

DA-Ada: Learning Domain-Aware Adapter for Domain Adaptive Object Detection

DA-Ada: 基于域感知适配器的域自适应目标检测算法

李昊宸, 张蕊, 姚涵涛, 张欣, 郝一帆, 宋新开, 李夏清, 赵永威, 李玲, 陈云霁

Accepted by The 38th Annual Conference on Neural Information Processing Systems, NeurIPS 2024 (CCF-A), vol. 37, pp. 103574-103598, 2024

联系人: 李昊宸 haochen2021@iscas.ac.cn 18810252076
李玲 liling@iscas.ac.cn 18500300237

导言

域自适应目标检测 (Domain Adaptive Object Detection, DAOD)

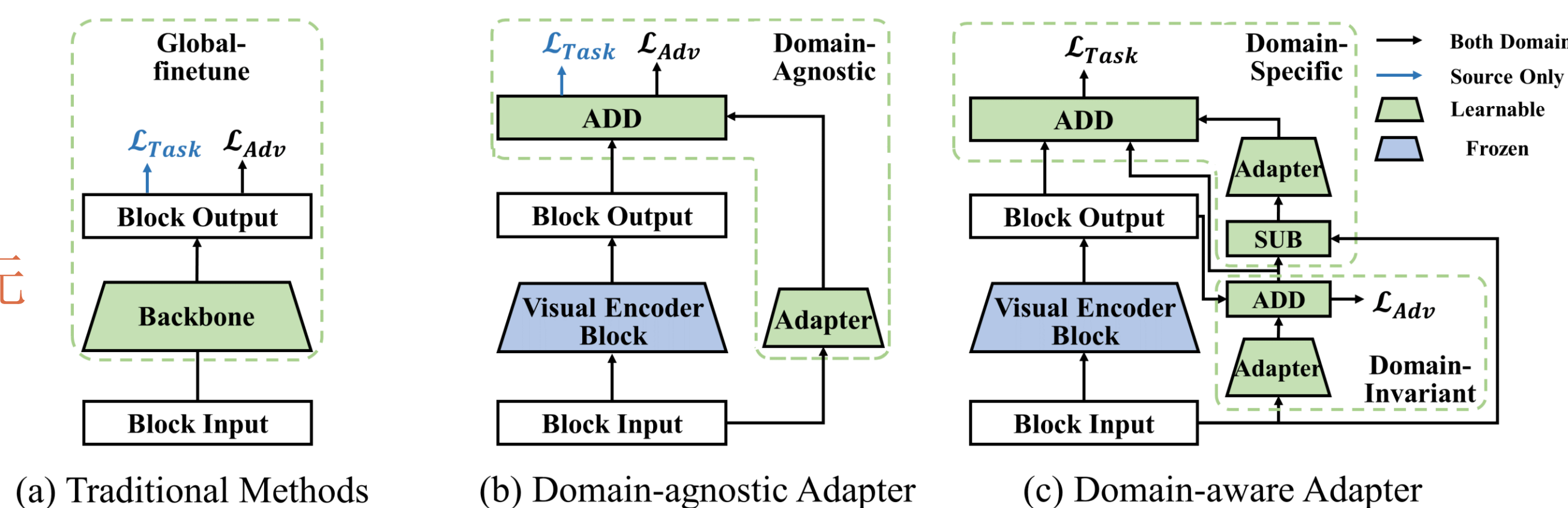
