FisherTune:
Fisher-Guided Robust Tuning of Vision Foundation Models for Domain Generalized Segmentation

Dong Zhao, Jinlong Li, Shuang Wang, Mengyao Wu, Qi Zang, Nicu Sebe, Zhun Zhong

University of Trento

Hefei University of Technology

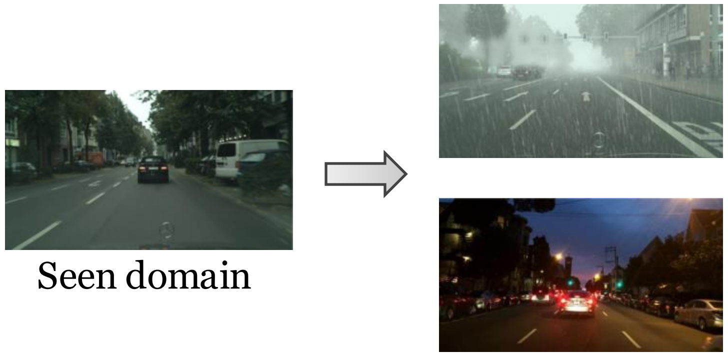
Outline

- Background
- Method
- Experiments

Background

Introduction of Domain Generalized Segmentation

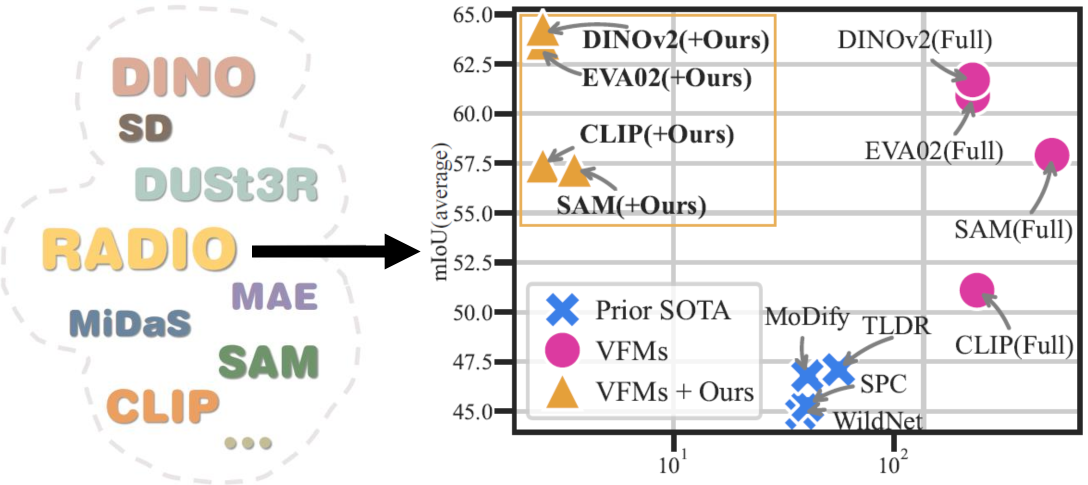
- Domain Generalized Segmentation aims to train a model on source domains that can generalize to unseen target domains without accessing their data during training.
- **Progress:** Enhancing local segmentation models → Enhancing pretrained VFMs.
- **Challenging:** Directly fine-tuning VFMs often compromises their inherent generalization ability.



Seen domain

Unseen domain

Domain Generalization



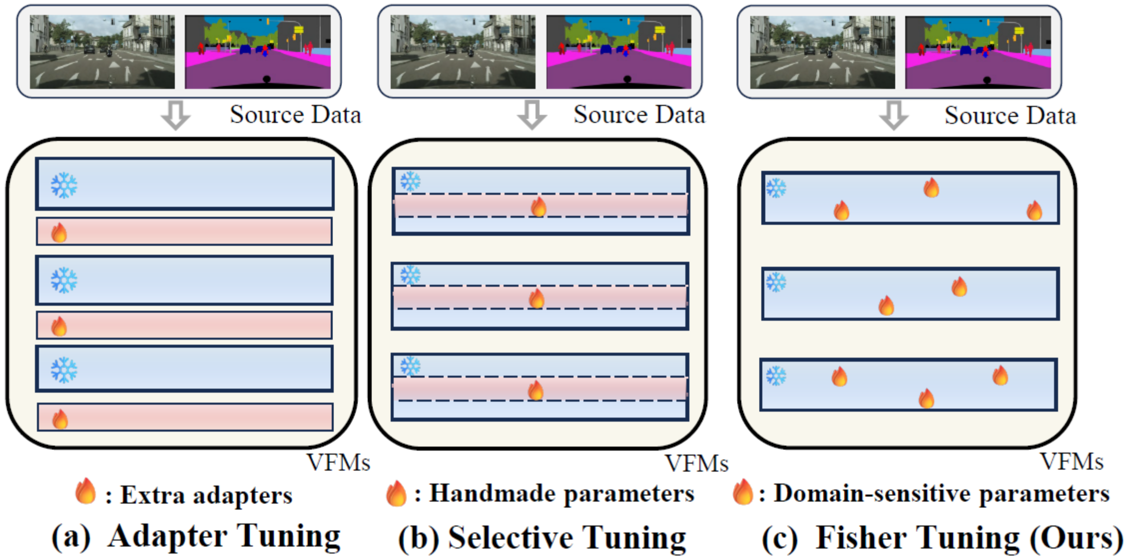
Visual Foundation Models

Better generalization ability !

