# SoMA: Singular Value Decomposed Minor Components Adaptation for Domain Generalizable Representation Learning

Seokju Yun Seunghye Chae Dongheon Lee Youngmin Ro*
Machine Intelligence Laboratory, University of Seoul, Korea

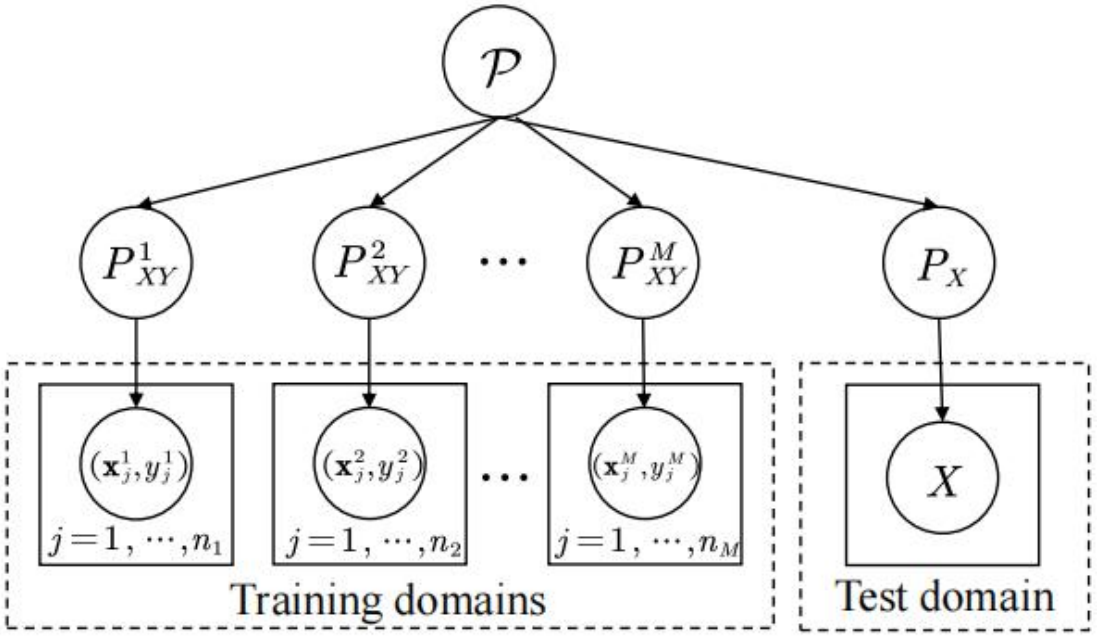
https://github.com/ysj9909/SoMA

CVPR 2025 Highlight

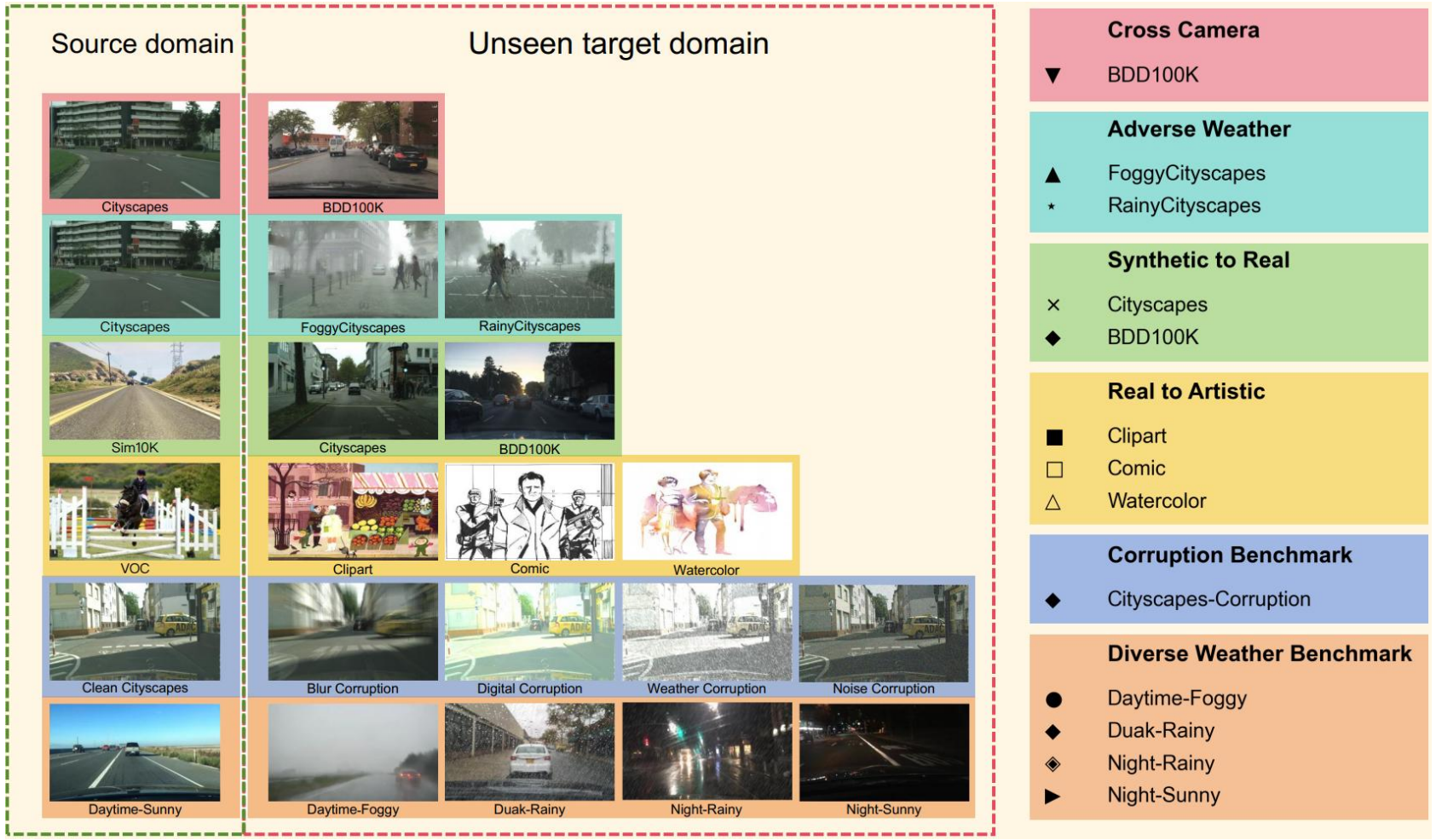
# Background - Domain Generalization

- Definition
  - Given: $M$ training domains $\mathcal{S} = \{\mathcal{S}_i | i = 1, \ldots, M\}$, where $\mathcal{S}_i = \{(x_j^i, y_j^i)\}_{j=1}^{n_i}$
  - Condition:
    - Joint distributions are different, i.e., $P_{XY}^i \neq P_{XY}^j$, $1 \leq i \neq j \leq M$
    - Test domain **cannot be accessed** in training
  - Goal:
    - Achieve minimum test error on test domain
    - $(P_{XY}^i \neq P_{XY}^{test})$