将在有标注的源域上训练的检测器泛化到无标注的目标域上, 缓解跨域性能下降问题

传统DAOD方法

- ✓思路: 通过对齐两个域的视觉特征或者生成可信的伪标签来微调特征骨干网
- ✗局限: 由于大量参数仅通过源域标签进行更新, 易于偏好源域知识

基于视觉-语言模型(Visual-Language Model, VLM)的DAOD方法

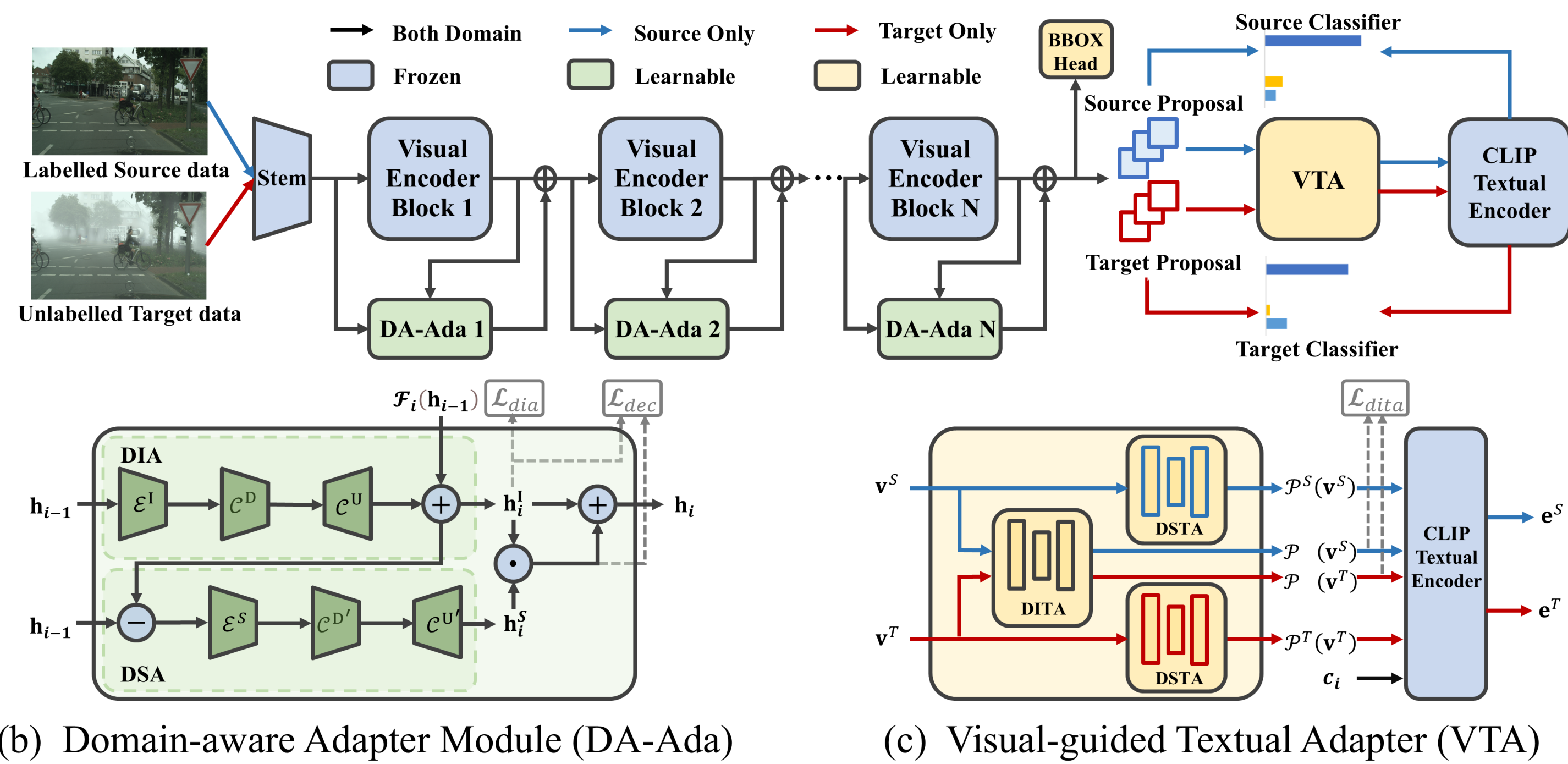
- ✓思路: 大规模预训练的编码器提取高泛化特征, 参数高效微调学习领域知识
- ✗局限: 域无关的适配器忽略了域内特性信息



域感知适配器 DA-Ada

- ✓全新框架: 我们提出全新的适配器微调框架DA-Ada, 基于特殊设计的域感知适配器, 利用视觉-语言模型提供的通用知识来探索跨域信息。这是首个将适配器微调思想引入域自适应目标检测的工作
- ✓域感知适配器: 我们设计了域感知适配器, 包含一个在域间学习域不变知识的域不变适配器, 和为每个域补充域内特性知识的域特定适配器, 在保护预训练知识的同时注入跨域信息

