 - 通常为图像块或者均值、最大值、熵等简单统计量
2. 超参数的选择依赖于具体的图像内容，对场景变化不够鲁棒
 - 模型越复杂，易引入越多的超参数，鲁棒性越差（过拟合）
 - 信号处理与机器学习对待优化的区别，即对泛化性的关注

1. 深度学习兴起之前，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、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。
2. 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。

泛化性之于红外小目标检测：

- 红外小目标检测的一大挑战在于如何应付复杂多变的场景，作为一个需求导向的应用研究，模型泛化能力应是其关键。

本章背景

深度学习屠榜的时代，单帧红外小目标检测的发展较为滞后：

1. 最大原因：迄今为止还缺少一个公开的基准数据集

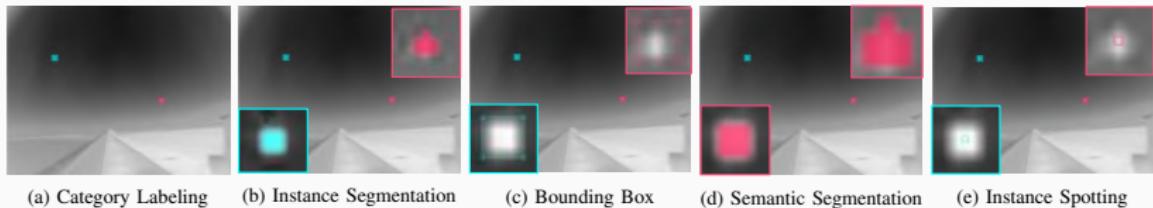
- 根源：收集真实的红外小目标图像非常困难
 - 短波、中波红外成像的器件受到各国严格管制
 - 高空高速飞行器等感兴趣目标也不易抓取
- 后果：
 - 深度学习方法缺少足够的训练样本
 - 也缺少一个所有方法可以公平比较的平台

2. 次要原因：成像后的绝对面积太小，缺少本征特征

- 红外成像的特点 + 成像距离太远 => 缺少纹理和形状特征
- 特征图语义程度与特征图分辨率大小的矛盾更加极端

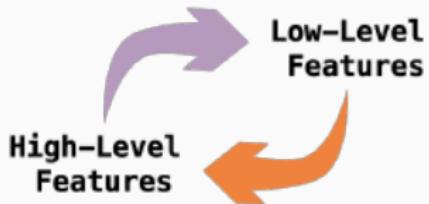
本章贡献

- 该领域首个开放且多种形式高质量标注的公开数据集



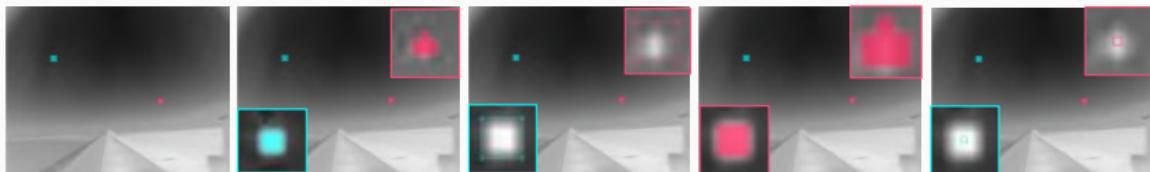
(a) Category Labeling (b) Instance Segmentation (c) Bounding Box (d) Semantic Segmentation (e) Instance Spotting

- 针对红外小目标的特点设计了一个跨层特征融合模块
 - 非对称的上下文调制 (Asymmetric Contextual Modulation, ACM)
 - 背后的想法：不应该只有自下而上的语义调制 [4]，高层语义信息与低层细节信息应当实现跨层的互相交换



图像收集和标注

- 427 帧图像，共 480 个目标实例
 - 50% 训练、20% 验证，30% 测试
 - 每个序列只挑选一幅代表性图像加入该数据集，以确保背景不重复（模型是真实检测出了目标，而非仅仅机械记住地了序列中的目标或者背景）
- 一共五种类型的标记，以支持多种任务建模
 1. 类别标记：图像分类，直接判定图像是否含有小目标
 2. 实例分割标记：支持实例分割任务
 3. 边框标记：支持目标检测任务
 4. 语义分割标记：支持语义分割任务
 5. 实例发现标记：支持实例发现任务（目前手头在做的工作）



(a) Category Labeling

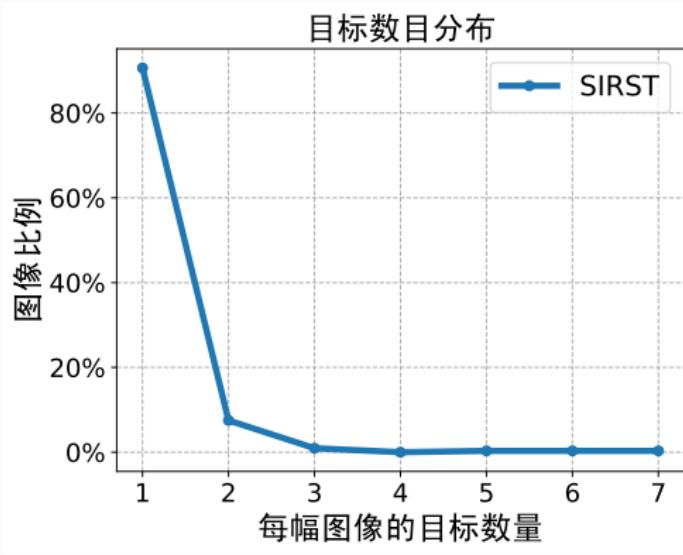
(b) Instance Segmentation

(c) Bounding Box

(d) Semantic Segmentation

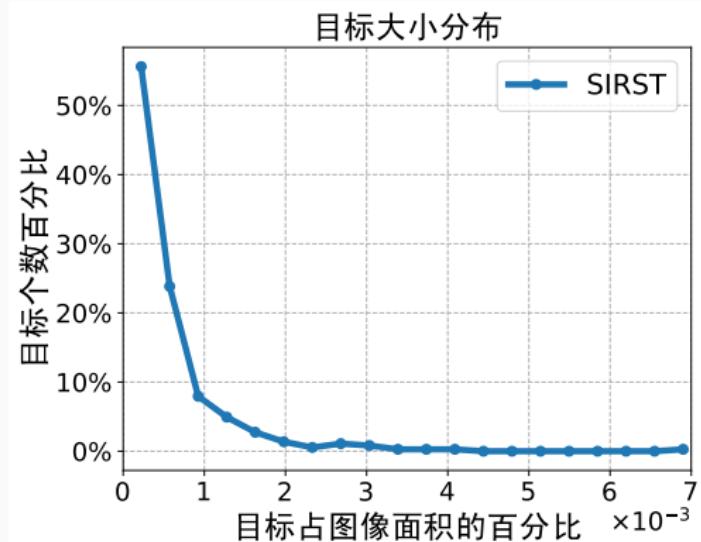
(e) Instance Spotting

统计信息 – 重新审视传统方法的一些假设



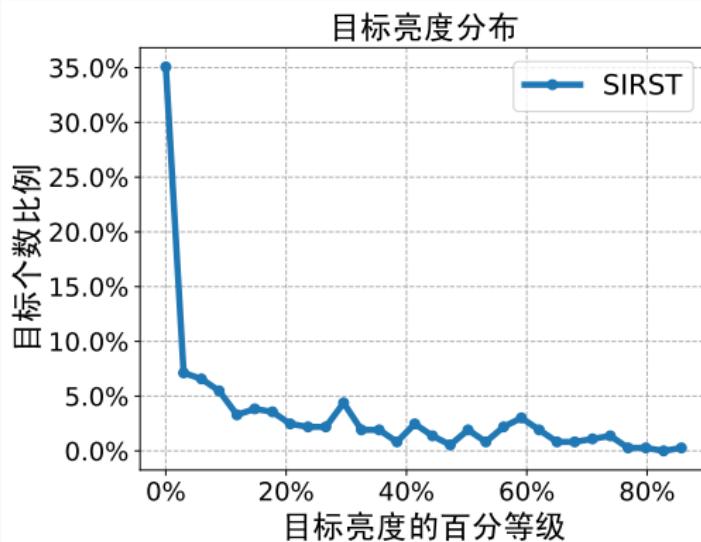
1. 90% 的图像中仅有单个目标：一定程度上支撑了全局唯一的稀疏性、显著性假设（假设忽略背景中的成分）
2. 但仍然有 10% 的图像含有多个目标，上述假设会导致漏检

统计信息 – 重新审视传统方法的一些假设



1. 55% 的目标仅有图像面积的 0.02%，作为对比，COCO 中小目标为图像面积 1%，非常稀疏
2. 目标自身特征的缺乏，随着网络的加深，目标特征容易被背景淹没

统计信息 – 重新审视传统方法的一些假设



- 仅有 35% 图像中目标亮度最大，简单的阈值化方法最理想也无法检测出剩余的 65%
- 65% 的目标与背景灰度类似，乃至 17% 的目标暗于图像平均灰度，关于红外小目标的显著性假设忽略了弱目标情形

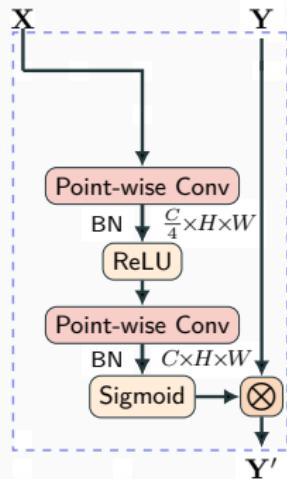
针对红外小目标的特点，做出如下改动：

1. 重新定制下采样方案：减少下采样倍数
2. 重新定制注意力模块：局部通道注意力机制¹
3. 重新定制跨层特征融合方法：非对称上下文调制模块（ACM）

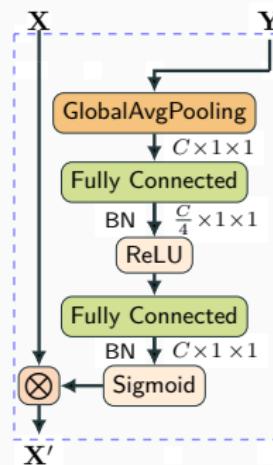
¹为下一章的主要内容

红外小目标检测网络

1. 自底向上的特征调制：局部通道注意力（提供细节信息）
2. 自顶向下的特征调制：全局通道注意力（提供全局信息）



(a) LCAM 模块

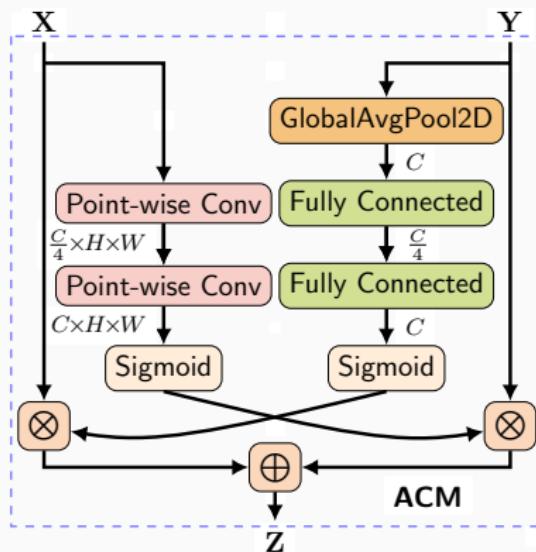


(b) GCAM 模块

红外小目标检测网络

非对称上下文调制模块 (ACM):

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



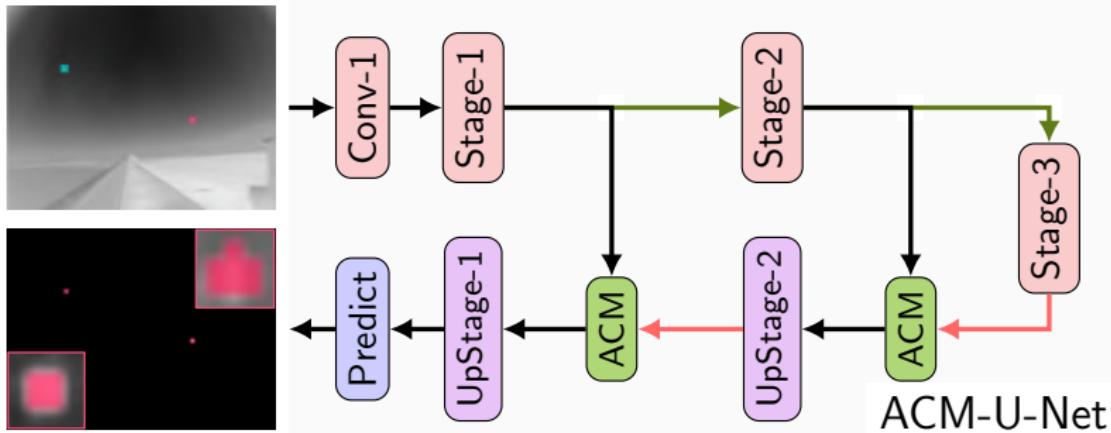
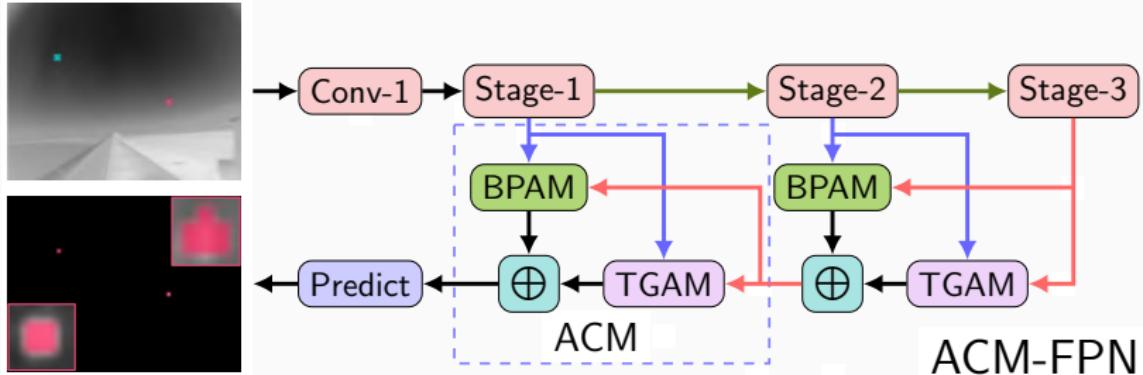
红外小目标检测网络

以 ResNet-20 为例，减少下采样倍数

Stage	Output	Backbone
Conv-1	480×480	$3 \times 3 \text{ conv}, 16$
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	120×120	$\begin{bmatrix} 3 \times 3 \text{ conv}, 64 \\ 3 \times 3 \text{ conv}, 64 \end{bmatrix} \times b$

b 为网络深度的伸缩参数，当 $b = 3$ 时，为标准的 ResNet-20.