Motivation

How to enhance task-specific adaptability of VFMs while preserving their generalization capability?

- **Simple way:** use PEFT methods like adapters (e.g., LoRA) or selectively fine-tuning small subset of parameters..



Disadvantages

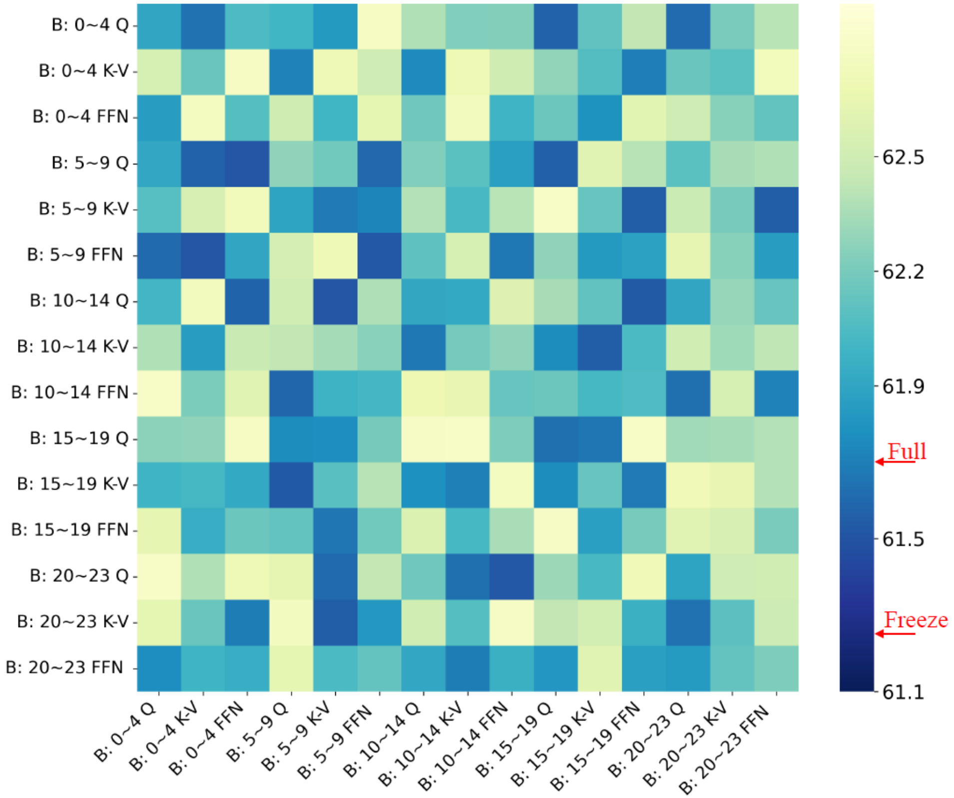
- ◆ Adapters does not fully leverage the internal representations of the VFM.
- ◆ Fine-tuning small subset of parameters fail to guarantee the generalization ability of the VFM.

Motivation

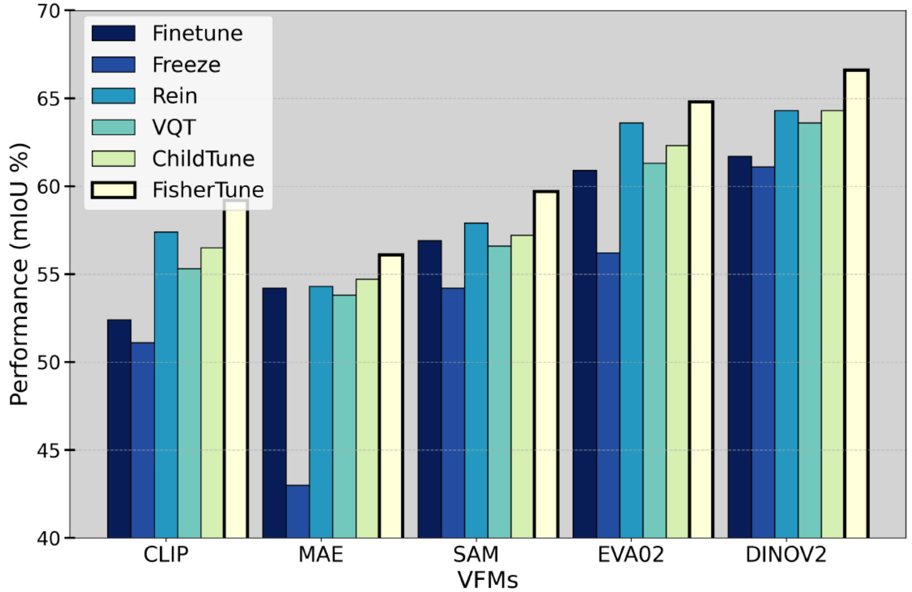
How to effectively fine-tune VFM for DG tasks ?

We find that,

- Fine-tuning different layers of VFMs yields varying impacts on generalization performance.
- Some parameters are crucial for task adaptation, while others are essential for preserving generalization.
- This suggests the existence of domain-sensitive parameters that should be selectively tuned for DG tasks.