方法



域不变适配器DIA

域间共享, 学习域不变视觉特征

$$\begin{aligned} \mathbf{h}_i^E &= \mathcal{E}^J(\mathbf{h}_{i-1}) \\ \mathbf{h}_i^L &= \mathcal{C}^D(\mathbf{h}_i^E) \\ \mathbf{h}_i^I &= \mathcal{C}^U(\mathbf{h}_i^L) + \mathcal{F}_i^J(\mathbf{h}_{i-1}) \end{aligned}$$

域特定适配器DSA

域间独立, 学习域特定视觉特征

$$\begin{aligned} \mathbf{h}_i^D &= \mathcal{E}^S(\mathbf{h}_{i-1} - \mathbf{h}_i^I) \\ \mathbf{h}_i^{L'} &= \mathcal{C}^{D'}(\mathbf{h}_i^D) \\ \mathbf{h}_i^S &= \mathcal{C}^{U'}(\mathbf{h}_i^{L'}) \end{aligned}$$

视觉引导的文本适配器VTA

利用视觉特征中的跨域信息来增强文本编码器, 缓解纯文本的局限性

$$\begin{aligned} \mathbf{e}_i^S &= \mathcal{T}(\mathcal{P}(\mathbf{v}^S), \mathcal{P}^S(\mathbf{v}^S), c_i) \\ \mathbf{e}_i^T &= \mathcal{T}(\mathcal{P}(\mathbf{v}^T), \mathcal{P}^T(\mathbf{v}^T), c_i) \end{aligned}$$

DA-Ada算法网络

$$\mathbf{h}_i = \mathbf{h}_i^I + \mathbf{h}_i^L \cdot \mathbf{h}_i^S$$

$$\mathbf{h}_i^I = \mathcal{A}_i^J(\mathbf{h}_{i-1}) + \mathcal{F}_i(\mathbf{h}_{i-1}) \quad \mathbf{h}_i^S = \mathcal{A}_i^S(\mathbf{h}_{i-1} - \mathbf{h}_i^I)$$

\mathcal{F}_i 视觉编码块 \mathcal{A}_i^J 域不变适配器 \mathcal{A}_i^S 域特定适配器

优化目标

$$\mathcal{L} = \mathcal{L}_{det} + \lambda_{dia} \mathcal{L}_{dia} + \lambda_{dita} \mathcal{L}_{dita} + \lambda_{dec} \mathcal{L}_{dec} + \mathcal{L}_{reg}$$

实验

四组测试集上的检测结果 (mAP)

Methods	Aero	Bike	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	Motor	Person	Plant	Sheep	Sofa	Train	TV	mAP
UaDAN [20]	35.0	73.7	41.0	24.4	21.3	69.8	53.5	2.3	34.2	61.2	31.0	29.5	47.9	63.6	62.2	61.3	13.9	7.6	48.6	23.9	40.2
FGRR [41]	30.8	52.1	35.1	32.4	42.2	62.8	42.6	21.4	42.8	58.6	33.5	20.8	37.2	81.4	66.2	50.3	21.5	29.3	58.2	47.0	43.3
UMT [10]	39.6	59.1	32.4	35.0	45.1	61.9	48.4	7.5	46.0	67.6	21.4	29.5	48.2	75.9	70.5	56.7	25.9	28.9	39.4	43.6	44.1
SIGMA [37]	40.1	55.4	37.4	31.1	54.9	54.3	46.6	23.0	44.7	65.6	23.0	22.0	42.8	55.6	67.2	55.2	32.9	40.8	45.0	58.6	44.5
TIA [78]	42.2	66.0	36.9	37.3	43.7	71.8	49.7	18.2	44.9	58.9	18.2	29.1	40.7	87.8	67.4	49.7	27.4	27.8	57.1	50.6	46.3
SIGMA++ [38]	36.3	54.6	40.1	31.6	58.0	60.4	46.2	33.6	44.4	66.2	25.7	25.3	44.4	58.8	64.8	55.4	36.2	38.6	54.1	59.3	46.7
CMT [60]	39.8	56.3	38.7	39.7	60.4	35.0	56.0	7.1	60.1	60.4	35.8	28.1	67.8	84.5	80.1	55.5	20.3	32.8	42.3	38.2	47.0
DA-Ada(Ours)*	42.3	75.1	48.9	45.9	49.0	71.8	55.6	15.4	50.7	56.6	19.9	20.6	61.3	80.7	73.0	29.2	37.5	21.5	52.5	52.9	48.0