网络实例



实验

整体实验的设计思路：

1. 消融实验

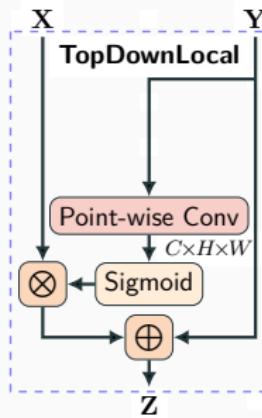
- 1.1 降低下采样倍数的有效性验证
- 1.2 双向调制的有效性验证
- 1.3 非对称双向调制的有效性验证

2. 与其他方法对比

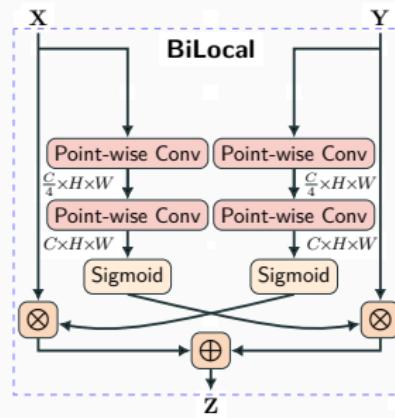
- 2.1 IoU 和 nIoU 指标比较
- 2.2 ROC 曲线性能比较

消融实验

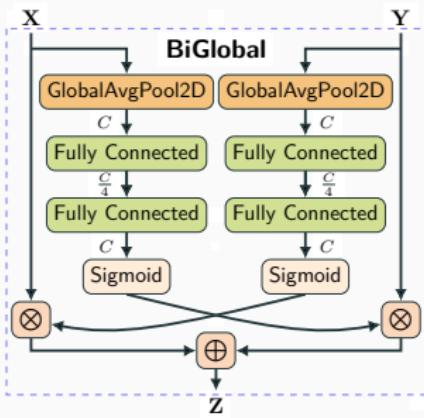
本章消融实验中所对比的模块结构图



(a) TopDownLocal



(b) BiLocal



(c) BiGlobal

控制变量：ACM 以及这三个模块均具有相同的参数数量

以 FPN 作为宿主网络的实验结果

Modulation Scheme	FPN as Host Network							
	IoU				nIoU			
	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 1$	$b = 2$	$b = 3$	$b = 4$
TopDownLocal	0.595	0.648	0.693	0.713	0.635	0.662	0.688	0.703
BiGlobal	0.599	0.660	0.685	0.693	0.645	0.674	0.696	0.684
BiLocal	0.591	0.662	0.713	0.722	0.657	0.694	0.709	0.714
Regular-ACM	0.683	0.703	0.711	0.711	0.661	0.671	0.680	0.675
ACM	0.645	0.700	0.714	0.731	0.684	0.702	0.713	0.721

以 U-Net 作为宿主网络的实验结果

Modulation Scheme	U-Net as Host Network							
	IoU				nIoU			
	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 1$	$b = 2$	$b = 3$	$b = 4$
TopDownLocal	0.648	0.710	0.713	0.718	0.673	0.692	0.694	0.697
BiGlobal	0.682	0.716	0.723	0.730	0.688	0.708	0.707	0.719
BiLocal	0.670	0.715	0.718	0.742	0.680	0.710	0.713	0.720
Regular-ACM	0.684	0.700	0.692	0.692	0.637	0.650	0.646	0.643
ACM	0.707	0.732	0.741	0.743	0.709	0.720	0.726	0.731

从实验结果中可以得出如下结论，在相同网络参数量的情况下：

1. Regular-ACM 和 ACM 的对比结果表明，减少下采样倍数更有利于小目标检测（验证了感受野匹配的重要性）
2. TopDownLocal 和 BiGlobal 的对比结果表明，双向调制比单独的自顶向下调制更加有效（验证了高低层互相交换信息的合理性，也是本文网络结构背后的动机）
3. BiGlobal, BiLocal 和 ACM 的对比结果表明，非对称调制比对称调制更加有效（验证了不同层的特征需要不同尺度的通道注意力机制²）

²启发了第五章的内容

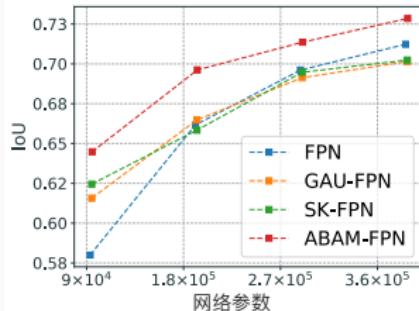
与其他方法对比

方法	模型驱动方法												数据驱动方法						
	局部对比度量			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	ACM	TBC	U-Net	SK	GAU	ACM			
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			

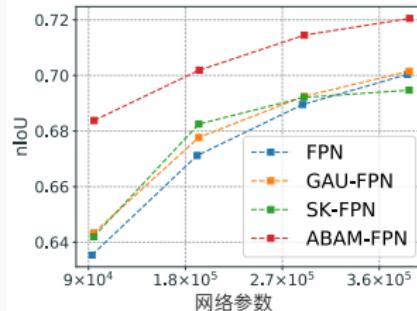
1. 在模型驱动方法中，稀疏低秩分解方法具有比局部对比度度量方法更好的表现（NIPPS 我的另一个工作）
2. 全体深度网络的检测性能均好于传统的模型驱动的方法，彰显了数据驱动方法的优势和潜力
3. 所提出的 ACM 模块在对比的跨层特征融合方案中取得了最好的性能表现

与其他方法对比

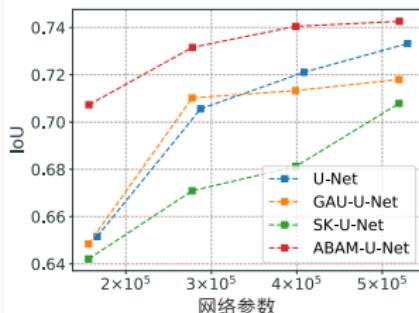
随着网络加深，检测性能的变化情况：



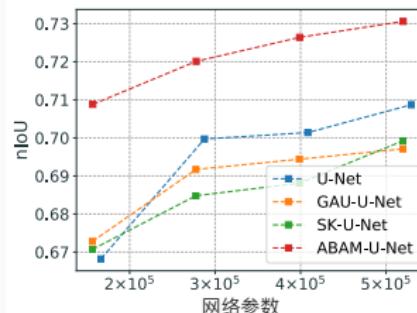
(a) 以 FPN 为宿主网络的 IoU 比较



(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

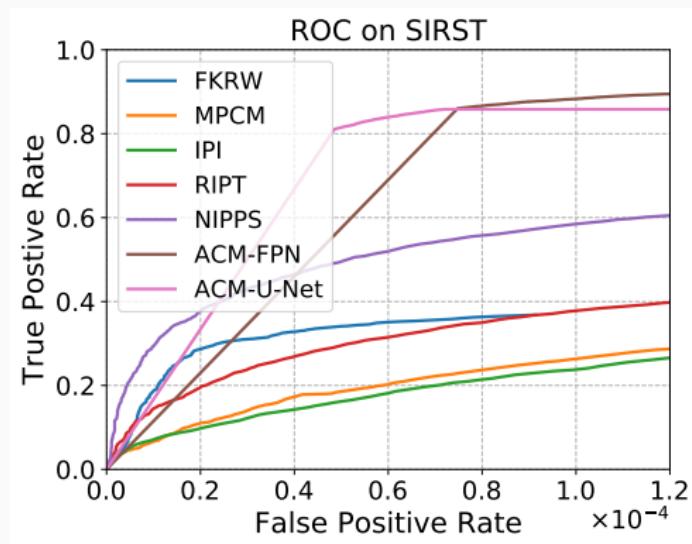
与其他方法对比

随着网络加深，检测性能的变化情况 – 实验表明：

1. 在红外小目标检测任务（No Free Lunch）上，ACM 模块优于其他跨层融合的注意力调制方案，可以在网络层数更少的情况下取得相当乃至更好的性能
2. 对于红外小目标检测任务，与其一味加深网络，远不如设计更加合理的跨层融合模块来得更为经济、高效

与其他方法对比

ROC 曲线比较也印证了本章网络在红外小目标检测上性能良好：



本章小结

1. 构造了首个开放的单帧红外小目标检测数据集
2. 定制了非对称上下文调制模块用于跨层特征融合

基于注意力激活单元的图像分类与 小目标分割

- 相关成果：
[1] **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, Oral 比例为所有投稿的前 3%)
- 所贡献的弱小冰山检测数据集（撰写中）：
<https://github.com/YimianDai/diskobay>
- 代码、训练好的模型（Star 7 Fork 3）：
<https://github.com/YimianDai/open-atac>

注意力机制：从 Block 下沉到 Layer 的动机

1. 不仅上一章，还有大量近年来的工作都表明，在网络中加入注意力模块可以提升网络性能
2. 很自然的问题：如果越多的注意力模块效果越好，那么如何才能添加更多的注意力模块？毕竟 SENet 已经将注意力做到了每个残差块。
3. 打开残差块，更小的单元只有卷积（Conv）、激活单元（ReLU）、批归一化（BN）构成的每一层了。
4. 那么是卷积、激活单元还是批归一化呢？

本章动机

观察：注意力机制与激活函数的统一框架

1. 注意力机制的调制过程可以被抽象表示为

$$\mathbf{X}' = \mathbf{G}(\mathbf{X}) \otimes \mathbf{X}, \quad (12)$$

2. 给定元素位置 (c, i, j) , 式 (12) 的标量形式可以被表示为

$$\mathbf{X}'_{[c, i, j]} = \mathbf{G}(\mathbf{X})_{[c, i, j]} \cdot \mathbf{X}_{[c, i, j]} = g_{c, i, j}(\mathbf{X}) \cdot \mathbf{X}_{[c, i, j]}. \quad (13)$$

3. 激活函数也可以被统一表述成如下的门控函数形式

$$\mathbf{X}'_{[c, i, j]} = g'(\mathbf{X}_{[c, i, j]}) \cdot \mathbf{X}_{[c, i, j]}. \quad (14)$$

- ReLU (指示函数)、SINREN 单元 (Sinc 函数)、其他也类似

本章动机

观察：注意力机制与激活函数的统一框架

1. 两者都可以被表述成非线性自适应的门控函数
2. 区别在于激活函数中的门控函数输入为标量，而注意力机制中的门控函数则为整个特征图
3. 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
4. 而注意力机制则可以被看作是一个结构复杂、上下文感知的激活单元

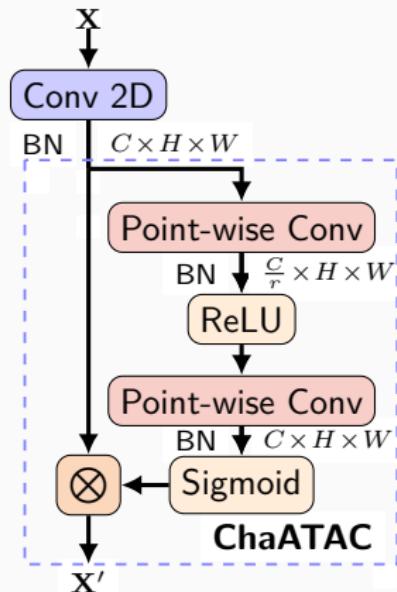
动机：考虑到这两者的联系，轻量级注意力模块作为激活单元：

1. 可以实现在网络中引入非线性的基本功能
2. 能够逐层地对卷积输出的特征进行动态、上下文感知的精炼

1. 提出了注意力激活单元这一构想
 - 1.1 给出了局部通道注意力单元、局部空间注意力单元、混合注意力激活单元等具体构造
 - 1.2 给出了一条构造全注意力网络的可行途径
2. 贡献了首个公开的弱小冰山检测数据集
 - 2.1 更大规模的小目标（Small Target）数据集的构建
 - 2.2 数据特点与红外小目标很相似，且也有红外波段，为后续利用迁移学习缓解红外小目标检测小样本问题创造更好的条件（对比 ImageNet，源数据集与目标数据集的分布差异更小）

注意力激活单元

局部通道注意力单元



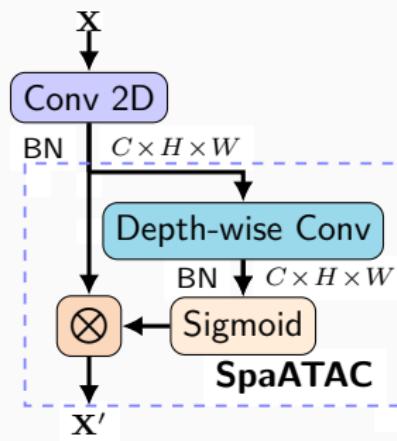
局部通道注意力单元

可以看作是 SENet 模块的局部版本，区别在于：

1. 从结构本身上，ChaATAC 单元移除了 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 模块对整个特征图施加相同的全局权重
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