Tuning different layers of VFM



Outline

- Background
- Method
- Experiments

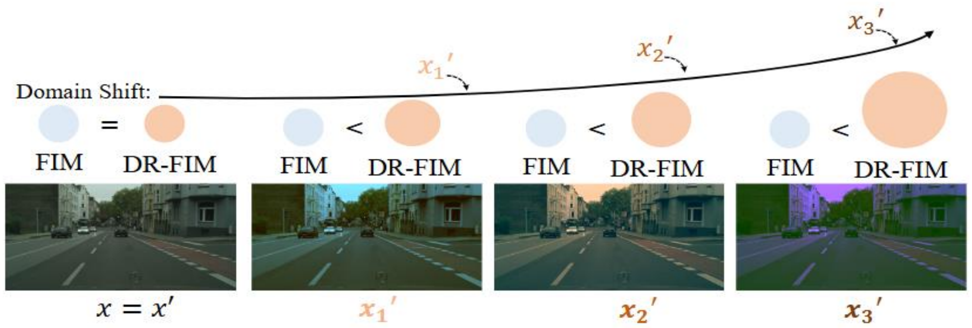
Method

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathbb{E}_x \left[\mathbb{E}_{y \sim f_\theta(y|x)} \nabla_{\boldsymbol{\theta}} \mathcal{L}(f_{\boldsymbol{\theta}}(x), y) \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(f_{\boldsymbol{\theta}}(x), y)^\top \right]$$

The Fisher Information Matrix captures task-sensitive parameters, not the domain-sensitive parameters

Higher FIM, more information $\rightarrow$ sensitive to changing
Lower FIM, less information $\rightarrow$ insensitive to changing

New metric



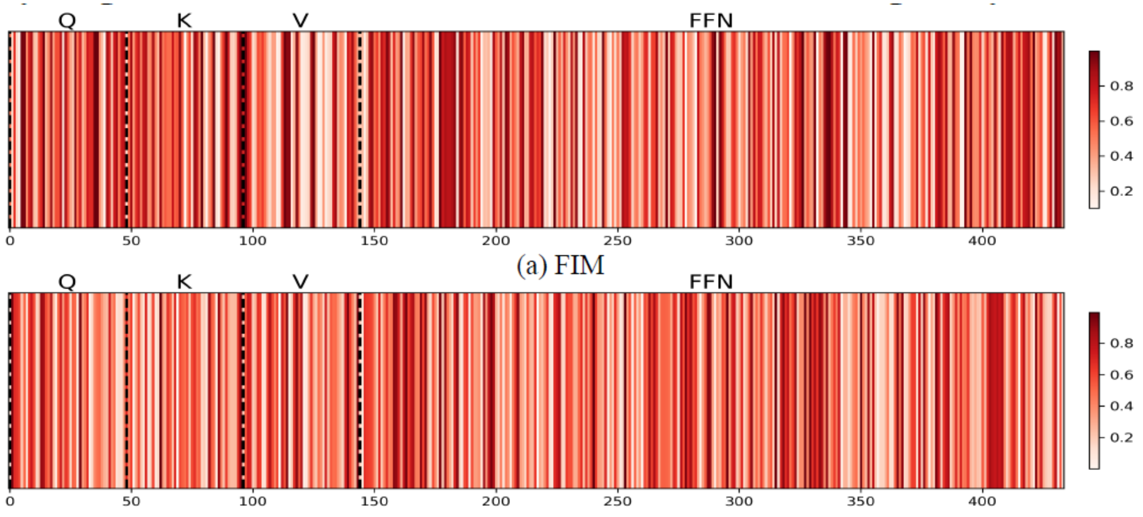
$$\Delta \mathbf{F}_{\boldsymbol{\theta}} = \frac{|\mathbf{F}_{\boldsymbol{\theta}}(x, y) - \mathbf{F}_{\boldsymbol{\theta}}(x', y)|}{\min(\mathbf{F}_{\theta_i}(x), \mathbf{F}_{\theta_i}(x')) + \epsilon}$$

Reflecting the model's varying sensitivity to parameter changes in data distributions.

$$\mathbf{DRF}_\theta = \underbrace{\mathbf{F}_\theta(x, y)}_{\text{task-sensitive}} + \underbrace{e^{-(\epsilon_\mu + \epsilon_\sigma)} \frac{|\mathbf{F}_{\boldsymbol{\theta}}(x, y) - \mathbf{F}_{\boldsymbol{\theta}}(x', y)|}{\min(\mathbf{F}_{\theta_i}(x), \mathbf{F}_{\theta_i}(x')) + \epsilon}}_{\text{domain-sensitive}} .$$

We introduce a new metric, Domain-Related FIM (DR-FIM), to account for both task-sensitive and domain-sensitive parameters

Method



(a) FIM

(b) DR-FIM without Stable Estimation

Directly estimating FIM and DR-FIM is often unstable, due to high gradient noise and sensitivity.

Stable estimation

Variational Estimation:

$$L(\hat{\theta}, \Lambda^{-1}) = \mathbb{E}_{\boldsymbol{\theta} \sim q(\boldsymbol{\theta})} \left[\mathcal{L}(\boldsymbol{\theta})\right] + \gamma\, KL(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta})).$$

We introduce the prior parameter distribution as a regularizer to prevent degradation during estimation.

