MERGETUNE: Continued Fine-tuning of Vision-Language Models

Wenqing Wang[1]*, Da Li[2,3]*, Xiatian Zhu[1]†, Josef Kittler[1]†

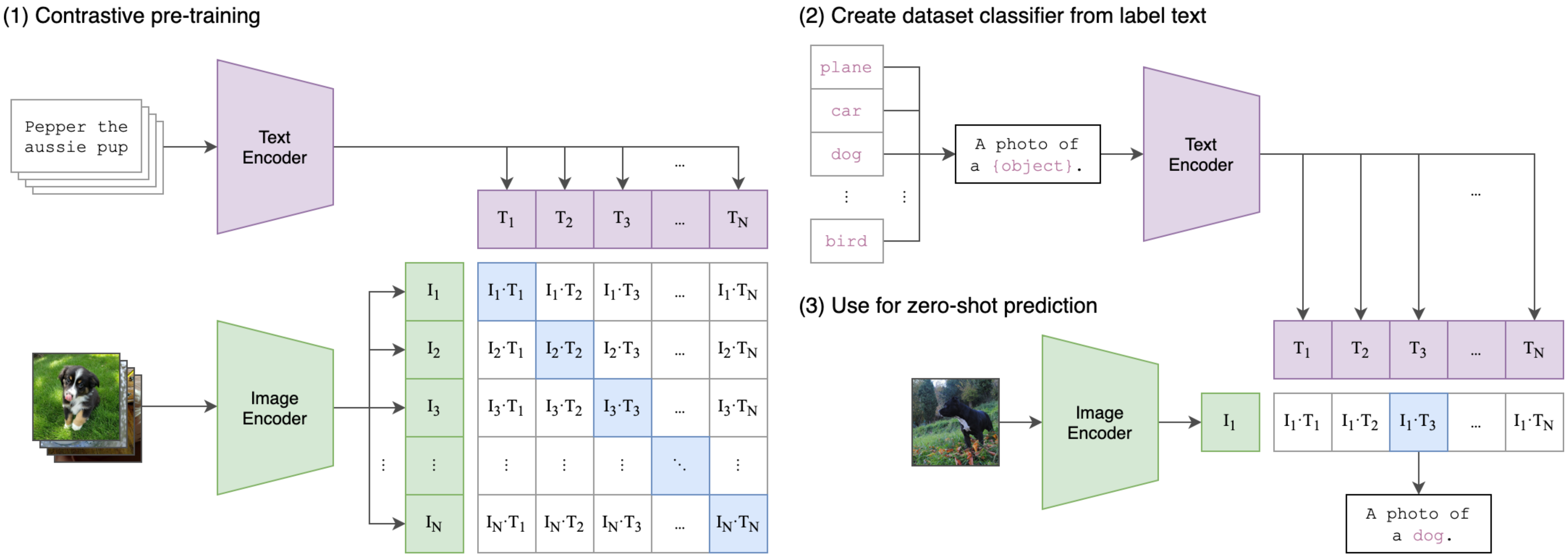
[1] University of Surrey
[2] Samsung AI Centre Cambridge
[3] Queen Mary University of London

*Equal contribution
†Joint last authorship

Contents

Background - Vision-Language models (VLMs)

➢ A vision–language model learn a joint representation of vision and language.



VLMs learn broad visual concepts from web-scale image–text pretraining and can do strong zero-shot transfer.

Background – VLMs Fine-tuning benefits

- Despite VLMs has strong zero-shot ability, still require fine-tuning for downstream tasks to get best performance.

Caltech101

Prompt	Accuracy
a [CLASS].	82.68
a photo of [CLASS].	80.81
a photo of a [CLASS].	86.29
$[V]_1 [V]_2 \ldots [V]_M$ [CLASS].	**91.83**

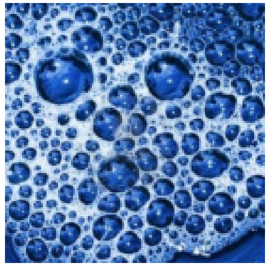
(a)

Flowers102

Prompt	Accuracy
a photo of a [CLASS].	60.86
a flower photo of a [CLASS].	65.81
a photo of a [CLASS], a type of flower.	66.14
$[V]_1 [V]_2 \ldots [V]_M$ [CLASS].	**94.51**

(b)

Describable Textures (DTD)



Prompt	Accuracy
a photo of a [CLASS].	39.83
a photo of a [CLASS] texture.	40.25
[CLASS] texture.	42.32
$[V]_1 [V]_2 \ldots [V]_M$ [CLASS].	**63.58**

(c)