$$\min_h \mathbb{E}_{(x,y)\in\mathcal{S}_{test}}[l(h(x), y)]$$



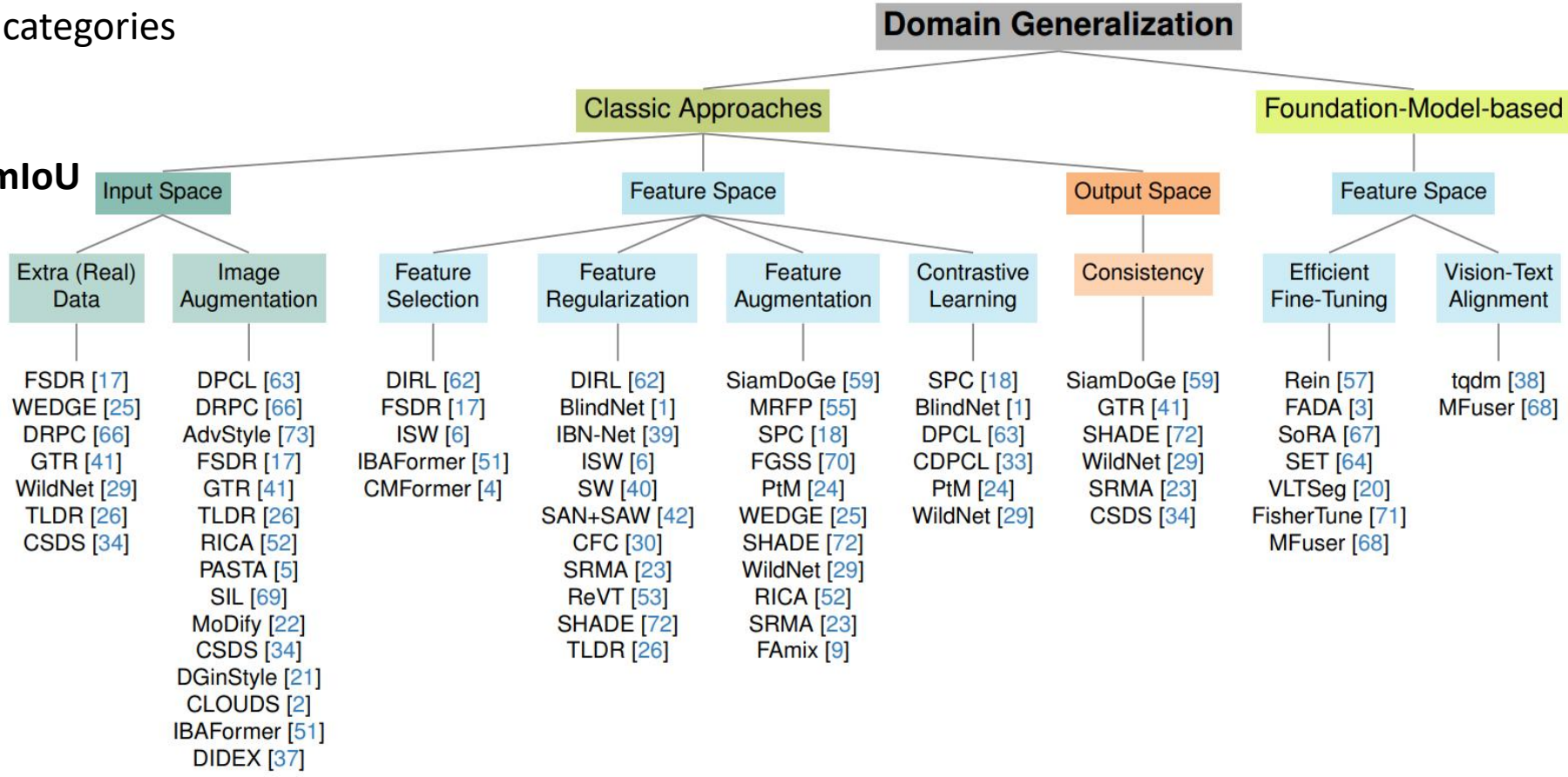
# Background - Dataset & Benchmark



# Background - Domain Generalization for Dense Predictions

- Settings
  - Input: a single image
  - Output:
    - Semantic Segmentation: pixel-level labels
    - Object Detection: bbox & categories
  - Evaluation Metrics:
    - Semantic Segmentation: **mIoU**
    - Object Detection: **mAP**

## Methodology

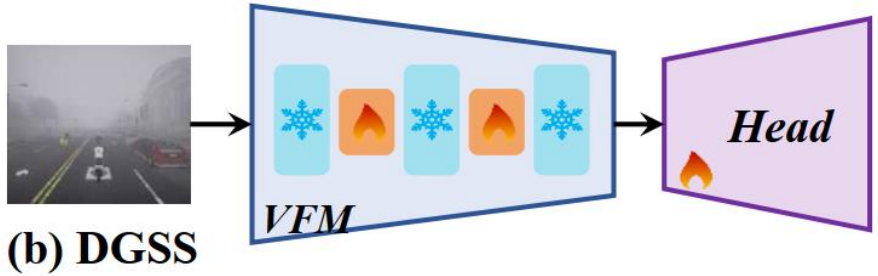


# Background - Vision foundation models (VFMs) for DG

- Motivation
  - Preserve the world knowledge of VFMs while effectively learning task-specific features.
- Architecture for DG
  - Frozen Backbone + PEFT modules + Task head
- Limitations
  - Generalized knowledge is easily disrupted during fine-tuning