According to the definition of FIM and its connection with the Hessian matrix, we have

$$\mathbb{E}_{\boldsymbol{\theta} \sim q(\boldsymbol{\theta})} \left[\mathcal{L}(\boldsymbol{\theta})\right] \approx \mathcal{L}(\hat{\boldsymbol{\theta}}) + \frac{1}{2} \operatorname{Tr}\left(\mathbf{F}_{\boldsymbol{\theta}} \Lambda^{-1}.\right)$$

Finally, the DR-FIM can be estimated as,

$$\mathbf{DRF}_{\theta} = \gamma \left(\Lambda_x - \tau^{-2} I + e^{-(\epsilon_\mu + \epsilon_\sigma)} \frac{|\Lambda_x - \Lambda_{x'}|}{\min(\Lambda_x, \Lambda_{x'}) + \frac{\epsilon}{\gamma}} \right)$$

Outline

- Background
- Method
- Experiments

Experiments | Main Results

VFM type	Fine-tune Method	Trainable Params	Citys	BDD	Map	Avg.
		GTAV → Cityscapes (Citys) + BDD100K (BDD) + Mapillary (Map)				
CLIP [49] (ViT-Large)	Full	304.20M	51.3	47.6	54.3	51.1
	Freeze	0M	53.7	48.7	55.0	52.5
	LoRA [22]	0.79M	54.0	49.8	55.1	53.0
	VPT [25]	3.69M	54.0	51.8	57.5	54.4
	Rein [63]	2.99M	57.1	54.7	60.5	57.4
	VQT [60]	3.01M	54.3	51.2	56.7	55.3
	ChildTune [66]	15.21M	57.9	53.4	58.2	56.5
	Ours	15.21M	**59.2**	**57.5**	**61.0**	**59.2**
MAE [19] (Huge))	Full	304.20M	53.7	50.8	58.1	54.2
	Freeze	0M	43.3	37.8	48.0	43.0
	LoRA [22]	0.79M	44.6	38.4	52.5	45.2
	VPT [25]	3.69M	52.7	50.2	57.6	53.5
	Rein [63]	2.99M	55.0	49.3	58.6	54.3
	VQT [60]	3.01M	53.3	50.3	57.7	53.8
	ChildTune [66]	15.21M	55.4	50.6	58.1	54.7
	Ours	15.21M	**56.6**	**51.9**	**59.7**	**56.1**
SAM [28] (Huge)	Full	632.18M	57.6	51.7	61.5	56.9
	Freeze	0M	57.0	47.1	58.4	54.2
	LoRA [22]	0.79M	57.4	47.7	58.4	54.5
	VPT [25]	3.69M	56.3	52.7	57.8	55.6
	Rein [63]	2.99M	59.6	52.0	62.1	57.9
	VQT [60]	3.01M	56.7	53.9	59.3	56.6
	ChildTune [66]	15.21M	60.8	49.6	61.2	57.2
	Ours	15.21M	**60.9**	**54.4**	**63.9**	**59.7**
EVA02 [15] (Large)	Full	304.20M	62.1	56.2	64.6	60.9
	LoRA [22]	0.79M	55.5	52.7	58.3	55.5
	AdaptFormer [7]	3.17M	63.7	59.9	64.2	62.6
	VPT [25]	3.69M	62.2	57.7	62.5	60.8
	Rein [63]	2.99M	65.3	61.1	63.9	63.4
	VQT [60]	3.01M	61.3	55.1	62.2	59.5
	ChildTune [66]	15.21M	61.6	59.3	62.3	61.1
	Ours	15.21M	**65.8**	**61.5**	**66.0**	**64.4**