EuroSAT

Prompt	Accuracy
a photo of a [CLASS].	24.17
a satellite photo of [CLASS].	37.46
a centered satellite photo of [CLASS].	37.56
$[V]_1 [V]_2 \ldots [V]_M$ [CLASS].	**83.53**

(d)

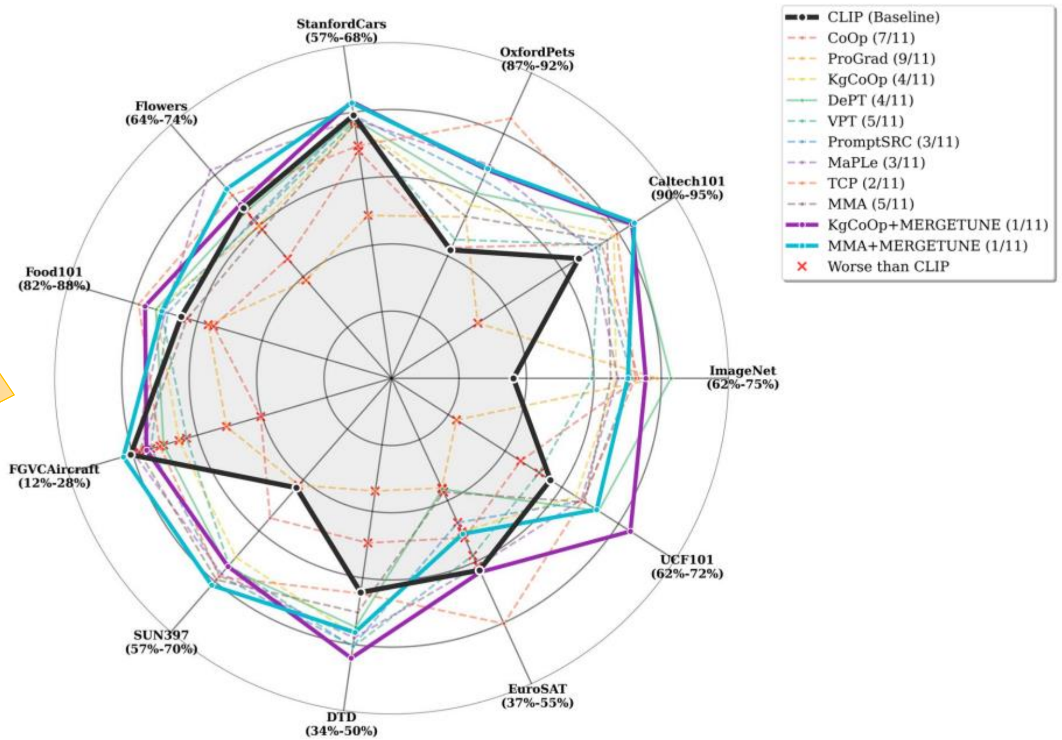
In practice, VLM fine-tuning often falls into two broad families: parameter-efficient tuning (PEFT, e.g., prompts/ adapters) and robust fine-tuning.

Background – VLMs fine-tuning methods and weaknesses

PEFT (prompts / adapters)
Training small modules only
Best for few-shot dataset
Pros efficient, strong in-domain gains

Robust fine-tuning (ensembling / soups / merging)
combine pretrained + tuned models
Best for many-shot / domain generalization
Pros better robustness & out-of-domain

However, both routes can still hurt zero-shot generalization (catastrophic forgetting)