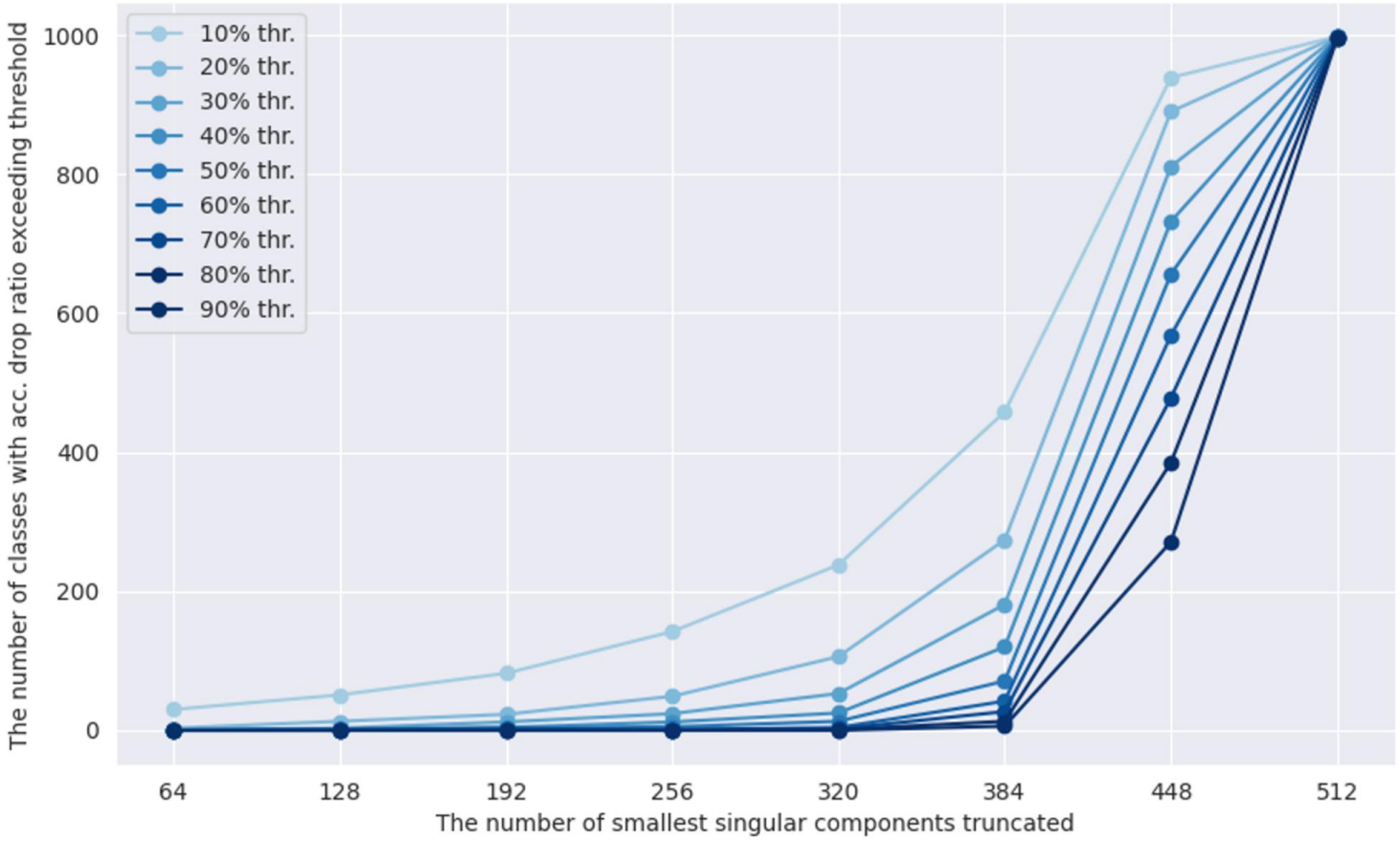

Figure 1: **Visualization of the first PCA components.** We compute a PCA between the patches of the images from the same column (a, b, c and d) and show their first 3 components. Each component is matched to a different color channel. Same parts are matched between related images despite changes of pose, style or even objects. Background is removed by thresholding the first PCA component.

**Architecture Paradigm**



# Preliminaries: SVD Analysis

- Testbed: ImageNet 1k Recognition