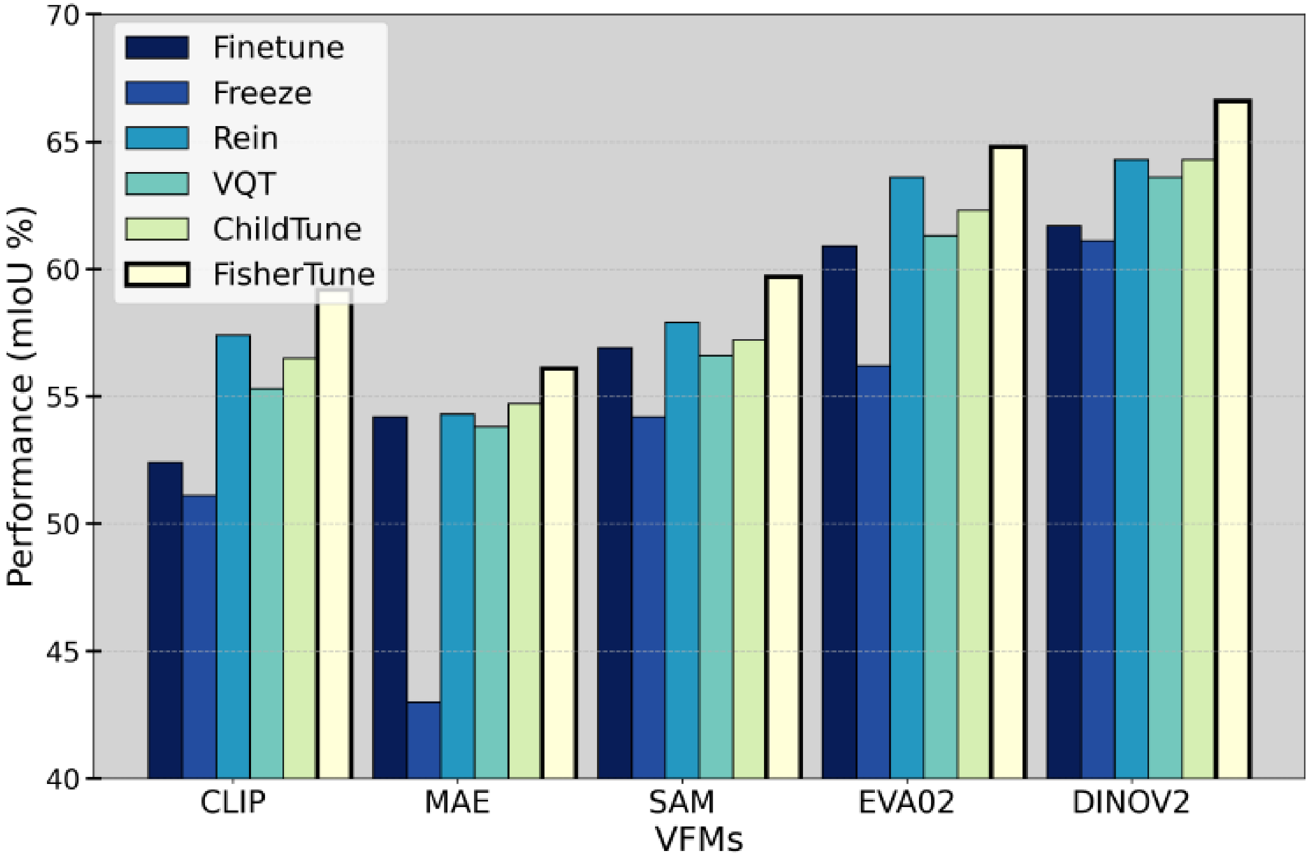


Table 1: Comparison of average performance across multiple VFMs in DGSS experiments on GTA → Cityscapes + BDD100K + Mapillary using different fine-tuning methods.

Experiments | Main Results

	Fine-tune Method	Trainable Params	road	side.	build.	wall	fence	pole	light	sign	vege	terr.	sky	pers.	rider	car	truck	bus	train	moto.	bicy.	mIoU
Cityscapes → BDD100K																						
DINOv2	Full	304.20M	89.0	44.5	89.6	51.1	46.4	49.2	**60.0**	38.9	89.1	47.5	91.7	**75.8**	48.2	91.7	52.5	82.9	**81.0**	30.4	49.9	63.7
	Freeze	0M	92.1	55.2	90.2	57.2	48.5	49.5	56.7	47.7	89.3	47.8	91.1	74.2	46.7	92.2	62.6	77.5	47.7	29.6	47.2	63.3
	REIN [63]	2.99M	**92.4**	**59.1**	90.7	58.3	**53.7**	51.8	58.2	46.4	89.8	49.4	90.8	73.9	43.3	92.3	64.3	81.6	70.9	40.4	54.0	66.4
	VQT [60]	3.01M	88.3	49.9	85.9	50.7	47.9	44.3	55.6	39.2	86.1	42.8	87.5	71.3	45.4	89.4	53.5	82.6	74.9	**46.1**	**57.4**	63.1
	ChildTune [65]	15.21M	92.1	56.1	**91.0**	58.8	46.9	52.0	58.6	47.2	90.8	47.9	**93.3**	72.0	47.1	**93.0**	63.9	76.2	47.9	28.8	48.3	63.8
	Ours	15.21M	92.1	55.4	90.2	**58.9**	50.9	**54.5**	59.8	**49.1**	**92.5**	**52.8**	91.0	73.7	**51.5**	92.7	**67.4**	**82.9**	72.8	44.3	54.1	**67.7**
EVA02	Full	304.20M	89.3	46.9	89.9	47.7	45.6	50.1	56.8	42.2	88.8	48.4	89.9	75.8	49.0	90.5	45.3	69.2	55.9	44.4	55.1	62.1
	REIN [63]	0M	**93.1**	**52.7**	88.0	47.4	31.1	41.7	46.0	39.6	85.7	41.4	89.5	67.5	39.7	89.0	47.0	72.8	46.3	19.2	35.2	56.5
	VQT [60]	2.99M	91.7	51.8	90.1	**52.8**	48.4	48.2	56.0	42.0	89.1	44.1	90.2	74.2	47.0	91.1	**54.5**	**84.1**	**78.9**	**47.2**	59.4	65.3
	ChildTune [65]	3.01M	90.1	46.6	91.1	46.9	46.4	51.7	56.5	43.2	89.3	49.6	92.3	75.0	50.3	90.3	44.6	71.8	57.4	44.0	55.8	62.8
	ChildTune	15.21M	91.4	50.7	88.9	47.9	47.4	**54.6**	56.3	**45.9**	91.2	50.0	91.2	76.1	**52.2**	92.3	48.0	69.3	55.2	43.9	59.8	63.8