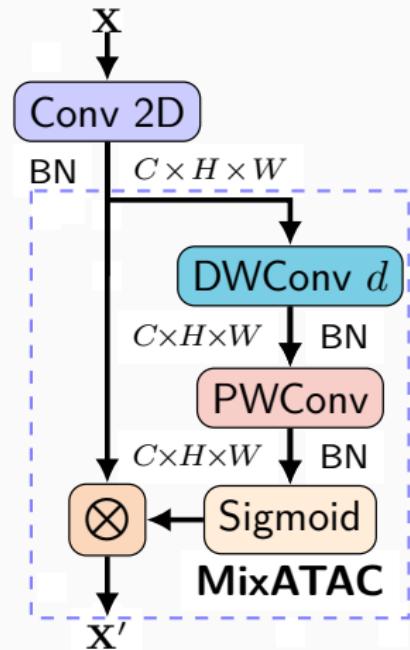
注意力激活单元

局部空间注意力单元



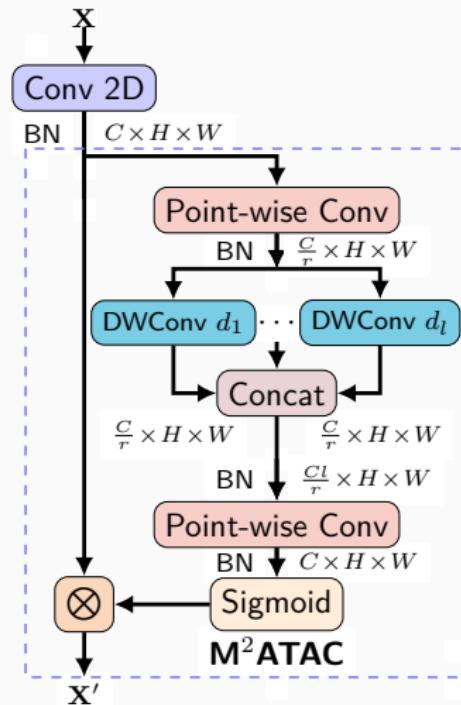
注意力激活单元

混合注意力激活单元



注意力激活单元

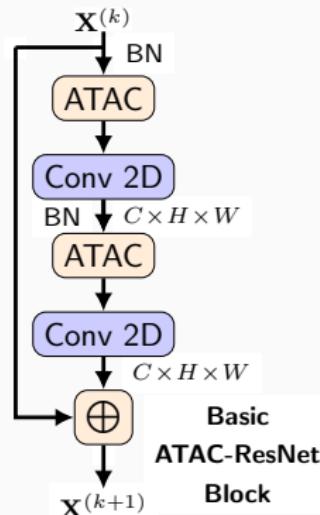
多尺度混合注意力激活单元



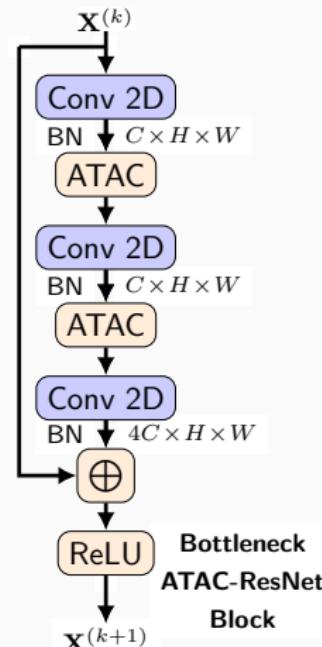
全注意力网络

将注意力激活单元嵌入残差块

基础残差块



瓶颈残差块



全注意力网络

构造全注意力网络的动机：

1. 注意力激活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
2. 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。

构建更大规模的小目标数据集

动机 – 再次回到红外小目标

1. 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
2. 数据分布与 ImageNet 等通用数据集的物体数据分布差异过大，难以像其他视觉任务那样通过预训练实现模型迁移 [5]

动机 – 降低源分布与目标分布之间的差距

1. 在可见光遥感影像中，存在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
2. 可见光遥感影像获取更为开放，易于构建较大规模的数据集
3. 额外的福利：走向更多的波段

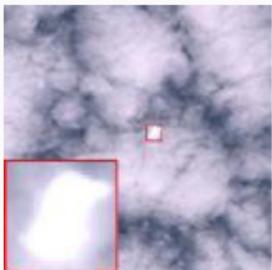
部分代表性图像



(a) 极小尺寸



(b) 低对比度



(c) 大尺寸



(d) 相似的背景干扰

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. 受限于传感器分辨率以及冰山自身的分布特性，大小集中于 2×2 至 10×10 个像素之间，缺少相应的纹理和形状特征
3. 存在特定类型的云层，在光谱和外形上类似，易造成虚警

Table 3: 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%

实验

整体实验的设计思路：

1. 实验设定

1.1 数据集

1.2 骨干网络

2. 消融实验

2.1 注意力激活所需要的上下文 – 全局还是局部？

2.2 微模块的范式选择：SENet、NiN 还是本章的 ATAC？

2.3 网络需要走向全注意力吗？

3. 与其他方法对比

实验数据集

图像分类：CIFAR-10 数据集

airplane



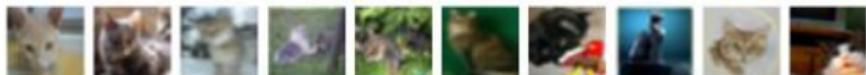
automobile



bird



cat



deer



dog



- 5 万幅训练图像和 1 万幅测试图像，一共 10 类
- 难度较小，数据量合适，作为消融实验的对象

实验数据集

图像分类：CIFAR-100 数据集



- 5 万幅训练图像和 1 万幅测试图像，一共 100 类
- 难度比 CIFAR-10 大，数据量合适，作为消融实验的主力

图像分类：ImageNet 数据集



- 128 万幅训练图像和 5 万幅测试图像，一共 1000 类
- 用于验证本章以及后一章的模型在大规模数据集上的性能

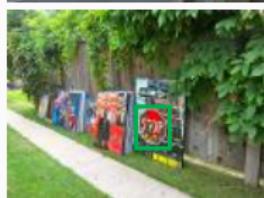
实验数据集

语义分割：DiskoBay 数据集



实验数据集

语义分割：StopSign 数据集（COCO 子集，受限于机器资源）



图像分类实验的骨干网络

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	$\begin{bmatrix} 3 \times 3 \text{ conv, 16} \\ 3 \times 3 \text{ conv, 16} \end{bmatrix} \times b$	112×112	$\begin{bmatrix} 3 \times 3 \text{ conv, 64} \\ 3 \times 3 \text{ conv, 64} \\ 3 \times 3 \text{ conv, 256} \end{bmatrix} \times 3$
Stage-2	16×16	$\begin{bmatrix} 3 \times 3 \text{ conv, 32} \\ 3 \times 3 \text{ conv, 32} \end{bmatrix} \times b$	56×56	$\begin{bmatrix} 3 \times 3 \text{ conv, 128} \\ 3 \times 3 \text{ conv, 128} \\ 3 \times 3 \text{ conv, 512} \end{bmatrix} \times 4$
Stage-3	8×8	$\begin{bmatrix} 3 \times 3 \text{ conv, 64} \\ 3 \times 3 \text{ conv, 64} \end{bmatrix} \times b$	28×28	$\begin{bmatrix} 3 \times 3 \text{ conv, 256} \\ 3 \times 3 \text{ conv, 256} \\ 3 \times 3 \text{ conv, 1024} \end{bmatrix} \times 6$
Stage-4			14×14	$\begin{bmatrix} 3 \times 3 \text{ conv, 512} \\ 3 \times 3 \text{ conv, 512} \\ 3 \times 3 \text{ conv, 2048} \end{bmatrix} \times 3$
		1×1	GAP, FC, Softmax	

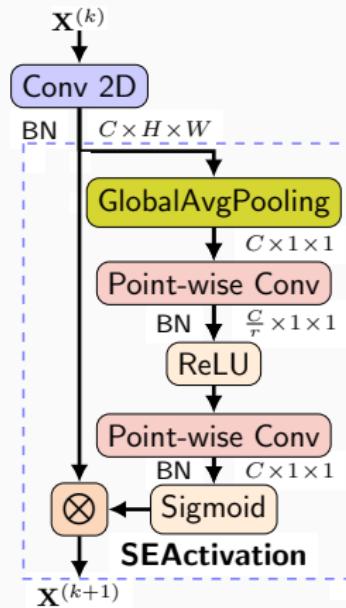
骨干网络

语义分割实验的骨干网络 (Basic ContextNet[6])

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

消融实验 – 特征上下文聚合尺度的重要性

局部通道注意力激活单元的消融实验架构：



验证对于 ChaATAC 单元，上下文聚合尺度应该全局还是局部？

消融实验 – 特征上下文聚合尺度的重要性

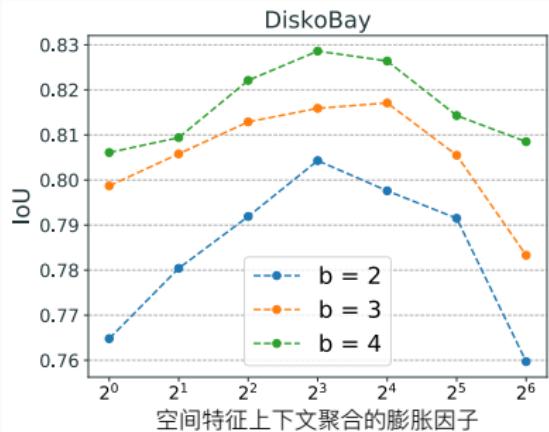
ChaATAC 单元与 SEActivation 单元的分类性能比较

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

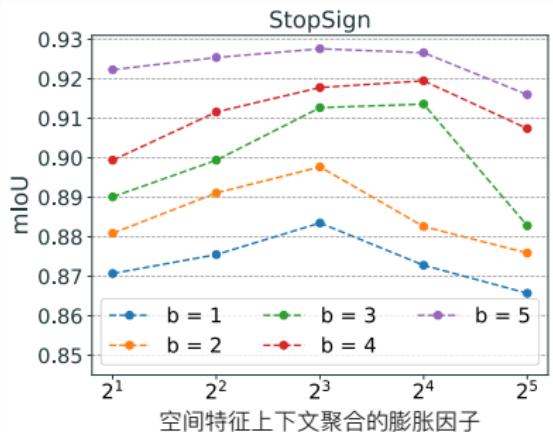
实验结果表明：相同网络参数数量的情况下，对于通道注意力激活单元，聚合上下文的局部性至关重要

消融实验 – 特征上下文聚合尺度的重要性

不同膨胀因子下 SpaATAC 单元的语义分割性能比较



(a) DiskoBay

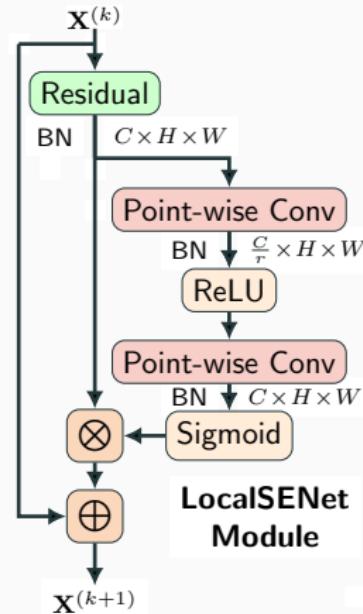
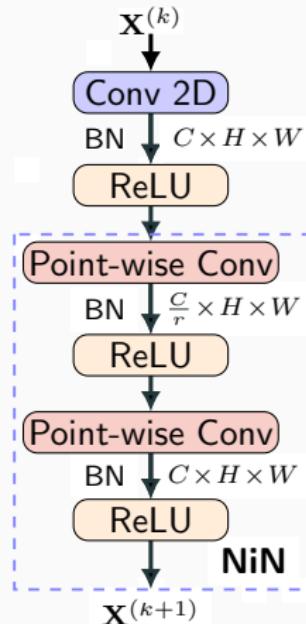


(b) StopSign

实验结果表明：相同网络参数数量的情况下，对于空间注意力激活单元，聚合上下文的尺度选择也同样至关重要

消融实验 – 微模块结构的选取

消融实验架构：



为保证网络参数相同，LocalSENet 的 r 大小为其他结构的 2 倍

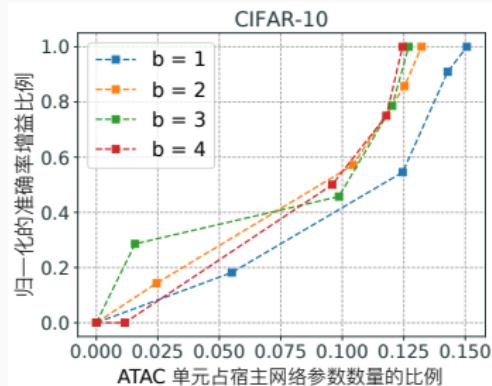
消融实验 – 微模块结构的选取

ChaATAC 单元与其他两种微模块结构的分类性能比较

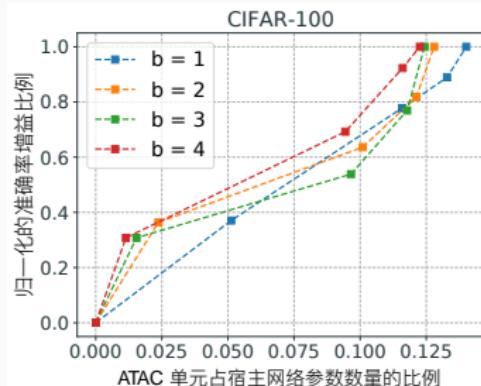
微模块类型	CIFAR-10				CIFAR-100			
	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 1$	$b = 2$	$b = 3$	$b = 4$
NiN	0.893	0.917	0.922	0.926	0.743	0.776	0.792	0.796
LocalSE	0.906	0.926	0.931	0.937	0.762	0.794	0.805	0.811
ChaATAC	0.906	0.927	0.936	0.939	0.764	0.796	0.812	0.821