- Computes the SVD of pre-trained weights across all layers of the DINOv2 ViT-large model

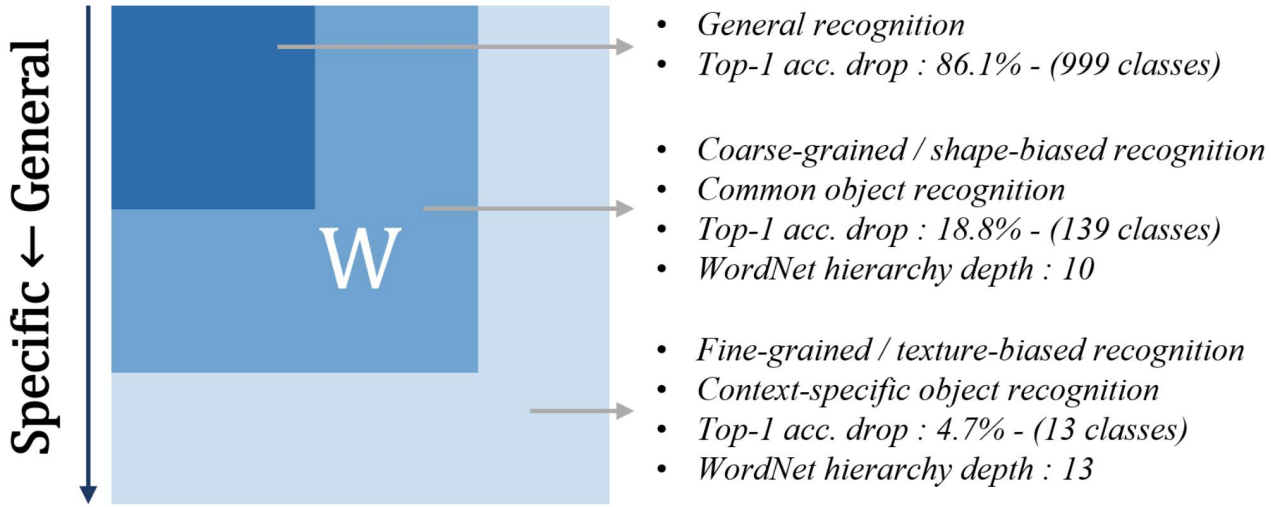
$$W = U \sum V^T = \sum_{i=1}^{R} \sigma_i u_i v_i^T$$

$$\text{in which } U = [u_1, u_2, \cdots, u_m] \in \mathbb{R}^{m \times m}$$
$$V = [v_1, v_2, \cdots, v_m] \in \mathbb{R}^{n \times n}$$

- **Truncation**: Set specific singular values and their corresponding singular vectors to 0
- 现象：随着截断的最小奇异分量的增加，误分类的类的数量呈指数增长
- 推论：具有较大奇异值的成分倾向于捕获跨越多个类的更一般的特征