	Ours	15.21M	92.6	49.9	**95.9**	51.1	**53.0**	50.8	**59.8**	45.7	**92.9**	**54.6**	**94.0**	**83.5**	52.2	**93.9**	45.1	69.4	57.1	47.2	**62.4**	**65.8**
Cityscapes → ACDC																						
DINOv2	Full	304.20M	92.8	75.0	87.4	55.7	54.1	55.6	71.2	69.6	82.4	56.0	92.2	66.8	45.6	89.0	79.7	87.9	87.5	51.4	62.7	71.7
	Freeze	0M	86.0	68.1	80.2	52.4	47.8	48.2	65.5	65.3	80.0	54.7	86.2	65.0	44.9	86.4	73.3	80.5	86.9	50.1	60.9	67.5
	REIN [63]	2.99M	94.6	78.3	92.0	**61.9**	55.0	64.8	73.8	72.7	**88.4**	**67.4**	95.4	**77.1**	**60.2**	92.6	84.1	86.9	92.5	**67.6**	**68.6**	**77.6**
	VQT [60]	3.01M	93.3	76.4	89.2	55.0	53.9	53.9	72.0	67.3	83.4	55.3	95.1	67.7	47.0	90.5	81.6	86.3	88.2	50.1	61.9	72.0
	ChildTune [65]	15.21M	92.9	72.8	84.7	56.6	54.1	56.8	70.9	67.7	82.3	55.7	93.6	65.9	45.3	89.6	77.6	87.8	87.0	52.5	62.2	71.4
	Ours	15.21M	**95.6**	**79.0**	**96.5**	60.5	**58.3**	**64.9**	**75.6**	**77.7**	85.0	61.3	**98.6**	73.6	51.5	**94.8**	**85.4**	**94.7**	**93.8**	59.0	66.7	77.5
EVA02	Full	304.20M	90.2	68.8	81.0	53.7	49.9	48.1	68.7	64.2	80.1	57.4	88.1	68.8	41.8	89.7	74.1	82.1	89.7	50.0	56.8	68.6
	Freeze	0M	86.0	60.5	76.3	49.0	41.7	46.1	60.5	61.0	72.1	49.8	77.7	56.7	40.6	80.3	68.3	77.2	85.5	46.7	56.4	62.8
	REIN [63]	2.99M	88.7	71.8	81.7	55.2	51.7	50.5	70.5	64.9	83.7	59.0	**90.3**	**72.0**	48.3	**93.0**	79.3	83.3	91.3	50.8	62.0	70.9
	VQT [60]	3.01M	90.3	71.2	81.4	54.3	53.1	49.1	67.9	64.3	82.0	**60.5**	86.9	66.8	41.3	89.3	76.6	81.7	91.3	47.2	55.7	69.0
	ChildTune [65]	15.21M	86.4	68.8	81.0	54.4	50.6	48.9	69.6	64.5	83.2	57.8	88.2	69.0	47.9	90.2	74.8	82.8	90.3	51.0	61.4	69.5
	Ours	15.21M	**90.5**	**75.2**	**83.6**	**58.8**	**54.6**	**52.2**	**73.1**	**66.6**	**85.7**	60.5	90.2	70.7	**51.5**	92.3	**82.6**	**88.2**	**91.9**	**54.0**	**62.4**	**72.9**