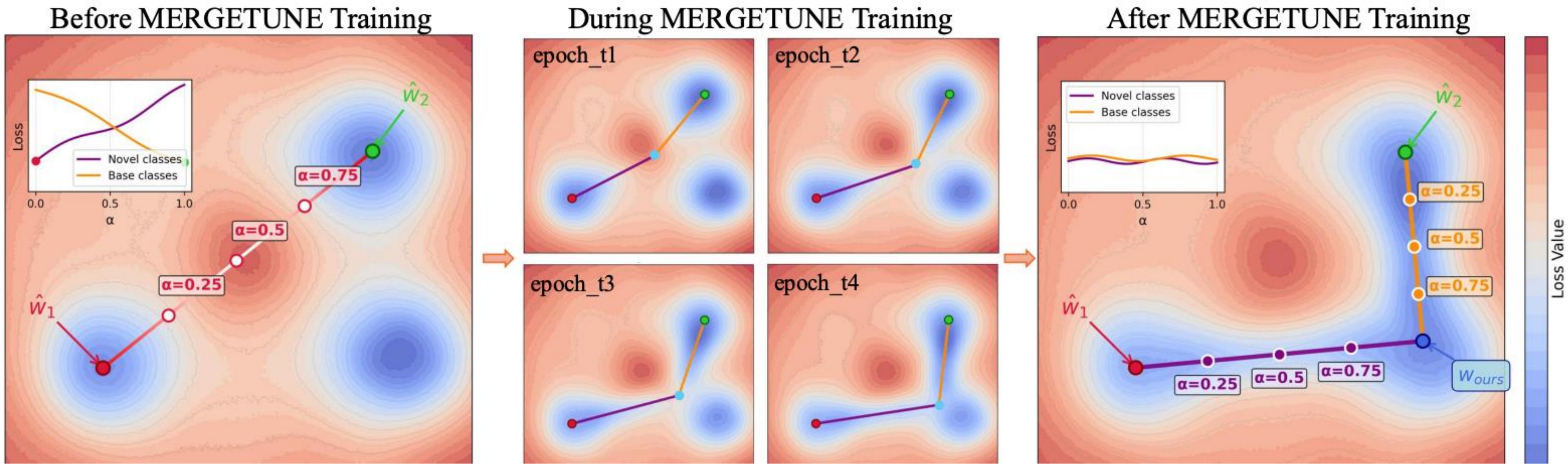
Method	Imagenet	Distribution shifts -V2	-S	-A	-R	ObjectNet	Avg-D
Zero-shot (CLIP)	68.34	61.90	48.27	50.12	77.60	54.23	58.42
Linear Probing (CVPR 22)	79.79	70.02	46.99	46.48	71.16	52.28	57.39
+ Weight ens. (CVPR 22)	79.80	70.45	48.41	47.89	73.00	53.07	58.56 (+1.17)
+ VRF (NeurIPS 24)	79.84	70.36	48.67	48.08	73.87	53.36	58.87 (+1.48)
+ TIES (NeurIPS 23)	79.75	70.33	48.25	48.32	73.78	53.13	58.76 (+1.37)
+ DARE (ICML 24)	79.14	70.26	48.14	47.74	73.11	53.01	58.45 (+1.06)
+ MERGETUNE	79.96	70.22	49.47	49.21	75.98	53.43	59.66 (+2.27)
+ Weight ens.	79.88	70.27	50.14	50.04	76.69	54.01	**60.23** (+2.84)
E2E-FT (CVPR 22)	81.31	70.61	45.12	36.62	65.63	50.51	53.70
+ Weight ens. (CVPR 22)	82.51	73.11	51.62	47.61	75.13	55.71	60.64 (+6.94)
+ VRF (NeurIPS 24)	82.32	72.12	52.93	48.41	78.72	56.41	61.72 (+8.02)
+ TIES (NeurIPS 23)	82.27	72.84	51.67	47.81	74.48	54.87	60.33 (+6.63)
+ DARE (ICML 24)	81.09	70.21	45.79	35.55	65.23	50.13	53.38 (-0.32)
+ MERGETUNE	82.26	72.98	52.76	51.61	78.01	56.22	62.29 (+8.59)
+ Weight ens.	82.18	73.21	53.10	52.68	78.68	56.84	**62.90** (+9.20)

Contents

ICLR
International Conference On Learning Representations

Motivation

- **High-level goal:** forgetting often unavoidable, so we trying to restore forgotten knowledge after the model has been fine-tuned.
- **Straightforward approach**: reuse the pretrained model, for example, naïve merging (e.g., weight averaging / interpolation) between pretrained and fine-tuned models.
- **Challenge**: pretrained and fine-tuned models can be far apart in weight space.
- **Idea**: find a continued model has mode-connected that connect both pretrained endpoints and fine-tuned endpoints.



Contents

- Background
- Motivation
- **Methods**
- Experiments

Methods - Preliminaries

➢ Mode connectivity:

When two models are trained on the same task with the loss $\mathcal{L}(\cdot)$ but differ in initialisation or training trajectory, they converge to different solutions $\hat{w}_1, \hat{w}_2$ that both achieve low loss, can merge them through weight averaging:

$$w = \gamma(\alpha) = (1-\alpha)\hat{w}_1 + \alpha\hat{w}_2, \quad \alpha \in [0,1].$$

The linear interpolation can sometimes yield a model with performance comparable to its endpoints $\hat{w}_1, \hat{w}_2$:

$$\mathcal{L}(\gamma(\alpha)) \approx 0,$$

Empirical studies have shown that seemingly distinct optima discovered by independent training runs can be linked by continuous low-loss paths in parameter space. This indicates that neural network solutions are not totally isolated minima but could lie in connected valleys of the loss landscape. Mode connectivity thus provides a theoretical basis for why interpolation between models can preserve low loss.