# Preliminaries: SVD Analysis

- 不同层级的奇异成分的行为分析




Top 8 principle components: $U_{[:,:8]}\Sigma_{[:8]}(V^T)_{[:8,:]}$,

Middle 160 singular components: $U_{[:,384:544]}\Sigma_{[384:544]}(V^T)_{[384:544,:]}$,

Bottom 320 minor components: $U_{[:,-320:]}\Sigma_{[-320:]}(V^T)_{[-320:,:]}$

**核心假设**:

VFM 权重的高奇异值成分编码跨域泛化知识，低奇异值成分编码任务 / 领域特定知识

Top-component truncated model:

*functionality*

*co-occurrence*

Middle-component truncated model:

Bottom-component truncated model:

# Minor Singular Components Adaptation

**Summary**: 通过奇异值分解（SVD）拆分预训练权重，仅微调最小 r 个奇异值对应的次要奇异成分、冻结其余主成分，以保留VFMs的泛化能力并高效学习任务特定特征的参数高效微调（PEFT）方法

1. Perform SVD on the pre-trained weight matrices

$$W = U\sum V^T = \sum_{i=1}^{R}\sigma_i u_i v_i^T \quad \text{in which } U = [u_1, u_2, \cdots, u_m] \in \mathbb{R}^{m\times m}, V = [v_1, v_2, \cdots, v_m] \in \mathbb{R}^{n\times n}$$

2. Initialize the adapter with QR-type reconstruction of minor singular components

$$\text{The Minor Singular Components: } U_{[:,-r:]}\Sigma_{[-r:]}(V^T)_{[-r:,:]}$$

$$B = U_{[:,-r:]}\sqrt{\Sigma}_{[-r:]} \in \mathbb{R}^{m\times r}, A = \sqrt{\Sigma}_{[-r:]}(V^T)_{[-r:,:]} \in \mathbb{R}^{r\times n}$$

3. Update

$$y = (W \underbrace{-BA + B'A'}_{\Delta W_{SoMA}})x = (W_{res} + B'A')x = W'x.$$

# Discussion

$$y = (W \underbrace{-BA + B'A'}_{\Delta W_{SoMA}})x = (W_{res} + B'A')x = W'x.$$

1. **Aim**: Quantify the interference of learned $\Delta W$ on Pretrained Weight $W_0$
2. **Method**: Quantify the extent of interference of $\Delta W$ on $W_0$ through **singular modulation ratio** (SMR): The projection of $\Delta W$ on singular vector of the $W_0$
3. **Advantage**: SoMA effectively preserves the structure of the VFM's generalizable knowlege.

$$\mathrm{SMR}_i = \left| \frac{u_i^T \Delta W v_i}{\sigma_i} \right|$$

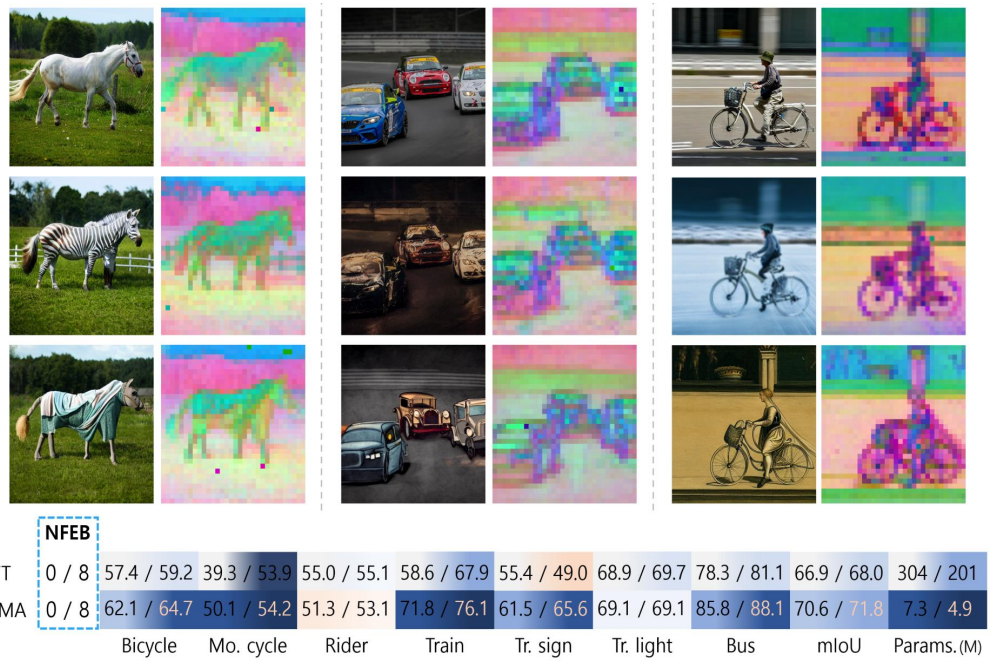
| SMR | Block | $i$ | $\Delta W_{SoMA}$ | $\Delta W^*_{SoMA}$ | $\Delta W_{LoRA}$ |
|---|---|---|---|---|---|
| $\lvert \frac{u_i^\top \Delta W v_i}{\sigma_i} \rvert$ | 12th | 0–255 | 0.075 | 0.084 (↓ 17%) | 0.101 |
| | | 256–511 | 0.094 | 0.094 (↓ 43%) | 0.166 |
| | | 512–767 | 0.157 | 0.165 (↓ 40%) | 0.275 |
| | | 768–1023 | 4.104 | 6.152 (↑ 21%) | 5.095 |
| | 24th | 0–255 | 0.097 | 0.095 (↓ 24%) | 0.125 |
| | | 256–511 | 0.105 | 0.108 (↓ 35%) | 0.166 |
| | | 512–767 | 0.206 | 0.189 (↓ 41%) | 0.323 |
| | | 768–1023 | 2.781 | 3.620 (↑ 26%) | 2.865 |

