SEEN-DA: SEmantic ENtropy guided Domain-aware Attention for Domain Adaptive Object Detection

SEEN-DA :语义熵引导的域感知注意力算法

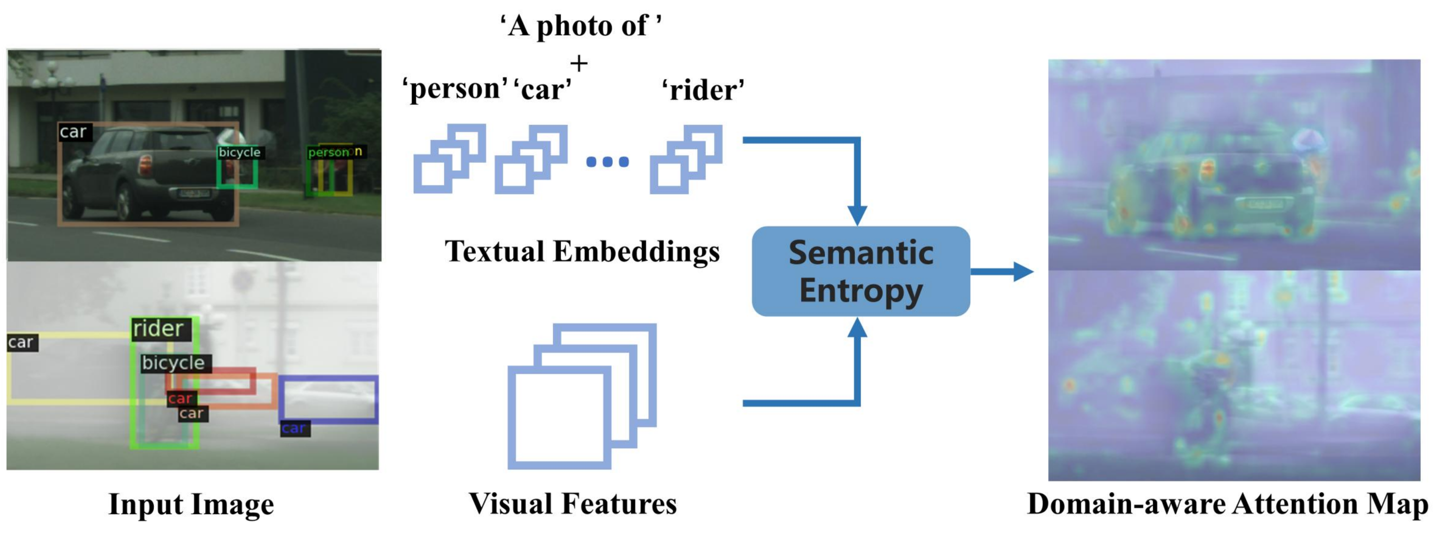
李昊宸，张蕊，姚涵涛，张欣，郝一帆，宋新开，
彭少辉，赵永威，赵琛，武延军，李玲

Proceedings of the Computer Vision and Pattern Recognition Conference,
CVPR 2025 (CCF-A), pp. 25465-25475, 2025

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

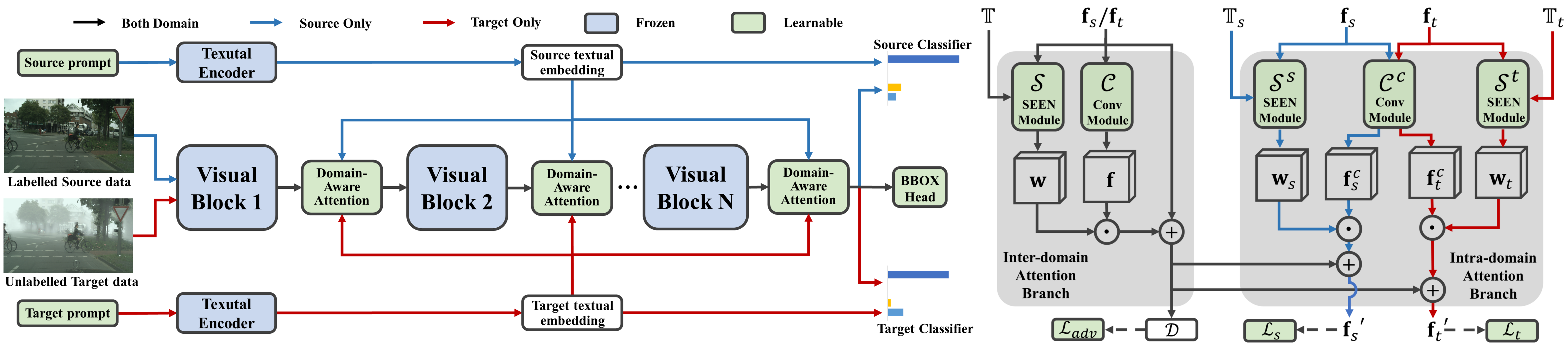
导言

- 域自适应目标检测 (Domain Adaptive Object Detection, DAOD)方法
  - ✓思路: 通过对齐两个域的视觉特征来提取域间共享的域不变特征，从而将有标注的源域上训练的检测器泛化到无标注的目标域上
  - ×局限: 语义不可知的类别标签（比如独热码）忽略了类别名称蕴含的语言属性以及域间共性和差异
- 基于视觉-语言模型(Visual-Language Model, VLM) 的DAOD方法
  - ✓思路: 微调文本提示词，来学习域间共享的域不变语义和域内特有的域特定语义
  - ×局限: 语义信息仅在检测头中被有效利用，限制了视觉特征的判别性