➢ Mode connectivity of different tasks:

Model merging can be extended to integrate knowledge from different but related tasks, the goal is to find a merged solution w that preserves performance on both tasks. Given two models $\hat{w}_1$ and $\hat{w}_2$ rained with task losses $\mathcal{L}_1$ and $\mathcal{L}_2$, one can seek a new model w whose interpolation paths $\hat{w}_1$ and $\hat{w}_2$

$$\gamma_1(\alpha) = \hat{w}_1 + \alpha(w - \hat{w}_1), \quad \gamma_2(\alpha) = \hat{w}_2 + \alpha(w - \hat{w}_2), \quad \alpha \in [0,1],$$

satisfy

$$\mathcal{L}_1(\gamma_1(\alpha)) \approx 0, \quad \mathcal{L}_2(\gamma_2(\alpha)) \approx 0.$$

Throughout the interpolation, model w maintains two smooth, low-loss connections to both endpoints, ensuring that knowledge from both tasks is preserved in the single w

Methods - Linear mode connectivity as an objective.

Given a zero-shot checkpoint $\hat{w}_1$ (e.g., CLIP) and a downstream fine-tuned checkpoint $\hat{w}_2$ (e.g., CoOp), we seek a continued solution w that integrates both. We require w to remain linearly connected to both $\hat{w}_1$ and $\hat{w}_2$ via low-loss interpolations. This leads to the objective:

$$w = \arg\min_{w} \mathbb{E}_{\alpha\sim\mathcal{U}[0,1]}\Big[\mathcal{L}_1(\hat{w}_1 + \alpha(w - \hat{w}_1)) + \mathcal{L}_2(\hat{w}_2 + \alpha(w - \hat{w}_2))\Big],$$

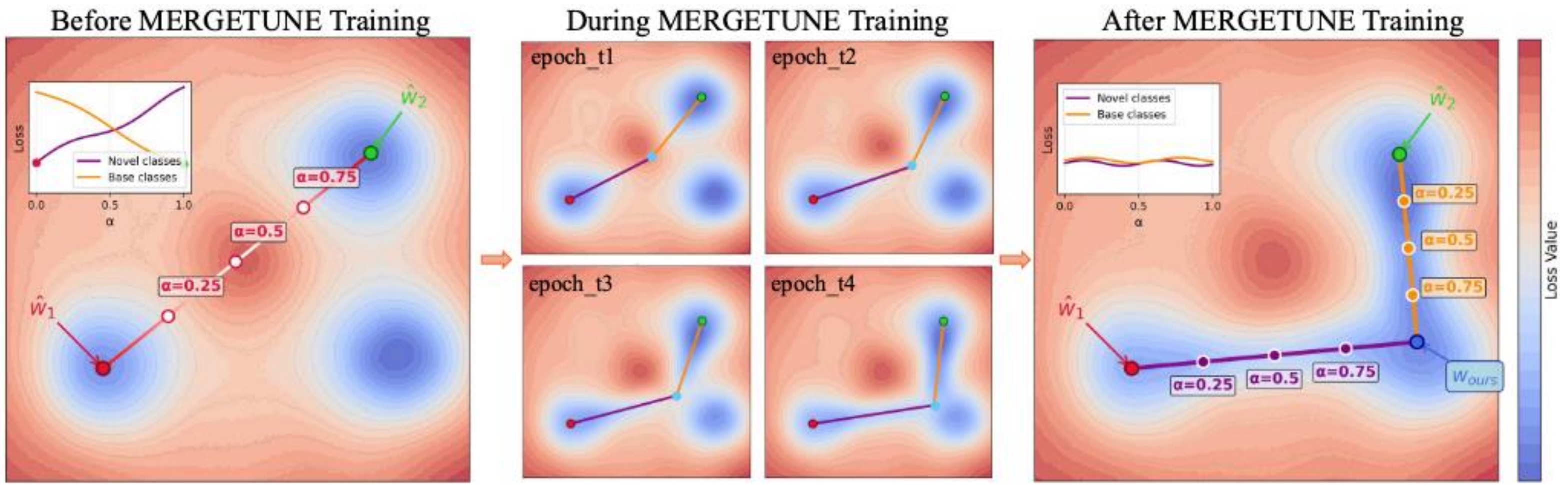
where $\mathcal{L}_1$ and $\mathcal{L}_2$ are the pretraining and downstream training losses for the zero-shot and finetuned models, and α is the interpolation coefficient uniformly sampled from [0, 1].

Our hypothesis is that if a model can be linearly connected to another solution through a consistently low-loss path, it can inherit and preserve the knowledge of that solution.

Challenge:

The direct implement of linear mode connectivity requires the computation of both $\mathcal{L}_1$ and $\mathcal{L}_2$.