1. NiN 不如 LocalSE 和 ChaATAC 这两个注意力模块，表明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 ($G(X) \otimes X$ 比 $G(X)$ 非线性更强)
2. ChaATAC 好于 LocalSE，表明在给定相同参数数量和计算量的情况下，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

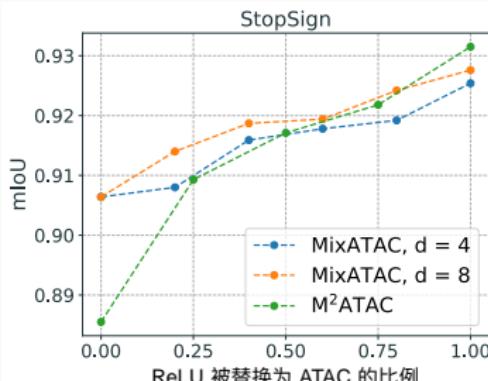
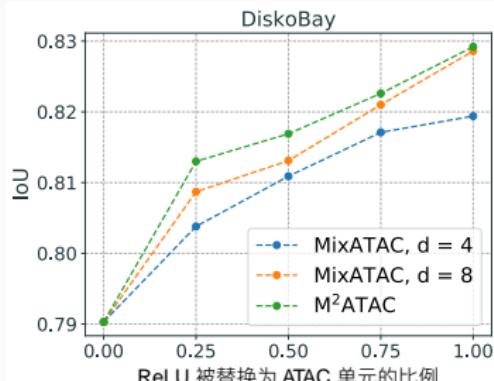
消融实验 – 全注意力网络的必要性验证



(a) CIFAR-10



(b) CIFAR-100

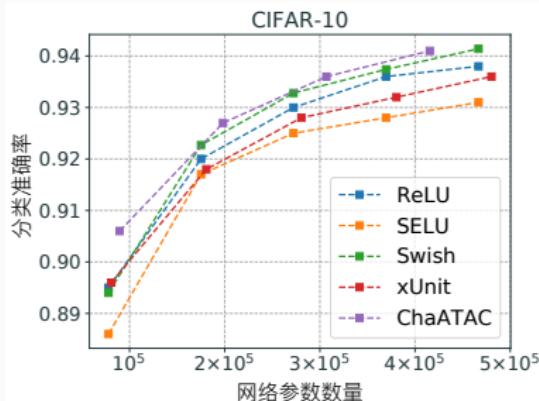


消融实验 – 全注意力网络的必要性验证

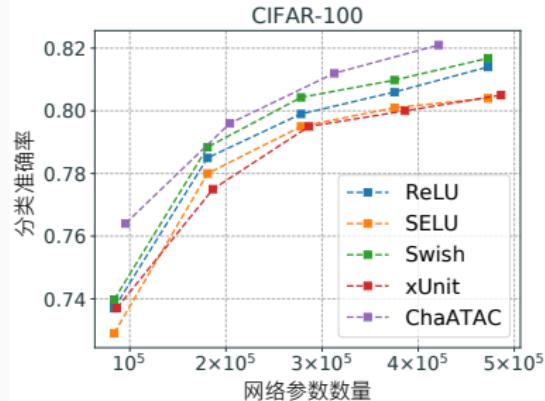
1. 对于 ResNet-20 这种深度的网络来说，随着 ATAC 单元比例的增加，网络性能在持续性地提升，直至变成全注意力网络
2. 且对于 CIFAR-10/100 上的实验来说，增长折线的最右侧，其斜率在不少折线中最大，即增加参数的单位性能增益最大
3. 印证了本章的动机，即网络早期阶段的注意力调制通过抑制无关的低层特征、突出相关特征，可以使得网络能够更为高效地编码图像的高层语义信息。

与其他方法对比 – 激活单元

多种激活单元在 CIFAR-10/100 数据集上的分类性能比较



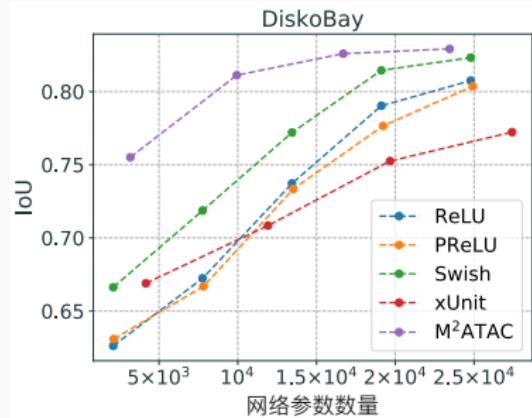
(a) CIFAR-10



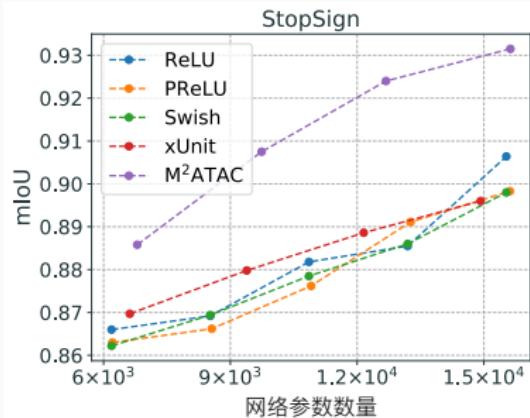
(b) CIFAR-100

与其他方法对比 – 激活单元

在 DiskoBay 和 StopSign 数据集上的分割性能比较



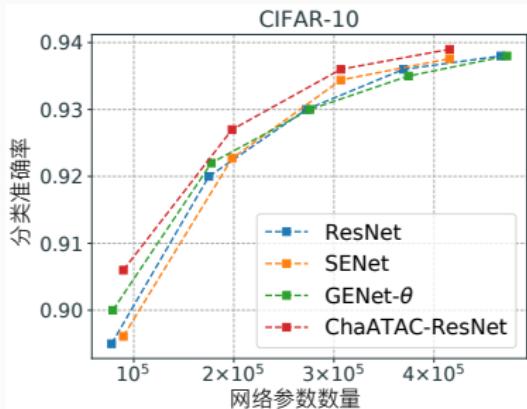
(a) DiskoBay



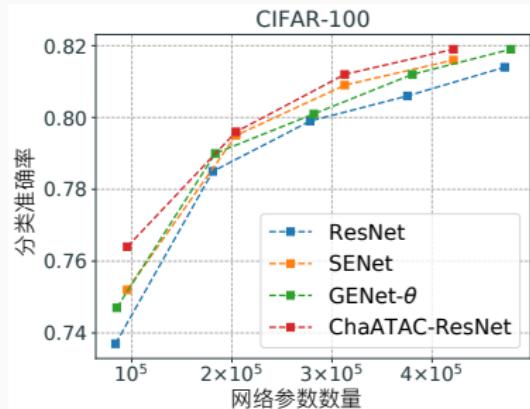
(b) StopSign

与其他方法对比 – 其他网络

与其他深度网络在 CIFAR-10/100 数据集上的分类准确率比较



(a) CIFAR-10 数据集



(b) CIFAR-100 数据集

与其他方法对比 – 其他网络

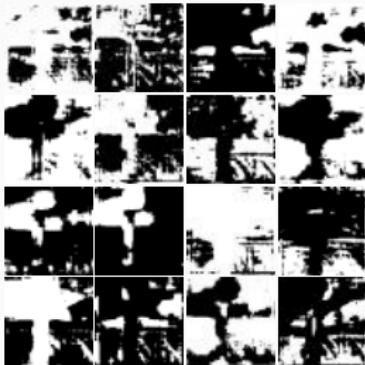
与其他深度网络在 ImageNet 数据集上的分类准确率比较

方法	GFLOPs	参数数	top-1 错误率 / %	top-5 错误率 / %
ResNet-50 [7]	3.86	25.6M	23.30	6.55
SE-ResNet-50 [8]	3.87	28.1M	22.12	5.99
AA-ResNet-50 [9]	8.3	25.8M	22.30	6.20
FA-ResNet-50 [10]	7.2	18.0M	22.40	/
GE- θ^+ -ResNet-50 [11]	3.87	33.7M	21.88	5.80
ChaATAC-ResNet-50	4.4	28.0M	21.41	6.02

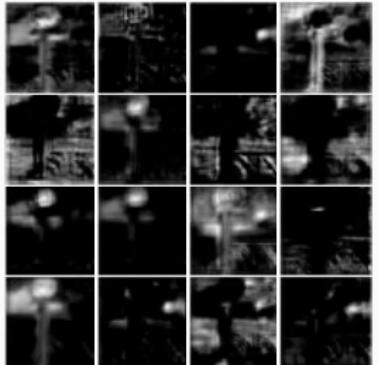
与其他方法对比 – 可视化比较



(a) ReLU 输入特征图



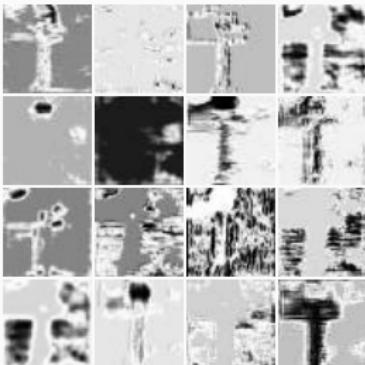
(b) ReLU 激活权重图



(c) ReLU 输出特征图



(d) xUnit 输入特征图



(e) xUnit 激活权重图

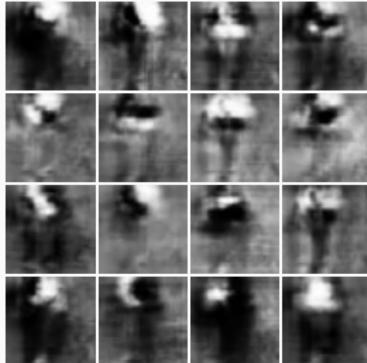


(f) xUnit 输出特征图

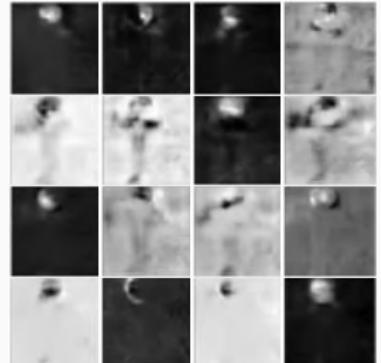
与其他方法对比 – 可视化比较



(g) M²ATAC 输入特征图



(h) M²ATAC 激活权重图



(i) M²ATAC 输出特征图

M²ATAC 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。

本章小结

1. 构造了首个开放的弱小冰山检测数据集
2. 提出了注意力激活单元
3. 第一次将尺度的概念引入通道注意力机制

基于注意力特征融合的图像分类与 小目标分割

- 相关成果：
[1] **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、训练好的模型 (157 Star, 31 Fork)：
<https://github.com/YimianDai/open-aff>

特征融合：另一个被大量研究、又总被忽视的领域

- 大量研究：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
- 总被忽视：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