Table 2: DGSS generalization performance for each category from the Cityscapes source domain to mixed-domain BDD100K and ACDC, with comparison methods including adaptor-based Rein and selective parameter fine-tuning methods.

Experiments | Visualization

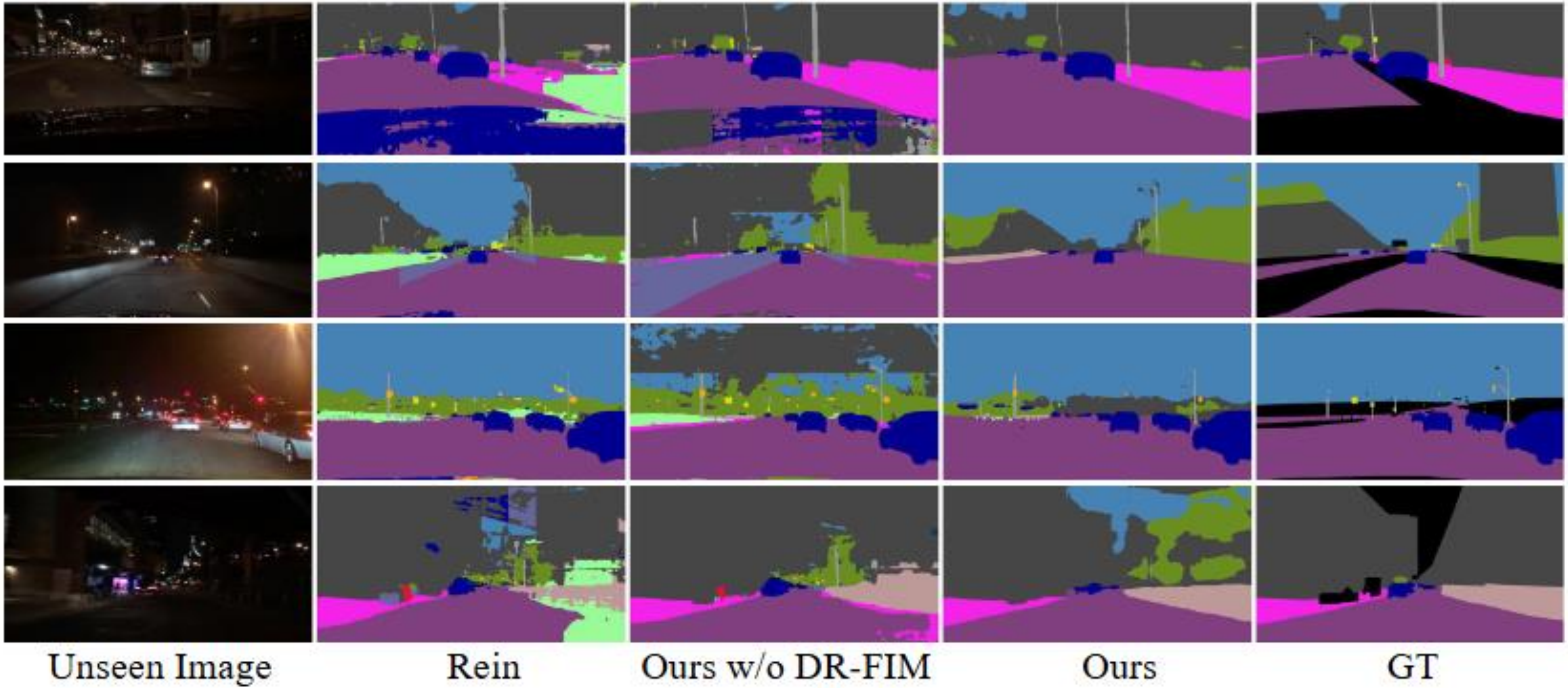


Figure 2: More structured and semantic segmentation on extreme scenarios.

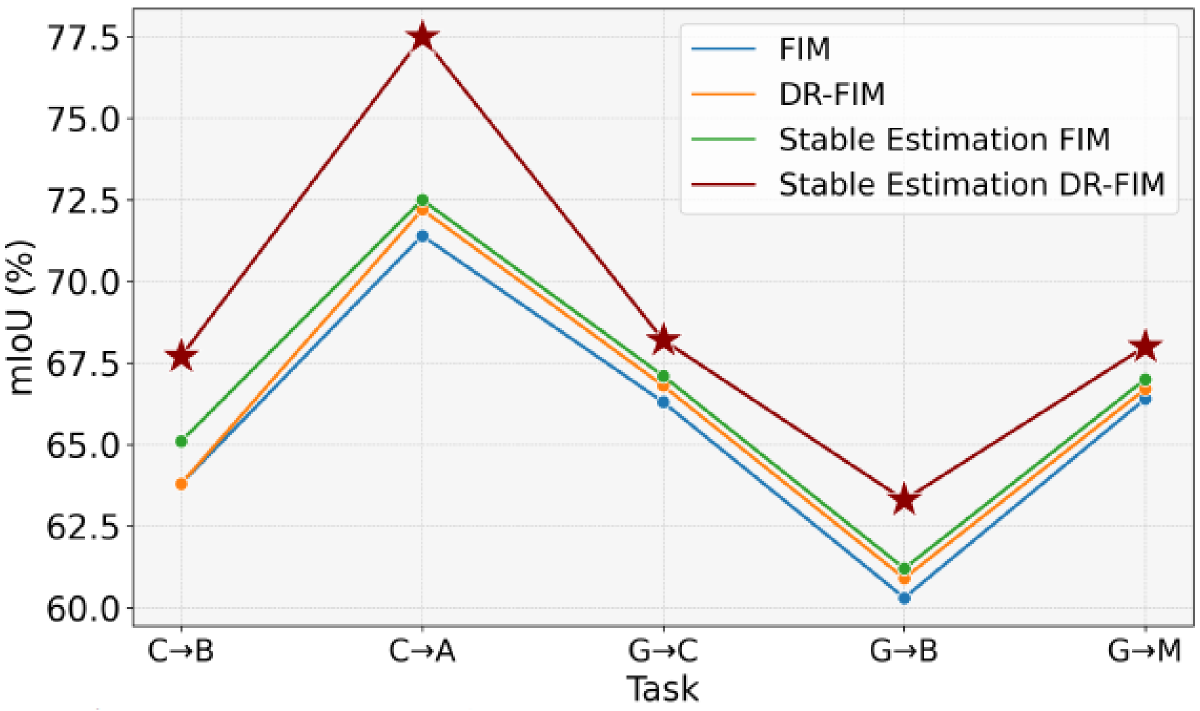
Experiments | Evaluation

Ablation Study

		Cityscapes →BDD100K	Cityscapes →ACDC
EVA02 [15] (Large)	Full	62.1	68.6
	Freeze	56.5	62.8
	Random	61.1	67.6
	Random Q	62.8	69.1
	Random K	61.9	68.1
	Random V	62.9	69.2
	$\mathbf{F}_{\theta}$	63.8	69.5
	$\Delta\mathbf{F}_{\theta}$	63.1	71.3
	$\mathbf{DRF}_{\theta}$	**65.8 (+3.7)**	**72.9 (+5.3)**
DINOv2 [5] (Large)	Full	63.7	71.7
	Freeze	63.3	67.5
	Random	62.7	71.0
	Random Q	63.2	72.0
	Random K	63.5	72.3
	Random V	63.2	72.9
	$\mathbf{F}_{\theta}$	63.8	71.4
	$\Delta\mathbf{F}_{\theta}$	64.5	76.1
	$\mathbf{DRF}_{\theta}$	**67.7 (+4.0)**	**77.5 (+5.8)**

DR-FIM outperforms both FIM and random selection strategies.