- 语义熵引导的域感知注意力算法 SEEN-DA
  - ✓语义熵: 我们提出全新的语义熵概念，根据特征所包含域和类别语义信息的丰度来估计视觉特征的重要性
  - ✓域感知注意力模块：基于语义熵，我们提出了域感知注意力模块。该模块作为轻量化的可学习组件，应用于冻结的视觉-语言模型的视觉编码器上，用于消除语义无关的冗余信息、激活域不变特征并补充域特定特征。

方法



- SEEN-DA 算法网络

$$\mathbf{h}_i = \mathcal{A}_i\left(\mathcal{V}_i(x)\right)$$

$\mathcal{V}_i$ 视觉编码块 $\mathcal{A}_i$ 域感知注意力模块

- 语义熵引导的域感知注意力算法 SEEN-DA

域间注意力：在域间共享，使用语义信息激活冻结的视觉编码器中的域不变特征

$$\mathbf{f} = \mathcal{C}(\mathbf{f}_d)$$

$$\mathbf{w} = SEAttention(\mathbb{T}, \mathbf{f}_d)$$

- 语义熵 Semantic Entropy

$$p(t_c, \mathbf{f}) = \frac{\exp(sim(t_c, \mathbf{f})/\tau)}{\sum_c \exp(sim(t_c, \mathbf{f})/\tau)}$$

$$SEAttention(\mathbb{T}, \mathbf{f}) = \sum_c p(t_c, \mathbf{f})\log(p(t_c, \mathbf{f})) + \log K$$

$$\mathbf{h}_i = \mathcal{A}_i\left(f_s\right) = \mathbf{w}\cdot\mathbf{f} + \mathbf{w}_{s/t}\cdot\mathbf{f}^c_{s/t} + \mathbf{f}_{s/t}$$

域内注意力：每个域独有，使用域特定语义信息丰富每个域的视觉特征

$$\mathbf{f}^c_{s/t} = \mathcal{C}^c\left(\mathbf{f}_{s/t}\right)$$

$$\mathbf{w}_{s/t} = SEAttention\left(\mathbb{T}_{s/t}, \mathbf{f}_{s/t}\right)$$

实验

- 四组测试集上的检测结果（mAP）

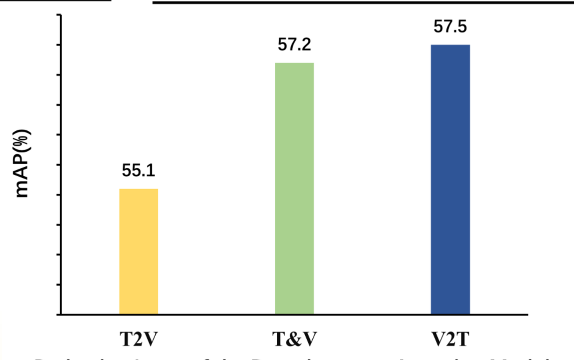
Methods	Aero	Bike	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Table	Dog	Horse	Motor	Person	Plant	Sheep	Sofa	Train	Tv	mAP
UaDAN [17]	35.0	**73.7**	41.0	24.4	21.3	69.8	53.5	2.3	34.2	61.2	31.0	**29.5**	47.9	63.6	62.2	**61.3**	13.9	7.6	48.6	23.9	40.2
TFD [54]	27.9	64.8	28.4	29.5	25.7	64.2	47.7	13.5	47.5	50.9	**50.8**	21.3	33.9	60.2	65.6	42.5	15.1	40.5	45.5	48.6	41.2
FGRR [1]	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 [6]	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 [29]	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
ATMT [27]	37.5	63.4	37.9	29.8	45.1	62.7	41.2	19.5	43.7	57.4	22.9	25.3	39.6	87.1	70.9	50.6	29.1	32.2	58.4	50.5	45.2
CIGAR [36]	35.2	55.0	39.2	30.7	60.1	58.1	46.9	31.8	47.0	61.0	21.8	26.7	44.6	52.4	68.5	54.4	31.3	38.8	56.5	**63.5**	46.2
TIA [64]	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++ [30]	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 [51]	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
RegionCLIP [65]	38.1	70.4	48.8	37.3	44.8	55.8	43.5	14.4	48.2	47.8	14.3	18.3	58.3	78.4	67.9	22.2	30.1	16.9	48.4	50.2	42.7
SEEN-DA(Ours)	**44.1**	73.4	**54.7**	**47.1**	45.1	**76.0**	51.6	20.4	51.7	53.0	18.5	17.3	61.8	86.8	72.2	22.8	**37.7**	21.1	**58.9**	52.7	**47.9**