无处不在的特征融合

本章背景 – 线性融合方式

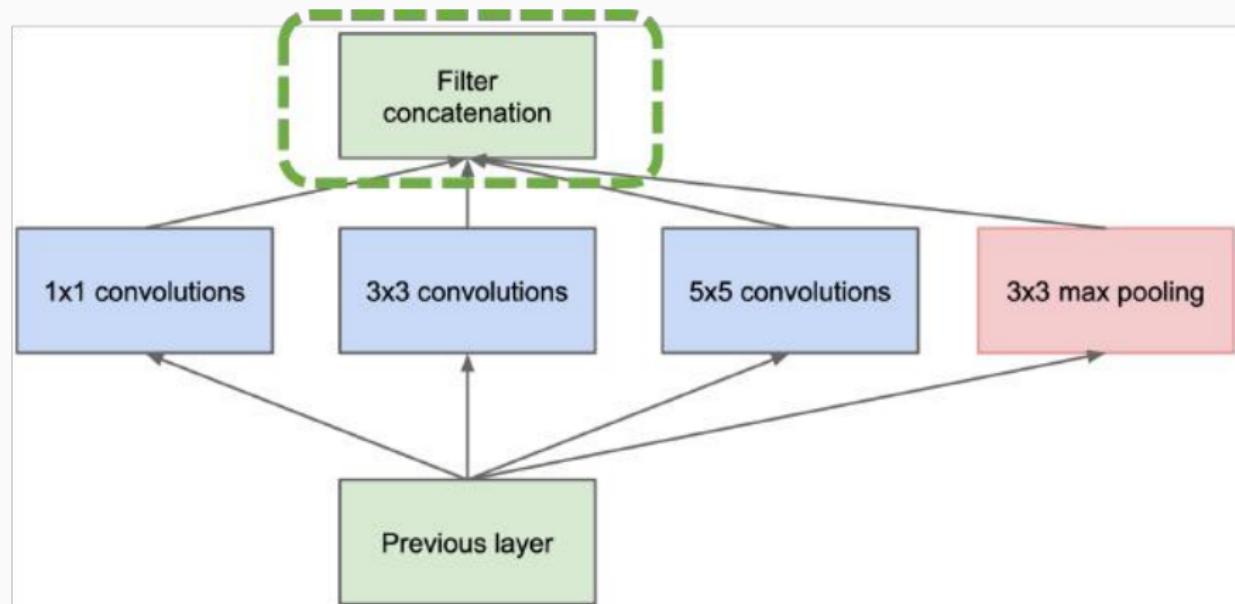


Figure 1: Inception 模块中的拼接操作 [12] – 同层融合

本章背景 – 线性融合方式

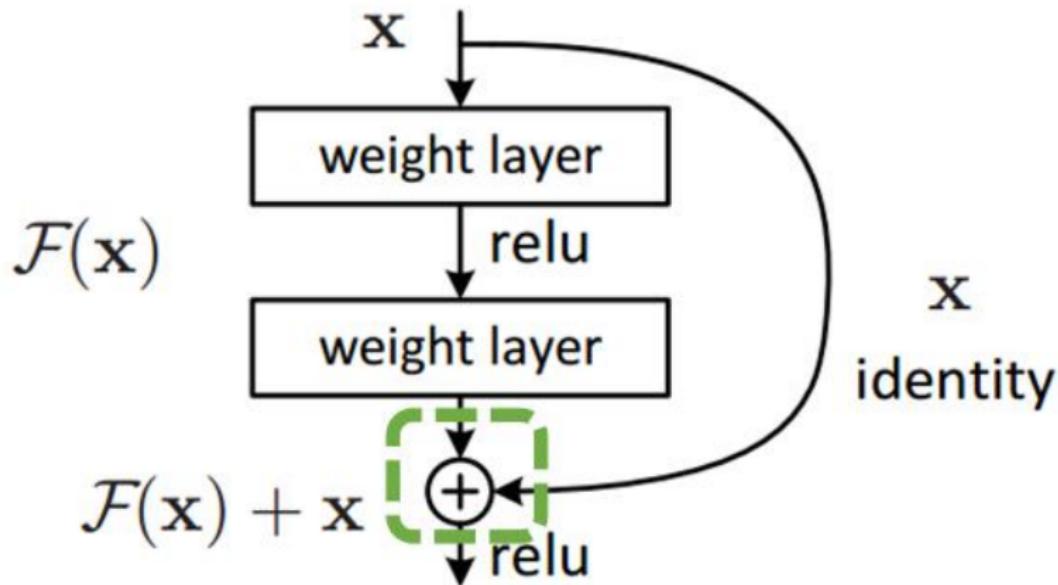


Figure 2: 残差块中的相加操作 [7] – 跨层融合中的短跳连接

本章背景 – 线性融合方式

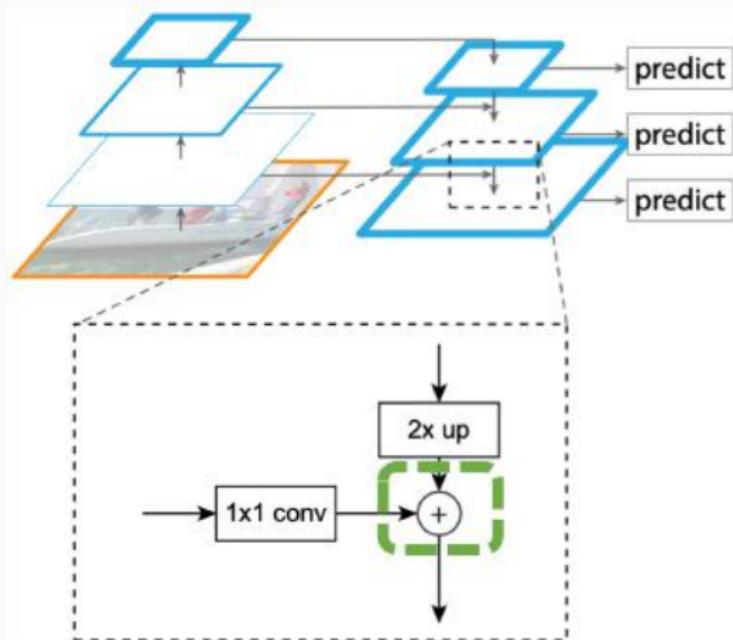


Figure 3: 特征金字塔中的相加操作 [13] – 跨层融合中的长跳连接

本章背景 – 线性融合方式

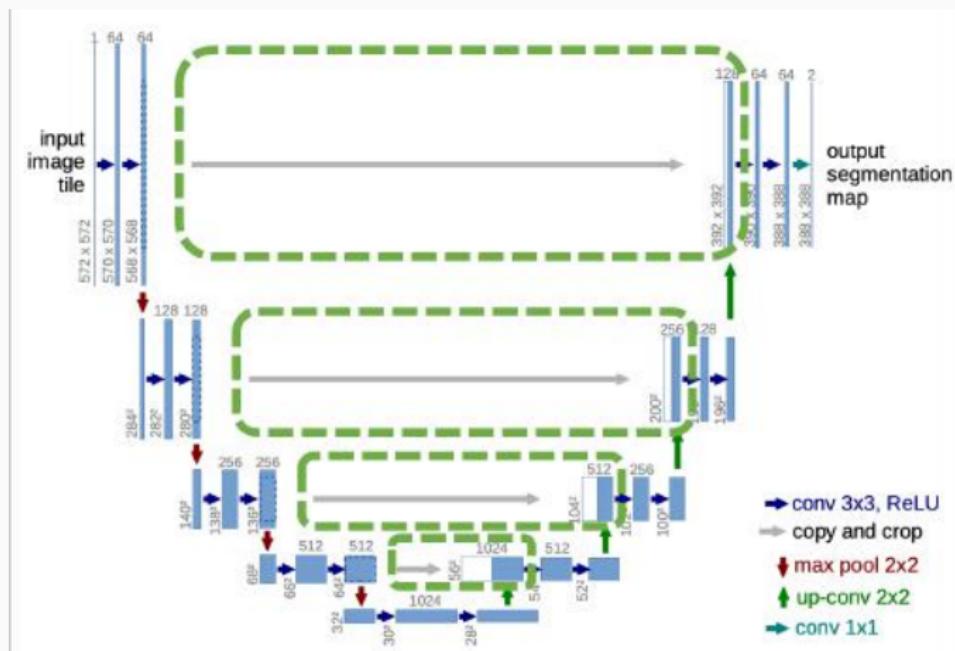


Figure 4: U-Net 中的拼接操作 [14] – 跨层融合中的长跳连接

本章背景 – 非线性融合方式

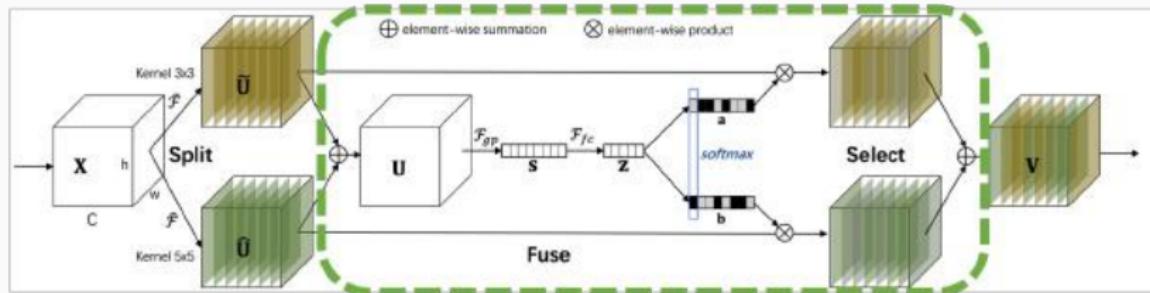


Figure 5: Selective Kernel Networks (SKNet) [15] – 同层融合

采用 SENet 中的注意力机制来融合同层特征

- 非线性、动态选择机制

SKNet 融合方式的不足：

1. 单一的融合场景
 - 跨层的特征融合尚未被讨论
2. 特征初始融合的忽视
 - SKNet 中的融合机制会不可避免地引入另一个特征融合阶段
 - 事实上，初始融合对后面的融合质量有很大的影响
3. 有偏的上下文聚合尺度
 - SENet 中的通道注意力偏向大目标
 - 问题：通道注意力可以是多尺度的吗？

1. 针对单一的融合场景
 - 将注意力特征融合拓展到短跳连接和长跳连接
 - 采用统一的融合方式来应对所有特征融合场景
2. 针对忽视特征初始融合
 - 引入另一个注意力模块来迭代地优化初始融合特征
3. 针对有偏的上下文聚合尺度
 - 想法：空间聚合（Pooling）大小控制着通道注意力的尺度
 - 构建了多尺度的通道注意力机制（Multi-Scale Channel Attention Module, MS-CAM）

一层层嵌套地优化初始融合特征



将通道注意力由单尺度推广为多尺度

方法 – 多尺度通道注意力模块

一个最极简的案例：

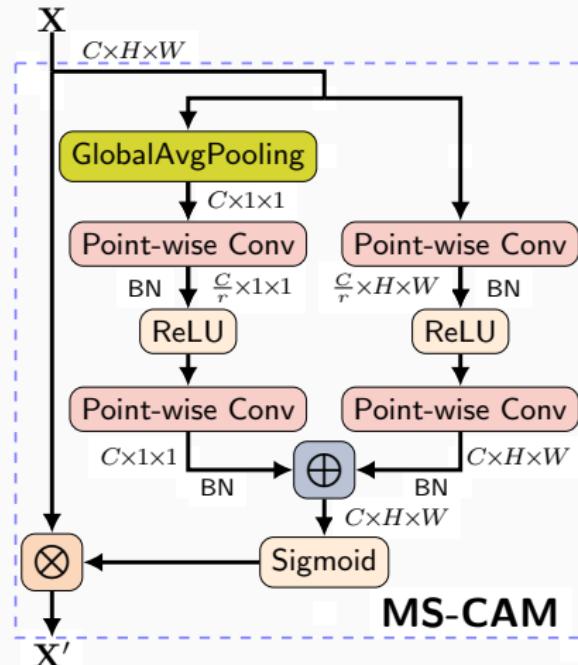


Figure 6: Multi-Scale Channel Attention Module (MS-CAM)

方法 – 注意力特征融合

需要注意的是，特征初始融合仍然是一个特征融合问题

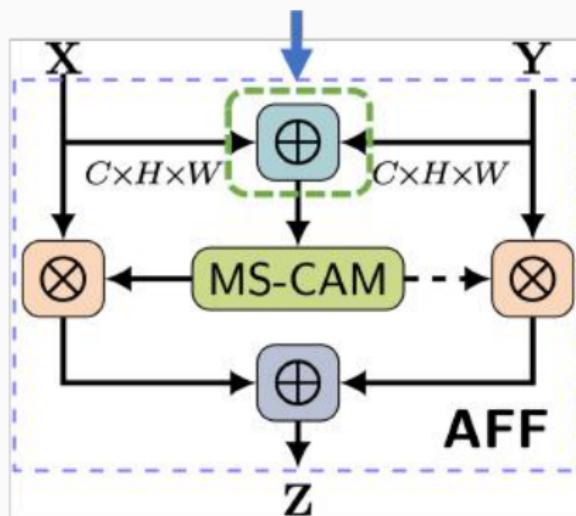


Figure 7: Attentional Feature Fusion (AFF)

方法 – 迭代的注意力特征融合

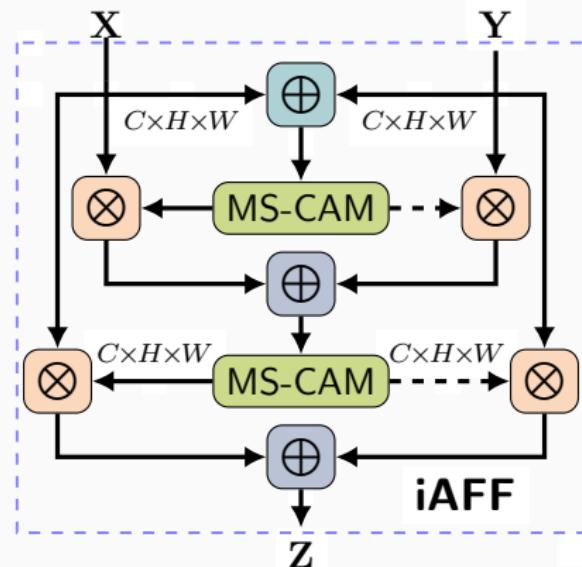


Figure 8: iterative Attentional Feature Fusion (iAFF)

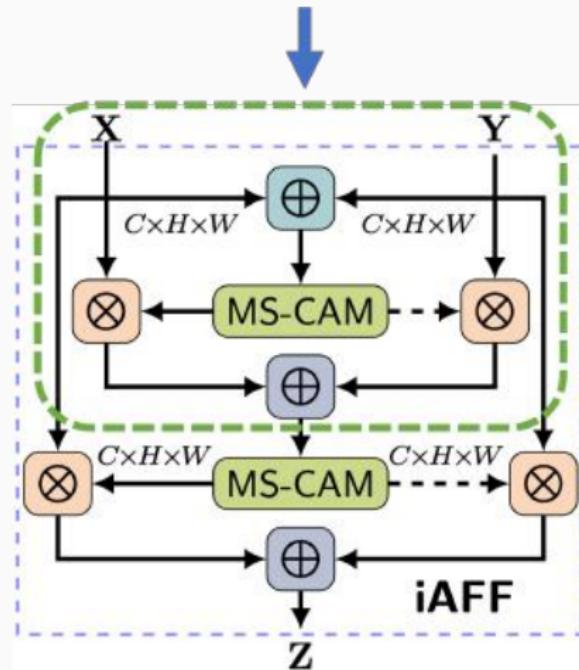
注意力特征融合，一个更加广义的特征融合框架

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]
	部分	精炼	$\mathbf{X} + \mathbf{G}(\mathbf{Y}) \otimes \mathbf{Y}$	短跳 ^[21,22,94,172]
动态、非线性	调制	调制	$\mathbf{G}(\mathbf{Y}) \otimes \mathbf{X} + \mathbf{Y}$	长跳 ^[189]
		选择	$\mathbf{G}(\mathbf{X}) \otimes \mathbf{X} + (1 - \mathbf{G}(\mathbf{X})) \otimes \mathbf{Y}$	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)