Table 1. **Singular modulation ratio comparison.** * indicates the adaptation matrix trained using the annealing weight decay strategy. The weight matrices are taken from the query projection layers of the 12th and 24th attention blocks in DINOv2-Large [58].

# Freezing Early Blocks

冻结视觉基础模型（VFMs）的浅层（Early Blocks）可保留领域不变语义特征、避免模型过拟合源域并减少可训练参数，其有效性经实验验证，且需控制冻结数量以平衡泛化性与判别性。

1. **核心假设**：VFMs Early Blocks（浅层layer）擅长提取定位精准的语义特征，**受风格或输入层面领域偏移影响小**；深层layer负责特征映射，需微调适配任务
2. **目的**：防止模型过拟合源域，保留预训练的泛化能力，同时减少可训练参数规模。
3. **原理性实验**：通过 PCA 分析证实Early Blocks特征的领域不变性，冻结后稀有类别 IoU 显著提升（如图）。
4. **超参数优化原则**：默认冻结生成最高分辨率特征图之前的块（如前 8 块），冻结过多会降低判别性。




Up: Visualization of PCA on 8th block's feature of the top 3 components
Bottom: Class-wise IoU comparison for rare classes

冻结ViT层数与性能的关系

| # frozen early blocks | 0 | 4 | 8 | 12 | 16 |
|---|---|---|---|---|---|
| *Citys.* perf. (mIoU in %) | 70.62 | 71.51 | **71.82** | 70.71 | 70.47 |
| Params.* | 7.3M | 6.1M | 4.9M | 3.7M | 2.4M |

Table 14. Performance comparison with **varying numbers of frozen early blocks** under *GTAV* → *Cityscapes* DGSS setting.

# Annealing Weight Decay

- **退火权重衰减**通过训练中逐步降低权重衰减系数，在不破坏视觉基础模型（VFM）预训练泛化能力的前提下提升判别性，是平衡泛化性与判别性的关键策略。
  - 核心假设：Weight Decay通过限制学习率防止训练早期模型泛化能力下降；常规Weight Decay使训练后期学习率过低导致判别性不足。
  - 核心策略：以较大权重衰减系数启动训练，后续按 cosine schedule等方式逐步降至零。
  - 作用机制：早期通过权重衰减抑制过拟合、保留预训练泛化性，后期降低正则化强度，让模型专注学习 task-specific 判别特征。
  - 实验效果：与 SoMA 其他组件协同，显著提升细粒度类别识别性能，且不增加额外计算开销。

| Methods | Params.* | DGSS Avg. | DGOD Avg. |
|---|---|---|---|
| Full fine-tuning (baseline) | 304.2M | 64.4 | 51.0 |
| └ + Freezing early blocks | 201.6M | 65.0 (↑ 0.6) | 51.4 (↑ 0.4) |
| └ + Tuning principal components | 4.9M | 66.1 (↑ 1.1) | 53.0 (↑ 1.6) |
| └ + Tuning minor components | 4.9M | 67.7 (↑ 2.7) | 53.8 (↑ 2.4) |
| └ + Annealing weight decay | 4.9M | 68.3 (↑ 0.6) | 54.3 (↑ 0.5) |

Table 8. **Effect of our changes** evaluated on DG benchmarks. See the full ablation study in the *Supplemental*.