However, $\mathcal{L}_1$ depends on the pretraining data (e.g., CLIP's web-scale corpus), which is often inaccessible and computationally prohibitive to replay.



Methods - Second-order surrogate loss

Propose a second-order surrogate loss to approximate the computation of Task 1 loss term in a replay-free manner. Specifically, approximate the Task 1 interpolation term using a second-order Taylor expansion:

$$\mathcal{L}_1(\hat{w}_1 + \alpha(w - \hat{w}_1)) \approx \mathcal{L}_1(\hat{w}_1) + \alpha \nabla \mathcal{L}_1(\hat{w}_1)^\top (w - \hat{w}_1) + \frac{\alpha^2}{2}(w - \hat{w}_1)^\top H_1 (w - \hat{w}_1),$$

We adopt two assumptions: $\nabla \mathcal{L}_1(\hat{w}_1) \approx 0$, as $\hat{w}_1$ lies near a local optimum for Task 1.

$H_1 \approx \mu I$ assuming isotropic curvature for tractability following common practices.

simplifies to $$\mathcal{L}_1(\hat{w}_1 + \alpha(w - \hat{w}_1)) \approx \mathcal{L}_1(\hat{w}_1) + \frac{\mu\alpha^2}{2}\|w - \hat{w}_1\|^2.$$

$\mathcal{L}_1(\hat{w}_1)$ is a constant value, $\frac{\mu\alpha^2}{2}$ can be represented by λ. The surrogate regulairiser can be formulated as:

$$\mathcal{R}_{\text{Task1}} = \lambda \|w - \hat{w}_1\|^2$$

Final replay-free objective

$$\mathcal{L}(w) = \mathcal{L}_2(w) + \lambda \|w - \hat{w}_1\|^2 + \beta\, \mathbb{E}_{\alpha \sim \mathcal{U}[0,1)} \mathcal{L}_2(\hat{w}_2 + \alpha(w - \hat{w}_2))$$

Methods – Training strategy

Our framework makes no limitation on which parameters are trainable, allowing it to be applied broadly.
We optimise w under the same training configurations as the base method

- ✓ In **prompt-based** methods such as CoOp and KgCoOp, $w = T(p)$ corresponds to the classifier weights derived from learnable prompts p, and T (·) is the text encoder.
- ✓ In **adapter-based** methods such as MMA, $w = T(p;\theta)$ includes fixed prompts and trainable multimodal adapters θ.
- ✓ In **many-shot regimes**, w can represent the linear classification head (linear probing) or the entire model parameters (end-to-end fine-tuning).

This flexibility allows our continued fine-tuning approach to act as a general post hoc enhancement, seamlessly integrating with diverse CLIP adaptation strategies.

Contents

- Background
- Motivation
- Methods
- **Experiments -> benenfit of, no more detail**

Experiments

Table 3: Domain generalisation results on ImageNet and four distribution shifts. Avg-D = average over domain-shifted datasets. MMA* is our reproduction.

Method	ImageNet	-V2	-S	-A	-R	Avg-D
CoOp (IJCV 22)	71.51	64.20	47.99	49.71	75.21	59.28
+TIES (NeurIPS 23)	62.84	56.26	41.14	44.65	70.77	53.20 (-6.08)
+DARE (ICML 24)	69.02	61.75	45.87	49.10	73.85	57.64 (-1.64)
+MergeTune	71.68	64.56	48.67	50.74	76.61	60.15 (+0.87)
KgCoOp (CVPR 22)	70.66	64.10	48.97	50.69	76.70	60.11
+TIES (NeurIPS 23)	67.77	61.47	46.36	49.86	75.91	58.40 (-1.71)
+DARE (ICML 24)	69.67	62.89	47.69	50.33	76.15	59.27 (-0.84)
+MergeTune	71.80	64.70	49.10	51.01	77.02	60.46 (+0.35)
MMA* (CVPR 24)	70.45	63.87	48.84	49.91	77.29	59.98
+TIES (NeurIPS 23)	64.01	57.13	42.32	45.23	72.11	54.20 (-5.78)
+DARE (ICML 24)	68.35	58.67	44.75	47.26	74.26	56.24 (-3.74)
+MergeTune	71.11	64.41	49.26	50.43	77.51	60.40 (+0.42)
PromptKD† (ICLR 24)	77.12	69.77	58.72	70.36	87.01	71.47
+TIES (NeurIPS 23)	74.63	64.22	54.85	67.47	83.85	67.60 (-3.87)
+DARE (ICML 24)	76.14	68.94	56.89	69.36	86.67	70.47 (-1.00)

Table 2: Cross-dataset gen
datasets. Avg-C = average

Method	Caltech101
CLIP (ICML 21)	93.30
CoOp (IJCV 22)	93.70
+TIES (NeurIPS 23)	93.31
+DARE (ICML 24)	89.67
+MergeTune	93.96
KgCoOp (CVPR 22)	93.92
+TIES (NeurIPS 23)	91.94
+DARE (ICML 24)	93.66
+MergeTune	94.24
MMA* (CVPR 24)	93.80

tly evaluated on other
xternal knowledge.