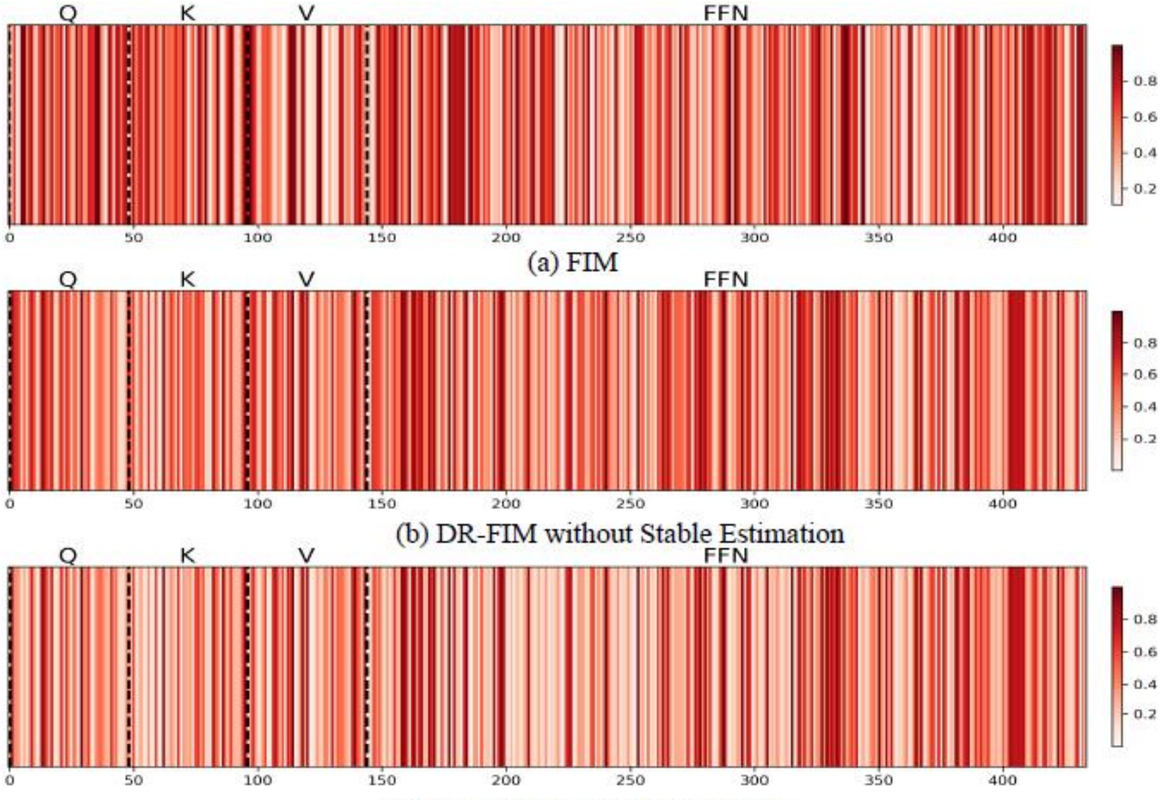
Different Estimation Methods



Stable estimation significantly improves the effectiveness of both FIM and DR-FIM

Experiments | Pre-training Transfer

Different Estimation Methods



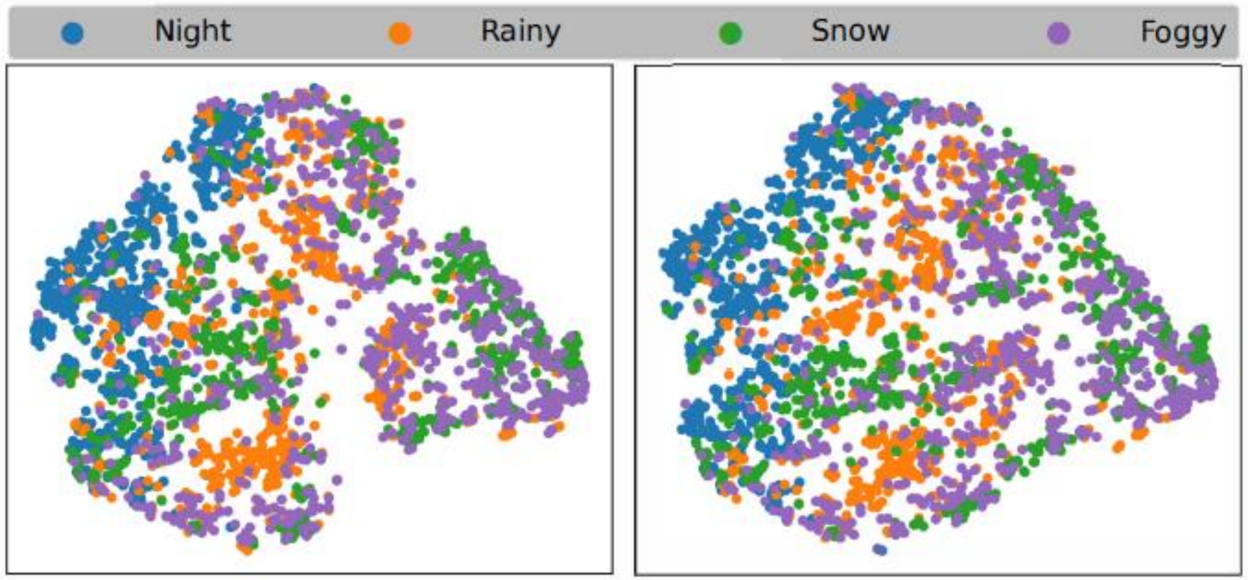
(a) FIM

(b) DR-FIM without Stable Estimation

(c) DR-FIM with Stable Estimation

Stable DR-FIM estimation more effectively highlights important parameters

Fine-tuning Transfer



FisherTune produces more balanced and domain-invariant feature distributions across unseen domains compared to Rein.

Thanks for Your Listen!