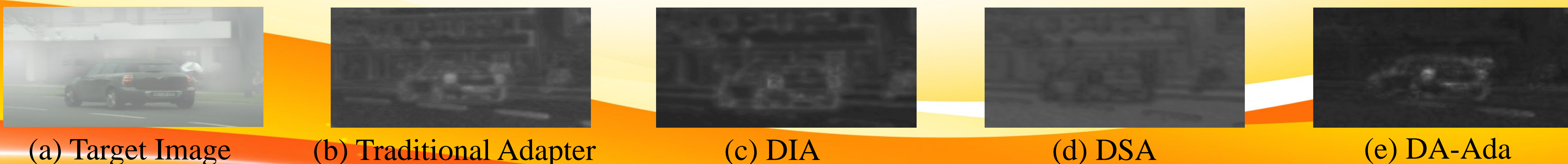
消融实验: 不同的适配器设计

Input of DIA	Input of DSA	Injection Operation	mAP
\mathbf{h}_{i-1}		$\mathcal{F}_i(\mathbf{h}_{i-1})$	52.6
\mathbf{h}_{i-1}		$\mathbf{h}_i^I = \mathcal{F}_i(\mathbf{h}_{i-1}) + \mathcal{A}_i(\mathbf{h}_{i-1})$	54.8
\mathbf{h}_{i-1}	$\mathcal{F}_i(\mathbf{h}_{i-1})$	$\mathbf{h}_i^I + \mathbf{h}_i^S$	55.2
\mathbf{h}_{i-1}	$\mathbf{h}_{i-1} - \mathcal{F}_i(\mathbf{h}_{i-1})$	$\mathbf{h}_i^I + \mathbf{h}_i^S$	56.2
\mathbf{h}_{i-1}	$\mathbf{h}_{i-1} - \mathbf{h}_i^I$	$\mathbf{h}_i^I + \mathbf{h}_i^S$	56.7
\mathbf{h}_{i-1}	$\mathbf{h}_{i-1} - \mathbf{h}_i^I$	Cross-Attention($\mathbf{h}_i^I, \mathbf{h}_i^S, \mathbf{h}_i^S$)	57.0
\mathbf{h}_{i-1}	$\mathbf{h}_{i-1} - \mathbf{h}_i^I$	$\mathbf{h}_i^I + \mathbf{h}_i^L \cdot \mathbf{h}_i^S$	57.1

Methods	C→F					K→C		S→C
	Person	Rider	Car	Truck	Bus	Train	Motor	mAP
DA-Faster [7]	29.2	40.4	43.4	19.7	38.3	28.5	23.7	32.0
SIGMA++ [38]	46.4	45.1	61.0	32.1	52.2	44.6	34.8	39.9
CIGAR [44]	46.1	47.3	62.1	27.8	56.6	44.3	33.7	41.3
CSDA [16]	46.6	46.3	63.1	28.1	56.3	53.7	33.1	39.1
HT [11]	52.1	55.8	67.5	32.7	55.9	49.1	40.1	50.3
D ² -UDA [84]	46.9	53.3	64.8	38.9	61.0	48.5	42.6	54.2
AT [39]	56.3	51.9	64.2	38.5	45.5	55.1	54.3	35.0
NSA-UDA [83]	50.3	60.1	67.7	37.4	57.4	46.9	47.3	54.3
DA-Pro [31]*	55.4	62.9	70.9	40.3	63.4	54.0	42.3	58.0
DA-Ada(Ours)*	57.8	65.1	71.3	43.1	64.0	58.6	48.8	58.5

Methods	C→F	Gains	K→C	Gains	Adapter	Source-only	Domain-agnostic	Domain-aware
Hand-crafted Prompt [79]	52.6	-	59.5	-	Cross-Weather	50.4	53.8	57.1
COOP [81]	53.5	+0.9	60.7	+1.2	Cross-FoV	57.9	63.2	65.8
DA-Pro [31]	55.1	+2.5	61.4	+1.9	Sim-to-Real	58.4	62.4	65.3
VTA(Ours)	55.8	+3.2	62.9	+3.4	Cross-Style	38.3	44.0	46.4

适配器特征可视化



(a) Target Image

(b) Traditional Adapter

(c) DIA

(d) DSA

(e) DA-Ada