AT	UCF101	Avg-C
30	67.60	65.24
39	66.55	63.88
22	65.24	63.80 (-0.08)
65	64.68	61.67 (-2.21)
12	68.71	65.80 (+1.92)
04	68.50	65.51
36	65.74	63.55 (-1.96)
38	66.06	64.55 (-0.96)
41	70.42	66.53 (+1.02)
70	68.82	65.41
69	64.13	61.13 (-4.28)
89	65.35	62.28 (-3.13)
16	69.23	66.30 (+0.89)
74	76.39	71.33
68	72.54	67.86 (-3.47)
03	75.19	70.46 (-0.87)
94	77.12	71.68 (+0.35)

Table 4: ID-OOD generaisation accuracy of various methods on ImageNet and distribution shif
ViT-B/16 in the robust fine-tuning evaluation. Avg-D = average over domain-shifted datasets.

Method	Imagenet	Distribution shifts -V2	-S	-A	-R	ObjectNet	Avg-D
Zero-shot (CLIP)	68.34	61.90	48.27	50.12	77.60	54.23	58.42
Linear Probing (CVPR 22)	79.79	70.02	46.99	46.48	71.16	52.28	57.39
+ Weight ens. (CVPR 22)	79.80	70.45	48.41	47.89	73.00	53.07	58.56 (+1.17)
+ VRF (NeurIPS 24)	79.84	70.36	48.67	48.08	73.87	53.36	58.87 (+1.48)
+ TIES (NeurIPS 23)	79.75	70.33	48.25	48.32	73.78	53.13	58.76 (+1.37)
+ DARE (ICML 24)	79.14	70.26	48.14	47.74	73.11	53.01	58.45 (+1.06)
+ MergeTune	79.96	70.22	49.47	49.21	75.98	53.43	59.66 (+2.27)
+ Weight ens.	79.88	70.27	50.14	50.04	76.69	54.01	**60.23** (+2.84)
E2E-FT (CVPR 22)	81.31	70.61	45.12	36.62	65.63	50.51	53.70
+ Weight ens. (CVPR 22)	82.51	73.11	51.62	47.61	75.13	55.71	60.64 (+6.94)
+ VRF (NeurIPS 24)	82.32	72.12	52.93	48.41	78.72	56.41	61.72 (+8.02)
+ TIES (NeurIPS 23)	82.27	72.84	51.67	47.81	74.48	54.87	60.33 (+6.63)
+ DARE (ICML 24)	81.09	70.21	45.79	35.55	65.23	50.13	53.38 (-0.32)
+ MergeTune	82.26	72.98	52.76	51.61	78.01	56.22	62.29 (+8.59)
+ Weight ens.	82.18	73.21	53.10	52.68	78.68	56.84	**62.90** (+9.20)

Table 1: Base-to-novel generalisation experiments on 11 datasets. Our method achieves consistent average performance improvement over different baselines. †: Using large language model or teacher model's knowledge.