Figure 9: 深度网络中不同的特征融合方案

方法 – 迭代的注意力特征融合

用另一层的 Attention 来代替 \oplus



方法 – 网络实例

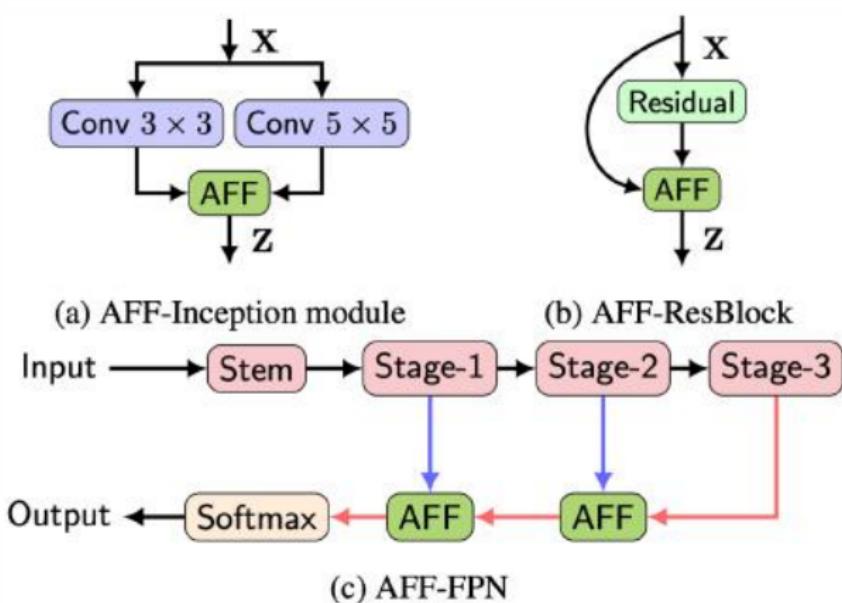


Figure 10: AFF-Inception 模块, AFF-ResBlock 和 AFF-FPN

整体实验的设计思路：

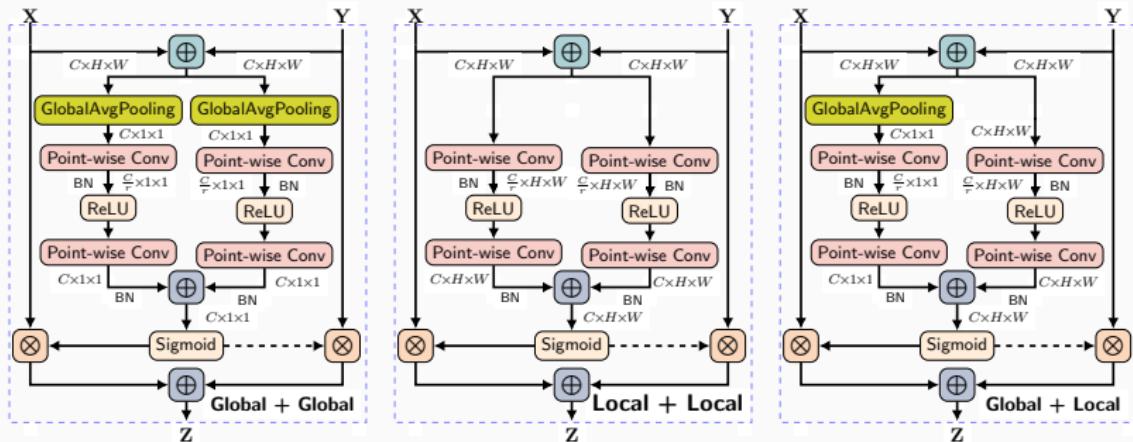
1. 消融实验

- 1.1 多尺度上下文聚合的必要性
- 1.2 对特征集成方式的比较研究
- 1.3 对定位与小目标识别性能的可视化比较

2. 与其他方法对比

消融实验 – 多尺度上下文聚合的必要性

用于验证多尺度上下文聚合必要性的消融实验架构：



注意：参数数量相同

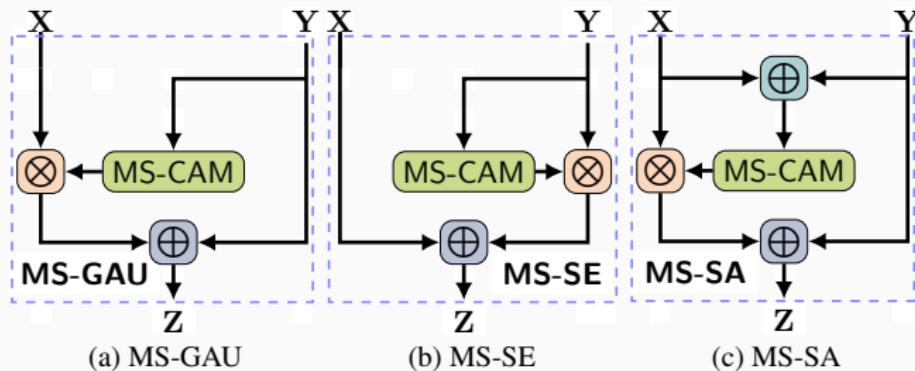
消融实验 – 多尺度上下文聚合的必要性

用于验证多尺度上下文聚合必要性的实验结果：

Aggregation Scale	InceptionNet on CIFAR-100				ResNet on CIFAR-100				ResNet + FPN on StopSign				ResNet on
	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 1$	$b = 2$	$b = 3$	$b = 4$	$b = 1$	$b = 2$	$b = 3$	$b = 4$	ImageNet
Global + Global	0.735	0.766	0.775	0.789	0.754	0.796	0.811	0.821	0.911	0.923	0.936	0.939	0.777
Local + Local	0.746	0.771	0.785	0.787	0.754	0.794	0.808	0.814	0.895	0.919	0.921	0.924	0.780
Global + Local	0.756	0.784	0.794	0.801	0.763	0.804	0.816	0.826	0.924	0.935	0.939	0.944	0.784

消融实验 – 对特征集成方式的比较研究

用于对特征集成方式进行比较研究的消融实验架构：



注意：参数数量相同

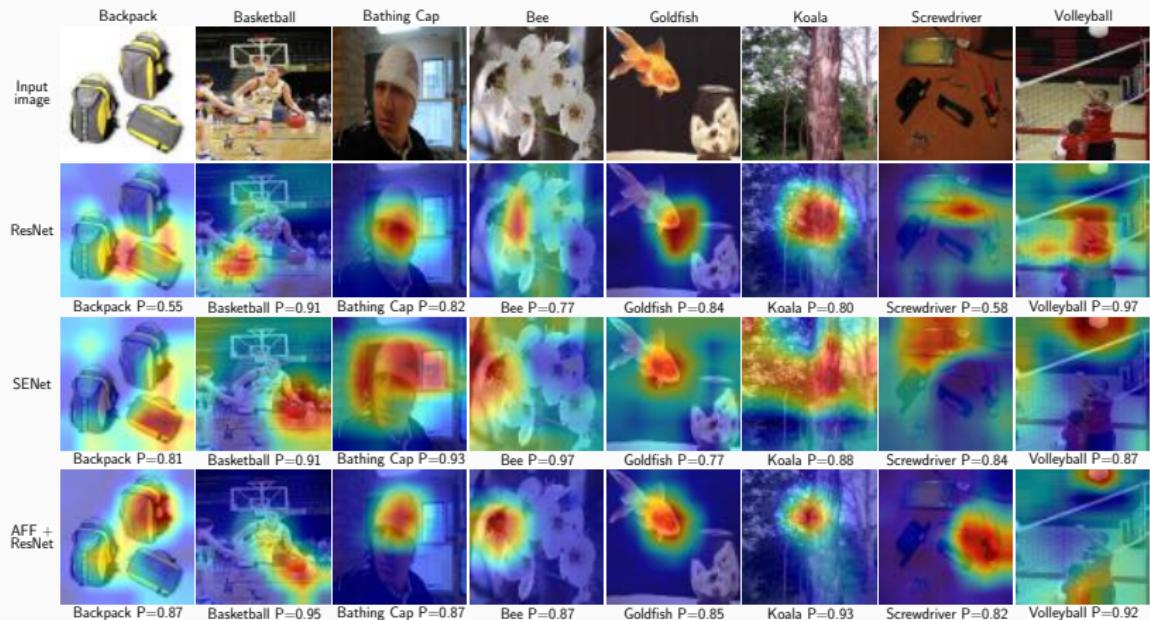
消融实验 – 多尺度上下文聚合的必要性

用于对特征集成方式进行比较研究的实验结果：

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

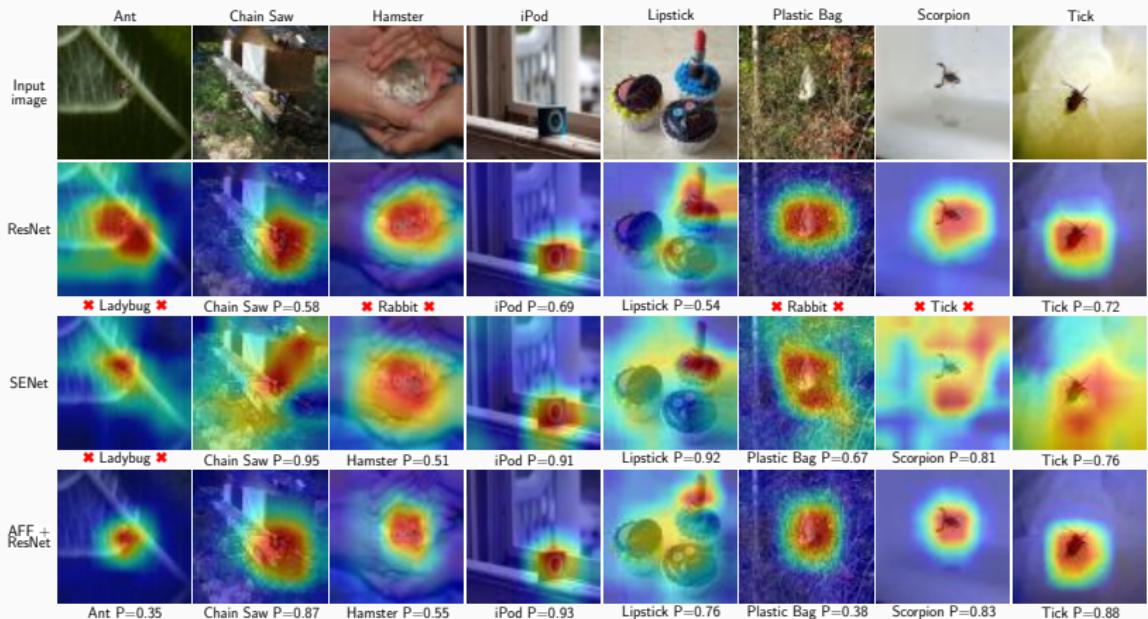
消融实验 – 对定位与小目标识别性能的可视化比较

更好的定位性能

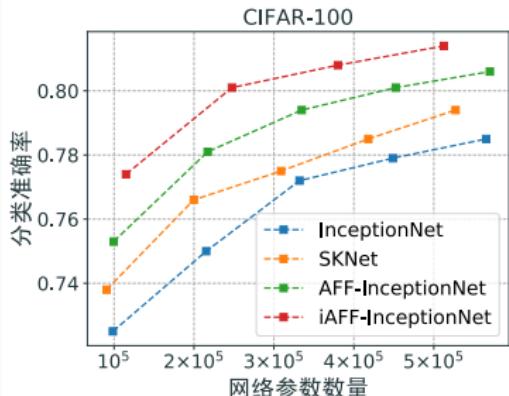


消融实验 – 对定位与小目标识别性能的可视化比较

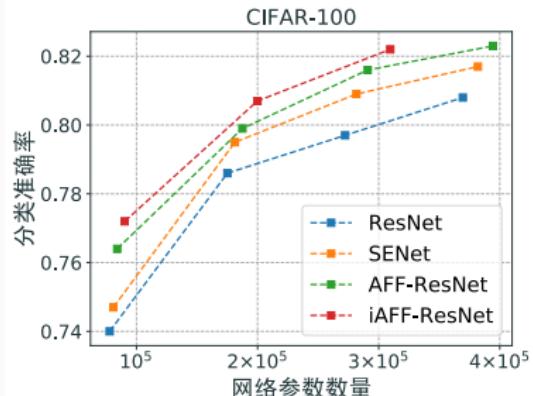
特别是对于小目标识别



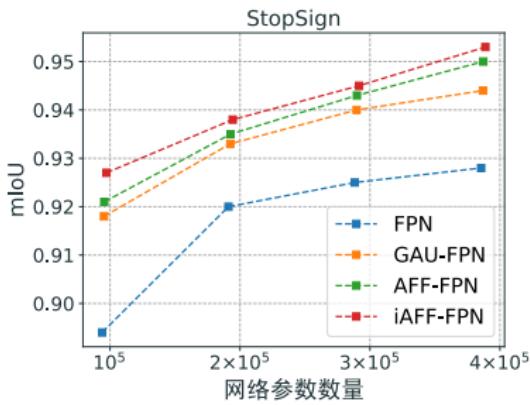
与其他方法对比 – 同层、短跳、长跳场景



(a) 同层融合



(b) 短跳连接



与其他方法对比 – ImageNet 数据集

网络	top-1 err.	参数数量
ResNet-101 [7]	23.2	42.5 M
Efficient-Channel-Attention-Net-101 [16]	21.4	42.5 M
Attention-Augmented-ResNet-101 [9]	21.3	45.4 M
SENet-101 [8]	20.9	49.4 M
Gather-Excite- θ^+ -ResNet-101 [11]	20.7	58.4 M
Local-Importance-Pooling-ResNet-101 [17]	20.7	42.9 M
<i>iAFF-ResNet-50 (本章)</i>	20.2	35.1 M
<i>iAFF-ResNeXt-50-32x4d (本章)</i>	19.8	34.7 M