# Experiment Results - Semantic Segmentation

| ***Synthetic-to-Real Generalization*** | | | Test Domains (mIoU in %) | | | |
|---|---|---|---|---|---|---|
| Methods | Backbone | Params.* | →*Citys.* | →*BDD* | →*Map.* | Avg. |
| *Single-source DGSS Trained on GTAV* | | | | | | |
| ○ CLOUDS [3] | CLIP-CN-L | 0.0M | 60.20 | 57.40 | 67.00 | 61.50 |
| ○ VLTSeg [36] | EVA02-L | 304.2M | 65.30 | 58.30 | 66.00 | 63.20 |
| ○ Rein [80] | EVA02-L | 3.0M | 65.30 | 60.50 | 64.90 | 63.60 |
| ○ FADA [4] | EVA02-L | 11.7M | 66.70 | 61.90 | 66.10 | 64.90 |
| ○ tqdm [59] | EVA02-L | 304.2M | 68.88 | 59.18 | 70.10 | 66.05 |
| ○ SoMA (Ours) | EVA02-L | 5.1M | 68.05 | 60.81 | 68.33 | 65.73 |
| ● SoMA (Ours) | EVA02-L | 5.1M | 69.94 | 62.48 | 68.33 | 66.92 |
| ○ DoRA [52] | DINOv2-L | 7.5M | 66.12 | 59.31 | 67.07 | 64.17 |
| ○ VPT [37] | DINOv2-L | 3.7M | 68.75 | 58.64 | 68.32 | 65.24 |
| ○ SET [86] | DINOv2-L | 6.1M | 68.06 | 61.64 | 67.68 | 65.79 |
| ○ FADA [4] | DINOv2-L | 11.7M | 68.23 | 61.94 | 68.09 | 66.09 |
| ○ AdaptFormer [12] | DINOv2-L | 6.3M | 70.10 | 59.81 | 68.77 | 66.23 |
| ○ SSF [51] | DINOv2-L | 0.5M | 68.97 | 61.30 | 68.77 | 66.35 |
| ○ LoRA [33] | DINOv2-L | 7.3M | 70.13 | 60.13 | 70.42 | 66.89 |
| ● Rein† [80] | DINOv2-L | 3.0M | 70.68 | 62.51 | 69.61 | 67.60 |
| ○ SoMA (Ours) | DINOv2-L | 4.9M | 71.82 | 61.31 | **71.67** | 68.27 |
| ● SoMA (Ours) | DINOv2-L | 4.9M | **73.63** | **63.33** | 70.98 | **69.31** |
| *Multi-source DGSS Trained on GTAV + SYNTHIA* | | | | | | |
| ○ Rein† [80] | DINOv2-L | 3.0M | 72.17 | 61.53 | 70.69 | 68.13 |
| ○ SoMA (Ours) | DINOv2-L | 4.9M | 73.16 | 61.90 | 72.73 | 69.26 |
| ● SoMA (Ours) | DINOv2-L | 4.9M | **74.85** | **63.59** | **73.92** | **70.79** |
| *Multi-source DGSS Trained on GTAV + SYNTHIA + UrbanSyn* | | | | | | |
| ○ FFT | DINOv2-L | 304.2M | 75.90 | 60.93 | 72.80 | 69.88 |
| ○ SoMA (Ours) | DINOv2-L | 4.9M | 77.33 | 62.78 | 74.93 | 71.68 |
| ○ FFT‡ | DINOv2-L | 307.3M | 77.06 | 61.81 | 75.09 | 71.32 |
| ○ Rein‡ [80] | DINOv2-L | 3.0M | 78.42 | 62.20 | 74.49 | 71.70 |
| ○ SoMA‡ (Ours) | DINOv2-L | 4.9M | 79.22 | 63.84 | **76.30** | 73.12 |
| ○ Freeze | DINOv2-G | 0.0M | 76.08 | 61.98 | 72.23 | 70.10 |
| ○ FFT | DINOv2-G | 1.1B | 76.90 | 61.69 | 73.53 | 70.71 |
| ○ SoMA (Ours) | DINOv2-G | 6.6M | 78.39 | 63.75 | 75.16 | 72.43 |
| ● SoMA (Ours) | DINOv2-G | 6.6M | **80.37** | **65.67** | 76.18 | **74.07** |

Table 2. Comparison of the proposed SoMA with existing DGSS ○ and PEFT ○ methods under various **synthetic-to-real settings**.

| ***Real-to-Real Generalization*** | | | Test Domains (mIoU in %) | | |
|---|---|---|---|---|---|
| Methods | Backbone | Params.* | →*BDD* | →*Map.* | Avg. |
| *Single-source DGSS Trained on Cityscapes* | | | | | |
| ○ HGFormer [19] | Swin-L | 196.0M | 61.50 | 72.10 | 66.80 |
| ○ CMFormer [5] | Swin-L | 196.0M | 62.60 | 73.60 | 68.10 |
| ○ tqdm [59] | EVA02-L | 304.2M | 64.72 | 76.15 | 70.44 |
| ○ FADA [4] | DINOv2-L | 11.7M | 65.12 | 75.86 | 70.49 |
| ○ Rein† [80] | DINOv2-L | 3.0M | 66.53 | 75.18 | 70.86 |
| ○ SoMA (Ours) | DINOv2-L | 4.9M | 67.02 | 76.45 | 71.74 |
| ● SoMA (Ours) | DINOv2-L | 4.9M | **68.08** | **77.87** | **72.98** |