Method	Average Base	Novel	HM	ImageNet Base	Novel	HM	Caltech101 Base	Novel	HM	OxfordPets Base	Novel	HM
CLIP (ICML 21)	69.34	74.22	71.70	72.43	68.14	70.22	96.84	94.00	95.40	91.17	97.26	94.12
CoOp (IJCV 22)	82.69	63.22	71.66	76.47	67.88	71.92	98.00	89.81	93.73	93.67	95.29	94.47
KgCoOp (CVPR 22)	80.73	73.61	77.01	75.83	69.96	72.78	97.72	94.39	96.03	94.65	97.76	96.18
MaPLe (CVPR 23)	82.28	75.14	78.55	76.66	70.54	73.47	97.74	94.36	96.02	95.43	97.76	96.58
PromptSRC (ICCV 23)	84.26	76.10	79.97	77.60	70.73	74.01	98.10	94.03	96.02	95.33	97.30	96.30
MMA (CVPR 24)	83.20	76.80	79.87	77.31	71.00	74.02	98.40	94.00	96.15	95.40	98.07	96.72
CoPrompt† (ICLR 24)	84.00	77.23	80.48	77.67	71.27	74.33	98.27	94.90	96.55	95.67	98.10	96.87
PromptKD† (CVPR 24)	86.96	80.73	83.73	80.83	74.66	77.62	98.91	96.65	97.77	96.30	98.01	97.15
CoOp + ATPrompt† (ICCV 25)	82.68	68.04	74.65 (+2.99)	76.27	70.60	73.33	97.95	93.63	95.74	94.77	96.59	95.67
PromptKD + ATPrompt† (ICCV 25)	87.05	81.82	84.35 (+0.62)	80.90	74.83	77.75	98.90	96.52	97.70	96.92	98.27	97.59
CoOp + TIES (NeurIPS 23)	66.95	65.71	66.32 (-5.33)	70.09	61.65	65.60	93.63	91.34	92.50	90.12	95.13	92.56
CoOp + DARE (ICML 24)	75.91	65.97	70.59 (-1.07)	74.25	64.32	68.93	96.36	91.12	93.67	93.41	96.33	94.84
CoOp + MergeTune	80.82	73.97	77.24 (+5.58)	75.96	69.91	72.81	97.91	94.65	96.25	95.09	97.75	96.40
KgCoOp + TIES (NeurIPS 23)	73.03	72.09	72.56 (-4.45)	74.06	68.06	70.93	97.27	94.43	95.83	93.99	96.51	95.23
KgCoOp + DARE (ICML 24)	78.15	72.41	75.17 (-1.84)	75.10	68.73	71.78	97.53	94.47	95.97	92.22	97.37	94.73
KgCoOp + MergeTune	81.85	74.46	77.98 (+0.97)	76.49	69.35	72.75	97.87	94.91	96.37	95.32	97.76	96.53
MMA + TIES (NeurIPS 23)	70.39	68.46	69.41 (-10.46)	73.16	66.48	69.66	95.36	92.23	93.77	91.24	93.11	92.17
MMA + DARE (ICML 24)	74.25	69.52	71.81 (-8.06)	74.17	66.76	70.27	94.95	92.49	93.70	90.52	95.53	92.96
MMA + MergeTune	84.27	76.94	80.44 (+0.57)	77.68	70.65	74.00	98.32	94.65	96.45	95.75	98.10	96.91
PromptKD† + TIES (NeurIPS 23)	82.03	77.16	79.52 (-4.21)	76.12	69.32	72.56	97.61	94.9	96.24	95.23	96.13	95.68
PromptKD† + DARE (ICML 24)	84.76	79.65	82.13 (-1.6)	78.29	72.54	75.31	98.12	96.19	97.15	95.56	98.00	96.76
PromptKD† + MergeTune	87.23	81.17	84.09 (+0.36)	80.89	74.88	77.77	98.93	96.64	97.77	96.63	98.34	97.48

Method	StanfordCars Base	Novel	HM	Flowers102 Base	Novel	HM	Food101 Base	Novel	HM	FGVCAircraft Base	Novel	HM
CLIP (ICML 21)	63.37	74.89	68.65	72.08	77.80	74.83	90.10	91.22	90.66	27.19	36.29	31.09
CoOp (IJCV 22)	78.12	60.40	68.13	97.60	59.67	74.06	88.33	82.26	85.19	40.44	22.30	28.75
KgCoOp (CVPR 22)	71.76	75.04	73.36	95.00	74.73	83.65	90.50	91.70	91.10	36.21	33.55	34.83
MaPLe (CVPR 23)	72.94	74.00	73.47	95.92	72.46	82.56	90.71	92.05	91.38	37.44	35.61	36.50
PromptSRC (ICCV 23)	78.27	74.97	76.58	98.07	76.50	85.95	90.67	91.53	91.10	42.73	37.87	40.15
MMA (CVPR 24)	78.50	73.10	75.70	97.77	75.93	85.48	90.13	91.30	90.71	40.57	36.33	38.33
CoPrompt† (ICLR 24)	76.97	74.40	75.66	97.27	76.60	85.71	90.73	92.07	91.40	40.20	39.33	39.76
PromptKD† (CVPR 24)	82.80	83.37	83.13	99.42	82.62	90.24	92.43	93.68	93.05	49.12	41.81	45.17
CoOp + ATPrompt† (ICCV 25)	77.43	66.55	71.58	97.44	67.52	79.77	88.74	87.44	88.09	40.38	27.22	32.52
PromptKD + ATPrompt† (ICCV 25)	82.51	84.03	83.26	99.15	82.03	89.78	92.48	93.86	93.22	49.63	42.35	45.70
CoOp + TIES (NeurIPS 23)	57.94	63.42	60.56	73.60	65.41	69.27	88.42	87.64	88.02	30.29	28.83	29.54
CoOp + DARE (ICML 24)	67.82	61.62	64.57	90.88	63.73	74.92	87.78	87.22	87.50	34.41	28.25	31.03
CoOp + MergeTune	71.94	75.12	73.49	95.32	74.30	83.51	90.56	91.75	91.15	36.17	34.35	35.24
KgCoOp + TIES (NeurIPS 23)	65.79	74.17	69.73	74.26	72.53	73.39	90.25	91.58	90.91	30.81	28.97	29.86
KgCoOp + DARE (ICML 24)	67.84	73.02	70.33	87.02	72.88	79.33	90.44	91.60	91.02	33.49	30.37	31.85
KgCoOp + MergeTune	73.21	75.14	74.17	96.49	74.68	84.20	90.51	91.83	91.17	37.09	35.11	36.07
MMA + TIES (NeurIPS 23)	60.72	70.12	65.08	73.46	70.74	72.07	88.79	89.83	89.31	29.27	28.06	28.65
MMA + DARE (ICML 24)	64.82	71.52	68.01	77.15	70.95	73.92	89.85	87.61	88.72	30.35	28.23	29.25
MMA + MergeTune	81.24	72.30	76.51	98.04	76.33	85.83	90.77	91.57	91.17	42.86	36.01	39.14
PromptKD† + TIES (NeurIPS 23)	75.13	76.99	76.05	94.53	79.22	86.20	90.96	92.10	91.53	43.68	38.75	41.07
PromptKD† + DARE (ICML 24)	80.13	81.95	81.03	97.73	80.93	88.54	91.57	93.28	92.42	46.96	40.79	43.66
PromptKD† + MergeTune	82.98	83.90	83.44	99.30	82.67	90.23	92.45	93.91	93.17	49.82	42.31	45.76