本章小结

1. 一个更加广义的特征融合框架
2. 通道注意力也可以有尺度这一概念，也可以做成多尺度
3. 应当重视特征初始融合 => 迭代的注意力特征融合

基于注意力局部对比度网络的 红外小目标检测

- 相关成果：

[1] **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)

- 代码、训练好的模型：

<https://github.com/YimianDai/open-alcnet>

- 所有检测结果的可视化：

[https://github.com/YimianDai/open-alcnet/tree/
master/results/pred](https://github.com/YimianDai/open-alcnet/tree/master/results/pred)

- 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」
 - 优点：没有需要学习的参数，无需建立较大规模的数据集
 - 缺点：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
 - 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
 - 优点：前几章实验结果已经表明，性能较模型驱动方法更好
 - 缺点：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进一步提升

- 模型驱动方法
 - 建模不精确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
 - 深度学习以端到端地方式自动学习具有语义判别性的特征
- 数据驱动方法
 - 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
 - 模型驱动方法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（先验知识）
- 想法：模型驱动 + 数据驱动 => 模型驱动的深度学习

同时利用模型驱动方法的先验知识与数据驱动方法的标注数据：

1. 将传统模型驱动方法中的局部对比度量方法模块化，作为特定的非线性特征变换层嵌入深度网络中
2. 两阶段的多尺度特征融合（同层、跨层）
3. 利用特征图循环移位的加速技巧

方法 – 模块化局部对比度先验

从图像块对比度到特征图对比度

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**

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

end

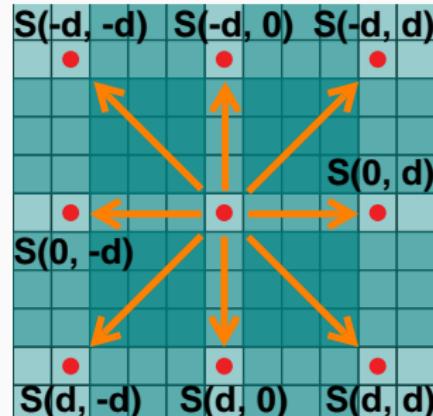
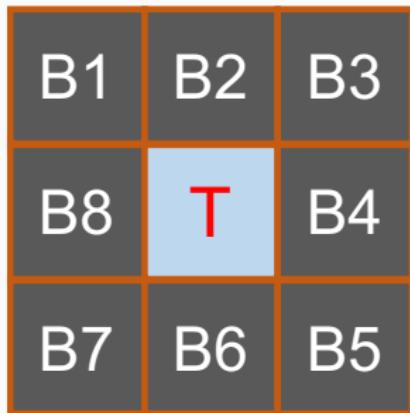
$$|\quad C_{(i,j)}^l = \min_{k=1,2,3,4} \tilde{d}_k^l(i, j);$$

end

end

方法 – 模块化局部对比度先验

从图像块对比度到特征图对比度



打破了「块大小 = 感受野 = 尺度」的硬性约束，可以在同一层特征（感受野大小相同）上进行多尺度度量

方法 – 模块化局部对比度先验

给定具体尺度 d 时，特征图上的局部对比度度量

$$\mathbf{D}_{[c,i,j]}^{(x,y)} = (\mathbf{F}_{[c,i,j]} - \mathbf{F}_{[c,i-x,j-y]}) \cdot (\mathbf{F}_{[c,i,j]} - \mathbf{F}_{[c,i+x,j+y]}), \quad (15)$$

$$\mathbf{C}_{[c,i,j]}^d = \min_{(x,y) \in \Omega} \left\{ \mathbf{D}_{[c,i,j]}^{(x,y)} \right\}. \quad (16)$$

循环很慢，如何提速？

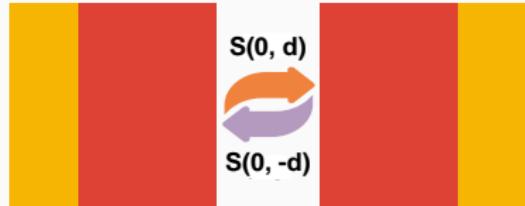
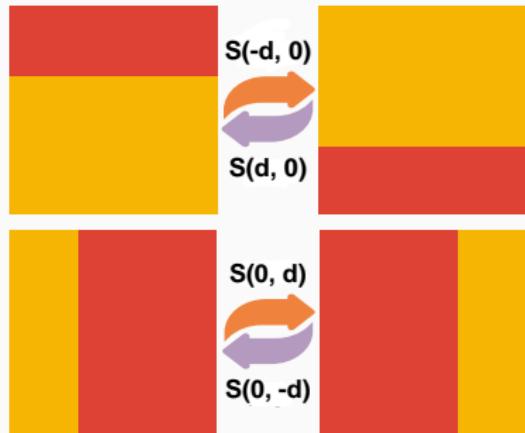
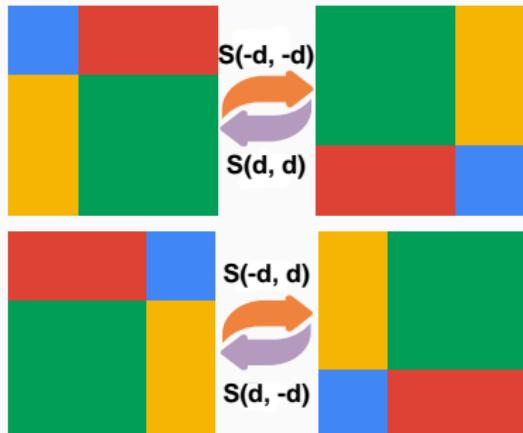
1. 用 CUDA 写像卷积那样的底层算子实现
2. 直接用卷积来实现 [18] (FLOPs 很高, $144k^2HW$)
3. 特征图的循环移位技巧 (本章方法, $8HW$)

方法 – 模块化局部对比度先验

特征图的循环移位：引入一个额外的假设

- 红外图像的边缘区域互相之间是相似且平滑过渡的

循环移位方案示意图：



方法 – 模块化局部对比度先验

特征图的循环移位：

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

$$\mathbf{D}^{(x,y)} = (\mathbf{F} - \mathbf{S}^{(x,y)}) \odot (\mathbf{F} - \mathbf{S}^{(-x,-y)}), \quad (17)$$

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

$$DLC(\mathbf{F}, d) = \min_{(x,y) \in \Omega} \left\{ \mathbf{D}^{(x,y)} \right\}. \quad (18)$$

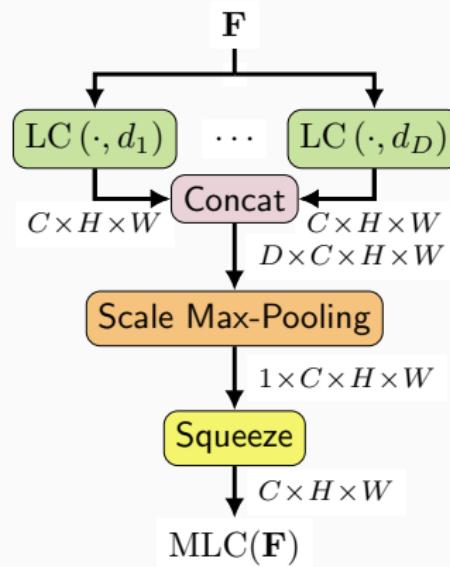
方法 – 模块化局部对比度先验

讨论：从 Attention Mechanism 的视角看该模块：

1. 在一定程度上，上式可以被看作是一种特殊且具有物理解释性的 Attention Mechanism，加强与先验知识（局部对比度）匹配的特征，抑制其余特征
2. 在较大的膨胀因子 d 下，特征图上的局部对比度量显式地打破了有限的有效感受野（ERF）[19]，并在特征图上编码了相对长程的上下文交互

方法 – 注意力局部对比度网络

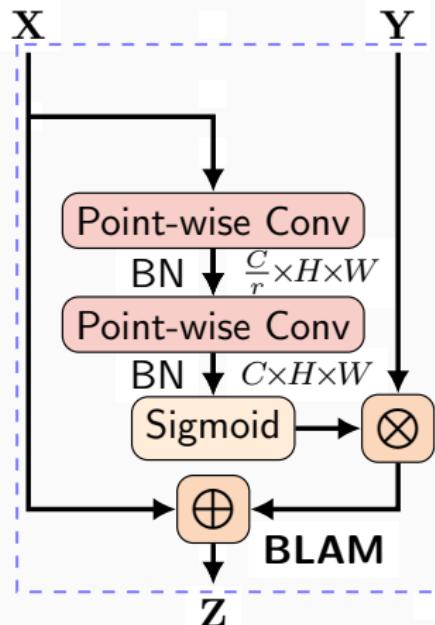
同一特征层上的多尺度局部对比度度量：变动膨胀因子 d



$$\text{MLC}(\mathbf{F}) = \text{Squeeze} (\text{SMP} (\text{Concat} (\text{DLC}(\mathbf{F}, d_1), \dots, \text{DLC}(\mathbf{F}, d_D))))$$

方法 – 注意力局部对比度网络

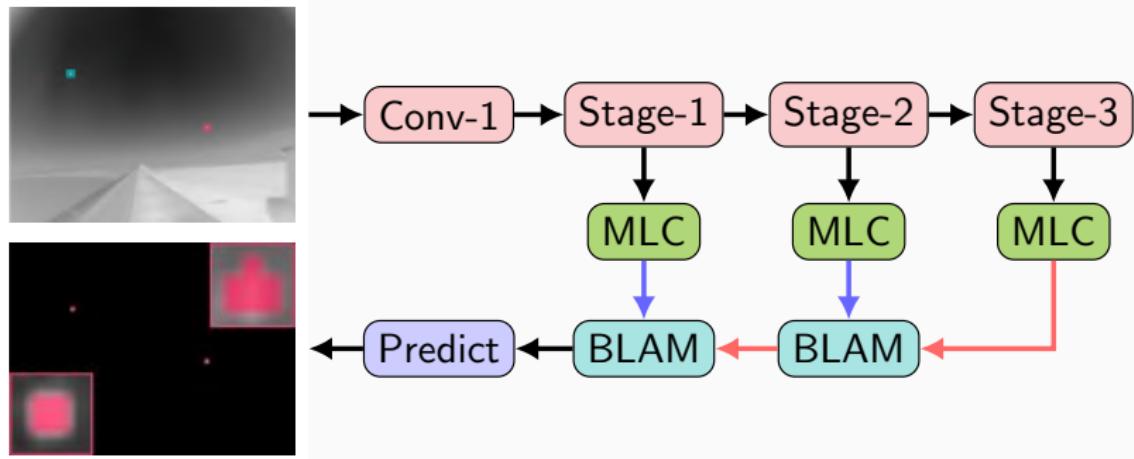
自底向上的局部对比度调制：



$$\mathbf{Z}' = \text{MLC}(\mathbf{X}) \uplus \text{MLC}(\mathbf{Y}) = \text{MLC}(\mathbf{X}) + \mathbf{L}(\text{MLC}(\mathbf{X})) \otimes \text{MLC}(\mathbf{Y})$$

方法 – 注意力局部对比度网络

网络整体架构：



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

方法 – 问题的形式化与优化

损失函数：

$$\ell_{\text{soft-IoU}}(p, y) = \frac{\sum_{i,j} p_{i,j} \cdot y_{i,j}}{\sum_{i,j} p_{i,j} + y_{i,j} - p_{i,j} \cdot y_{i,j}}, \quad (19)$$

优化目标：

$$\Theta = \arg \min_{\Theta} \sum_{n=1}^N \ell_{\text{soft-IoU}}(\sigma(M^2 LC(f, \Theta))_n, y_n), \quad (20)$$

整体实验的设计思路：

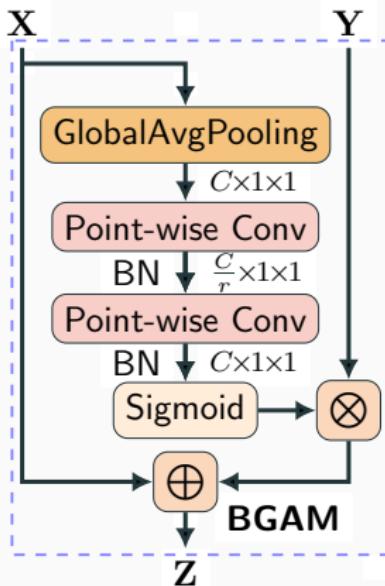
1. 消融实验

- 1.1 引入（单尺度）局部对比度先验的必要性验证
- 1.2 同层特征上的多尺度对比度度量的必要性验证
- 1.3 跨层特征融合的必要性验证
- 1.4 跨层特征融合方式的比较研究