Table 3. **Real-to-real DGSS comparison.**

# Experiment Results - Object Detection

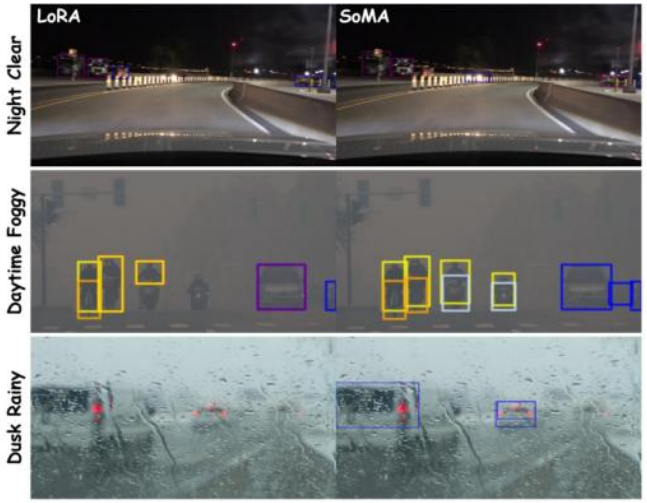



Figure 4. **DGOD qualitative results.**

| Methods | Params.* | DGSS Avg. | DGOD Avg. |
|---|---|---|---|
| Full fine-tuning (baseline) | 304.2M | 64.4 | 51.0 |
| └ + Freezing early blocks | 201.6M | 65.0 (↑ 0.6) | 51.4 (↑ 0.4) |
| └ + Tuning principal components | 4.9M | 66.1 (↑ 1.1) | 53.0 (↑ 1.6) |
| └ + Tuning minor components | 4.9M | 67.7 (↑ 2.7) | 53.8 (↑ 2.4) |
| └ + Annealing weight decay | 4.9M | 68.3 (↑ 0.6) | 54.3 (↑ 0.5) |

Table 7. **Effect of our changes** evaluated on DG benchmarks. See the full ablation study in the *Supplemental*.

| $\{W_q, W_k, W_v, W_o\}$ | $\{W_{up}, W_{down}\}$ | Params.* | →*Citys.* | →*BDD* | →*Map.* | Avg. |
|---|---|---|---|---|---|---|
| | ✔ | 2.7M | 68.76 | **61.50** | 70.00 | 66.75 |
| ✔ | | 2.2M | 70.82 | 61.36 | 69.86 | 67.35 |
| ✔ | ✔ | 4.9M | **71.82** | 61.31 | **71.67** | **68.27** |

Table 8. DGSS performance after applying SoMA to **different types of modules (Self-attention / MLP)** in DINOv2-Large.

| Rank $r$ | 4 | 8 | 16 | 32 | 64 |
|---|---|---|---|---|---|
| DGSS avg. (mIoU in %) | 66.91 | 67.71 | **68.27** | 67.76 | 67.59 |
| Params.* | 1.3M | 2.5M | 4.9M | 9.6M | 19.0M |

Table 9. DGSS performance with **different rank** $r$.

| ***Clear-to-Adverse Weather*** | | | S-DGOD [82] Test Domains (mAP@0.5 in %) | | | | |
|---|---|---|---|---|---|---|---|
| Methods | Params.* | DS | →*NC* | →*DR* | →*NR* | →*DF* | Avg. |
| *Single-source DGOD Trained on Daytime-Sunny (DS)* | | | | | | | |
| Backbone : ResNet101 [30] / Head : Faster R-CNN [66] | | | | | | | |
| S-DGOD [82] | 42.3M | 56.1 | 36.6 | 28.2 | 16.6 | 33.5 | 28.7 |
| CLIP-Gap [77] | 42.3M | 51.3 | 36.9 | 32.3 | 18.7 | 38.5 | 31.6 |
| OA-DG [47] | 42.3M | 55.8 | 38.0 | 33.9 | 16.8 | 38.3 | 31.8 |
| PDOC [50] | 42.3M | 53.6 | 38.5 | 33.7 | 19.2 | 39.1 | 32.6 |
| UFR [53] | 42.3M | 58.6 | 40.8 | 33.2 | 19.2 | **39.6** | 33.2 |
| DivAlign [17] | 42.3M | 52.8 | **42.5** | **38.1** | 24.1 | 37.2 | 35.5 |
| SoMA (Ours) | 3.1M | 49.3 | 41.9 | 37.9 | **24.5** | 38.2 | **35.6** |
| Backbone : DINOv2-L [58] / Head : Co-DETR [93] | | | | | | | |
| Freeze | 0.0M | 65.0 | 54.2 | 55.0 | 42.8 | 46.9 | 49.7 |
| FFT | 307.3M | 68.2 | 57.1 | 56.6 | 43.1 | 47.2 | 51.0 |
| DoRA [52] | 5.8M | 69.0 | 58.7 | 58.0 | 45.0 | 48.9 | 52.7 |
| AdaptFormer [12] | 6.3M | 68.9 | 58.8 | 58.3 | 44.4 | 49.8 | 52.8 |
| LoRA [33] | 5.5M | 69.6 | **59.6** | 58.1 | 46.1 | 49.5 | 53.3 |
| SoMA (Ours) | 4.9M | 69.4 | 59.3 | **59.3** | **47.6** | **51.0** | **54.3** |

Table 6. **Domain generalized object detection.**

# Conclusion