Methods	Venue	C→F Person	C→F Rider	C→F Car	C→F Truck	C→F Bus	C→F Train	C→F Motor	C→F Bicycle	C→F mAP	K→C mAP	S→ C mAP
DA-Faster [3]	CVPR 2018	29.2	40.4	43.4	19.7	38.3	28.5	23.7	32.7	32.0	41.9	38.2
VDD [59]	ICCV 2021	33.4	44.0	51.7	33.9	52.0	34.7	34.2	36.8	40.0	-	-
SCAN [28]	AAAI 2022	41.7	43.9	57.3	28.7	48.6	48.7	31.0	37.3	42.1	45.8	52.6
TIA [64]	CVPR 2022	52.1	38.1	49.7	37.7	34.8	46.3	48.6	31.1	42.3	44.0	-
LRA [42]	TNNLS 2024	45.6	47.1	59.7	31.2	52.4	44.6	28.1	39.5	43.5	49.4	55.7
SIGMA++ [30]	TPAMI 2023	46.4	45.1	61.0	32.1	52.2	44.6	34.8	39.9	44.5	49.5	57.7
CIGAR [36]	CVPR 2023	46.1	47.3	62.1	27.8	56.6	44.3	33.7	41.3	44.9	48.5	58.5
OADA [61]	ECCV 2022	47.8	46.5	62.9	32.1	48.5	50.9	34.3	39.8	45.4	47.8	59.2
MTM [57]	AAAI 2024	51.0	53.4	67.2	37.2	54.4	41.6	38.4	47.7	48.9	-	58.1
AT [31]	CVPR 2022	56.3	51.9	64.2	38.5	45.5	**55.1**	**54.3**	35.0	50.9	-	-
SOCCER [5]	ACMMM 2024	51.7	57.7	68.6	38.2	51.6	47.5	41.6	51.7	51.1	-	63.8
DSD-DA [11]	ICML 2024	49.0	59.6	65.3	35.7	61.0	46.5	43.9	57.3	52.3	49.3	52.5
CAT [22]	CVPR 2024	44.6	57.1	63.7	40.8	**66.0**	49.7	44.9	53.0	52.5	-	-
NSA-UDA [66]	ICCV 2023	50.3	60.1	67.7	37.4	57.4	46.9	47.3	54.3	52.7	55.6	56.3
REACT [24]	TIP 2024	51.4	57.9	67.4	37.7	58.4	52.8	44.6	54.6	53.1	59.5	58.6
DA-Pro [23]	NeurIPS 2023	55.4	62.9	70.9	40.3	63.4	54.0	42.3	58.0	55.9	61.4	62.9
RegionCLIP [65]	CVPR 2022	49.6	55.0	63.2	34.1	55.6	48.3	36.0	47.0	48.6	59.1	58.9
SEEN-DA (Ours)	-	**58.5**	**64.5**	**71.7**	**42.0**	61.2	54.8	47.1	**59.9**	**57.5**	**67.1**	**66.8**

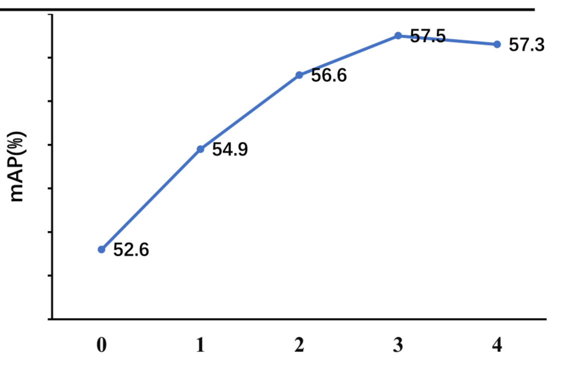
- 消融实验：不同的注意力机制设计

Benchmark	Self-Attention	Cross-Attention	Domain-aware Attention
Cross-Weather	54.8	55.5	**57.5**
Cross-FoV	61.9	63.5	**67.1**
Sim-to-Real	62.3	63.7	**66.8**
Cross-Style	43.1	45.2	**47.9**

Inter-domain Attention	$\mathcal{L}_{adv}$	Intra-domain Attention	mAP	Gains
			52.6	-
✓			54.9	+2.3
✓	✓		55.8	+0.9
✓	✓	✓	57.5	+1.7



Projection Layer of the Domain-aware Attention Module



Number of the Domain-aware Attention Modules

- 计算效率对比

Method	Backbone Parm(M)	Learnable Parm(M)	mAP	Abs. Gains
DSS [56]	26.834	26.834	40.9	+4.2
CSDA [13]	33.645	33.645	45.3	+6.9
REACT [24]	29.812	29.812	48.8	**+7.9**
AT [31]	**39.225**	18.723	50.9	**+7.9**
DA-Pro [23]	34.834	**0.008**	55.9	+3.3
SEEN-DA (Ours)	36.701	1.875	**57.5**	+4.9