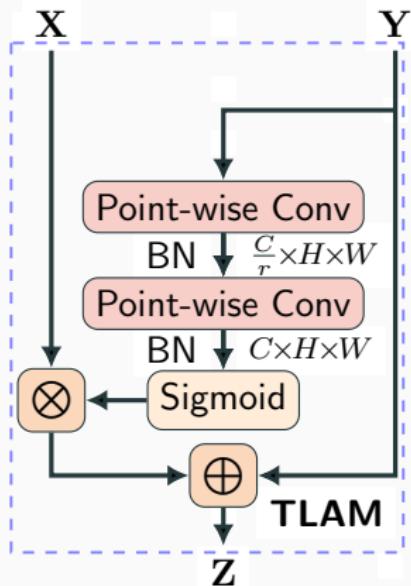
2. 与其他方法对比

消融实验

消融实验所需的跨层局部对比度特征融合方案



(a) BGAM 模块



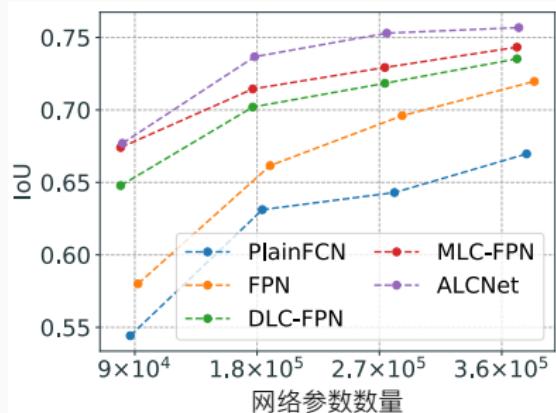
(b) TLAM 模块

消融实验

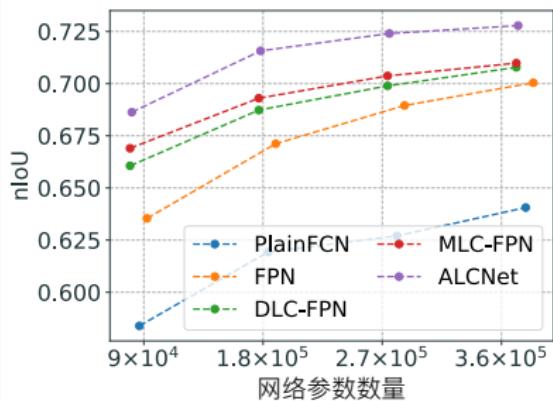
消融实验所构建网络的同层和跨层对比度度量方案

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

消融实验 – 编码局部对比度的重要性



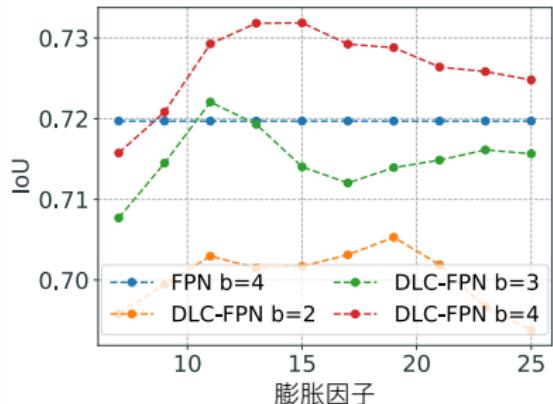
(a) IoU 比较



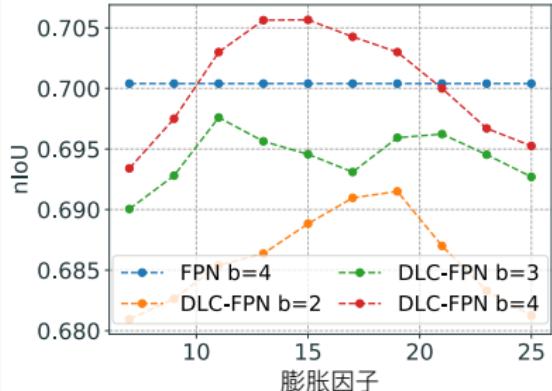
(b) nIoU 比较

1. FPN 和 DLC-FPN 网络参数数量相同，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
2. 表明编码局部对比度先验可以使得网络更加高效

消融实验 – 多尺度局部对比度特征融合的重要性



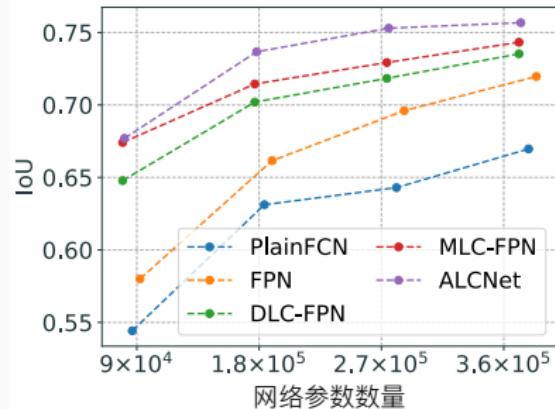
(a) IoU 比较



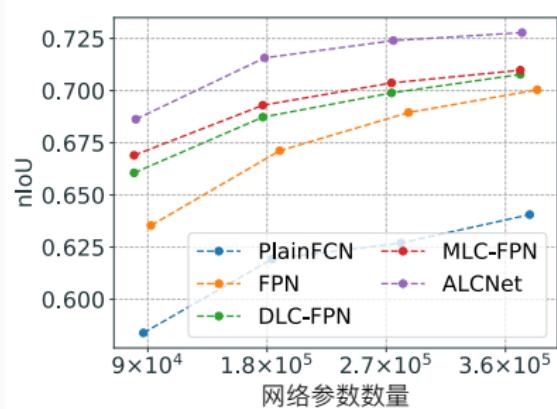
(b) nIoU 比较

- 对于 DLC-FPN 而言，合适的膨胀因子选取很重要

消融实验 – 多尺度局部对比度特征融合的重要性



(a) IoU 比较



(b) nIoU 比较

1. 对比 FCN 和 PlainFCN，表明在红外小目标检测任务上跨层的特征融合很重要
2. 对比 MLC-FPN 和 ALCNet，表明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

消融实验 – 跨层特征融合方案的重要性

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
		$\mathbf{L}(\mathbf{X}) \otimes \mathbf{X} + \mathbf{Y}$	TLA-FPN	0.688	0.729	0.750	0.753	0.688	0.708	0.722	0.718
局部	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

- 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。

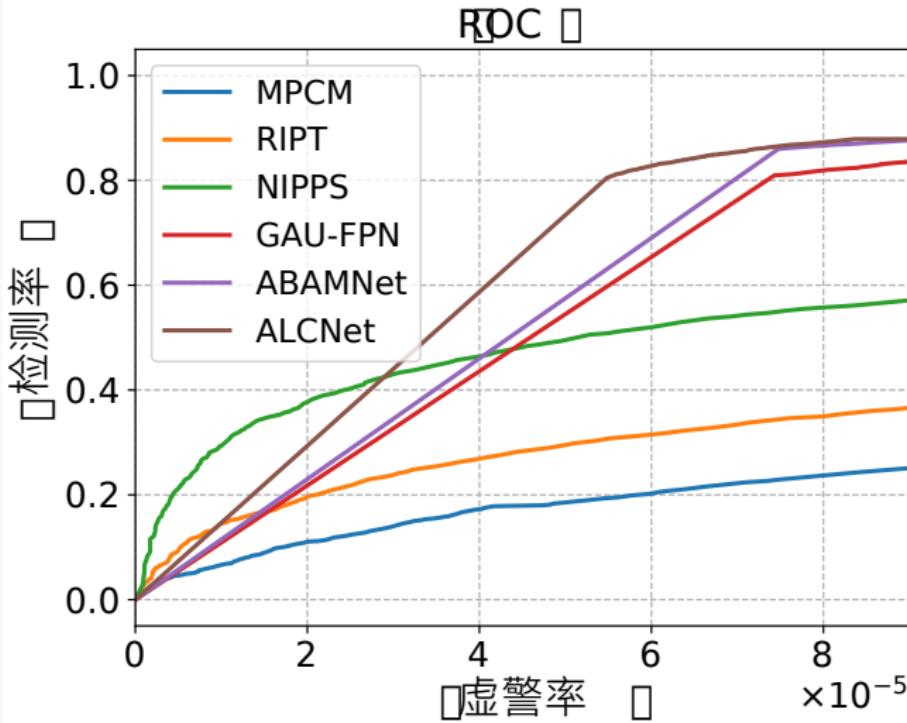
与其他方法对比

与其他 10 种方法在 IoU、nIoU 指标上的性能对比

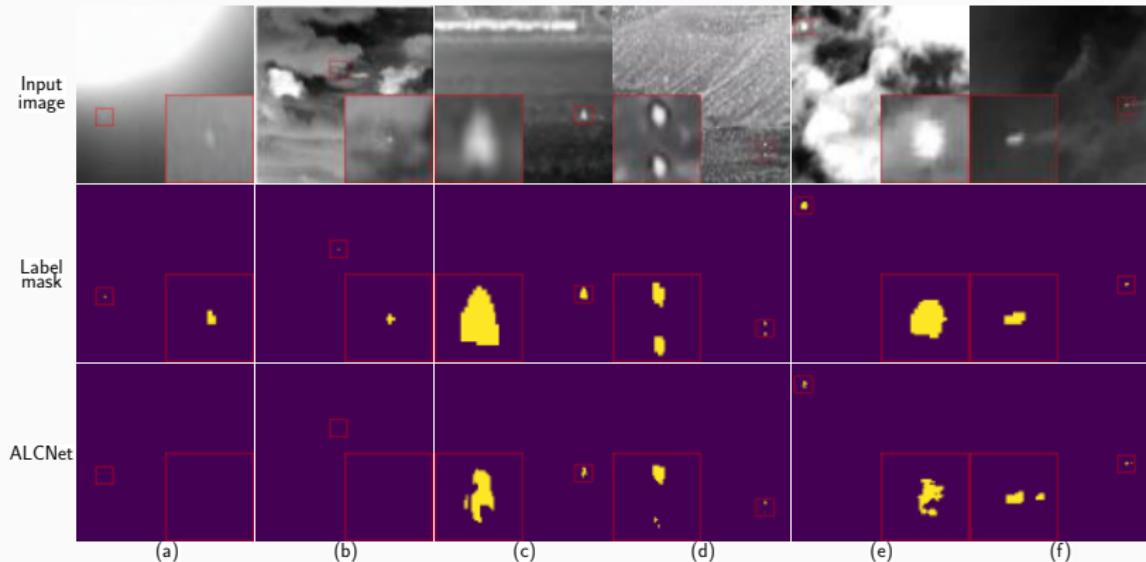
方法	SMSL	FKRW	MPCM	IPI	NIPPS	RIPT	FPN	SK-FPN	GAU-FPN	ABAMNet	ALCNet
IoU	0.081	0.268	0.357	0.466	0.473	0.146	0.720	0.702	0.701	0.731	0.757
nIoU	0.279	0.339	0.445	0.607	0.602	0.245	0.700	0.695	0.701	0.721	0.728

与其他方法对比

与其他方法在 ROC 指标上的性能对比



实验结果 – 失败案例分析



主要还是表现为漏检以及分割不完整，基本不存在虚警

本章小结

1. 构建了一个在卷积网络中嵌入局部对比度先验的检测框架
 - 1.1 采用更好的局部对比度模型应当能取得更好的效果
 - 1.2 将局部对比度模型看作是没有参数的、具有特定解释性的特殊注意力模块
2. 特征循环移位技巧
 - 2.1 传统模型也适用，例如可以将 MPCM[18] 加速 15% 以上

总结与展望

1. 构建了开放的小目标数据集 (SIRST、DiskoBay)
2. 针对小目标特点，构建了检测性能更好的模型或网络
 - RIPT 模型（第二章）以及前期工作 (WIPI、NIPPS)
 - ACM 模块（第三章）
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
 - ATAC 单元（第四章）
 - AFF 模块和 iAFF 模块（第五章）
4. 融合嵌入领域知识的传统模型和数据驱动的深度网络
 - ALCNet 单元（第六章）

1. 从小样本学习、自监督学习着手降低数据不易获取的影响
2. 无需奇异值分解的低秩稀疏分解模型深度展开
3. Roof-line 模型视角下的网络模型轻量化

谢谢各位老师

敬请批评指正

References i

- [1] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, Cham, 2014, pp. 740–755.
- [2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

References ii

- [3] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 4996–5009, 2013.
- [4] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," in British Machine Vision Conference (BMVC) 2018, Newcastle, UK, 2018, pp. 1–13.
- [5] B. Singh and L. S. Davis, "An analysis of scale invariance in object detection - SNIP," in 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, June 2018, pp. 3578–3587.

- [6] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2016, pp. 1–10.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.
- [8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 7132–7141.

- [9] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South), October 2019, pp. 3286–3295.
- [10] N. Parmar, P. Ramachandran, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," in Annual Conference on Neural Information Processing Systems (NeurIPS) 2019, Vancouver, BC, Canada, 2019, pp. 68–80.

References v

- [11] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in Annual Conference on Neural Information Processing Systems (NeurIPS) 2018, Montréal, Canada, 2018, pp. 9423–9433.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9.

- [13] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, “Feature pyramid networks for object detection,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 936–944.
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015, pp. 234–241.

- [15] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 510–519.
- [16] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 11534–11542.
- [17] Z. Gao, L. Wang, and G. Wu, "LIP: local importance-based pooling," in 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea (South). IEEE, 2019, pp. 3354–3363.

- [18] Y. Wei, X. You, and H. Li, “Multiscale patch-based contrast measure for small infrared target detection,” Pattern Recognition, vol. 58, pp. 216–226, 2016.
- [19] W. Luo, Y. Li, R. Urtasun, and R. S. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Annual Conference on Neural Information Processing Systems (NeurIPS) 2016, Barcelona, Spain, 2016, pp. 4898–4906.