Method	SUN397 Base	Novel	HM	DTD Base	Novel	HM	EuroSAT Base	Novel	HM	UCF101 Base	Novel	HM
CLIP (ICML 21)	69.36	75.35	72.23	53.24	59.90	56.37	56.48	64.05	60.03	70.53	77.50	73.85
CoOp (IJCV 22)	80.60	65.89	72.51	79.44	41.18	54.24	92.19	54.74	68.69	84.69	56.05	67.46
CoCoOp (CVPR 22)	79.74	76.86	78.27	77.01	56.00	64.85	87.49	60.04	71.21	82.33	73.45	77.64
KgCoOp (CVPR 22)	80.29	76.53	78.36	77.55	54.99	64.35	85.64	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MaPLe (CVPR 23)	80.82	78.70	79.75	80.36	59.18	68.16	94.07	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptSRC (ICCV 23)	82.67	78.47	80.52	83.37	62.97	71.75	92.90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MMA (CVPR 24)	82.27	78.57	80.38	83.20	65.63	73.38	85.46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CoPrompt† (ICLR 24)	82.63	80.03	81.30	83.13	64.73	72.79	94.60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptKD† (CVPR 24)	83.69	81.54	82.60	85.84	71.37	77.94	97.54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CoOp + ATPrompt† (ICCV 25)	80.84	68.64	74.24	80.83	45.49	58.22	90.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptKD + ATPrompt† (ICCV 25)	83.87	81.35	82.59	86.92	72.34	78.96	97.05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CoOp + TIES (NeurIPS 23)	62.41	62.99	62.70	57.33	41.87	48.40	50.17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CoOp + DARE (ICML 24)	70.33	63.01	66.47	72.49	47.02	57.04	69.71	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CoOp + MergeTune	80.37	76.19	78.22	78.51	56.84	65.93	85.15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
KgCoOp + TIES (NeurIPS 23)	73.05	73.88	73.46	60.77	53.54	56.93	68.66	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
KgCoOp + DARE (ICML 24)	77.47	74.81	76.12	72.46	54.19	62.01	84.48	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
KgCoOp + MergeTune	80.53	77.28	78.87	79.90	57.33	66.75	89.56	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MMA + TIES (NeurIPS 23)	69.46	65.77	67.56	55.38	48.36	51.63	65.69	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MMA + DARE (ICML 24)	71.43	72.61	72.02	67.65	50.37	57.74	77.42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MMA + MergeTune	82.46	78.48	80.42	83.37	65.34	73.26	90.09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptKD† + TIES (NeurIPS 23)	79.11	76.49	77.78	78.86	65.93	71.82	88.64	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptKD† + DARE (ICML 24)	80.34	79.66	80.00	84.22	70.04	76.48	92.48	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PromptKD† + MergeTune	83.92	81.49	82.69	86.77	72.10	78.76	97.81	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Few-shot setting →
Base-to-novel generalization
Cross-dataset generalization
Domain generalization

ID-OOD generalisation