Appendix

Proof of Proposition 1

Proof. By taking the expression of Eq. (4), Eq. (5) and Eq. (9) into the objective function (8), we can rewrite the objective function as

$$\min_{\mathbf{W}} -\frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\log p_{ik} + \sum_{k=1}^{K}\widehat{p}_k\log\widehat{p}_k$$
$$-\frac{\alpha-\beta}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}p_{ik}\log p_{ik} - \frac{\gamma}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}p_{ik}\log\hat{y}_{ik}.$$

$$(17)$$

This optimization problem can be reformulated by introducing auxiliary variables $\boldsymbol{q}=[q_{ik}]\in\mathbb{R}^{|\mathcal{Q}|\times K}$ and enforcing constraints $\boldsymbol{q}=\boldsymbol{p}$, alongside the pointwise simplex constraints:

$$\min_{\mathbf{W}} -\frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\log p_{ik} + \sum_{k=1}^{K}\widehat{q}_k\log\widehat{q}_k$$
$$-\frac{\alpha-\beta}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log p_{ik} - \frac{\gamma}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log\hat{y}_{ik},$$
$$\text{s.t.}\quad q_{ik}=p_{ik},\quad i\in\mathcal{Q},\quad k\in\{1,\cdots,K\}$$
$$\sum_{k=1}^{K}q_{ik}=1,\quad i\in\mathcal{Q}$$
$$q_{ik}\geq 0,\quad i\in\mathcal{Q},\quad k\in\{1,\cdots,K\}.$$

$$(18)$$

Notably, this constrained problem (18) can be resolved with penalty-based approach, by using the Kullback-Leibler(KL) divergence term to encourage equality constraints $\{\boldsymbol{q}_i=\boldsymbol{p}_i\mid i\in\mathcal{Q}\}$:

$$D_{\text{KL}}(\boldsymbol{q}_i\parallel\boldsymbol{p}_i)=\sum_{k=1}^{K}q_{ik}\log\frac{q_{ik}}{p_{ik}},\quad i\in\mathcal{Q},\qquad(19)$$

where $\boldsymbol{q}_i=[q_{i1},\cdots,q_{iK}]$ and $\boldsymbol{p}_i=[p_{i1},\cdots,p_{iK}]$ corresponds to the auxiliary pointwise prediction and our model's posterior prediction for sample $i\in\mathcal{Q}$, respectively. Thus, the constrained problem (18) becomes:

$$\min_{\mathbf{W},\,\boldsymbol{q}} -\frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\log p_{ik} + \sum_{k=1}^{K}\widehat{q}_k\log\widehat{q}_k$$
$$-\frac{\alpha-\beta}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log p_{ik} - \frac{\gamma}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log\hat{y}_{ik}$$
$$+\frac{1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log\frac{q_{ik}}{p_{ik}}$$
$$\text{s.t.}\quad \sum_{k=1}^{K}q_{ik}=1,\quad i\in\mathcal{Q}$$
$$q_{ik}\geq 0,\quad i\in\mathcal{Q},\quad k\in\{1,\cdots,K\}.$$

$$(20)$$

$$\square$$

Proof of Proposition 2

Proof. Plug the expression of p_{ik} in Eq. (2) into Eq. (11) and group terms together, we get:

$$(11)=\sum_{k=1}^{K}\widehat{q}_k\log\widehat{q}_k - \frac{\gamma}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log\hat{y}_{ik}$$
$$+\frac{1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log q_{ik} - \frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\log p_{ik}$$
$$-\frac{\alpha-\beta+1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log p_{ik}$$
$$=\sum_{k=1}^{K}\widehat{q}_k\log\widehat{q}_k - \frac{\gamma}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log\hat{y}_{ik}$$
$$+\frac{1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\log q_{ik}$$
$$+\frac{\lambda\tau}{2|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\parallel\boldsymbol{f}_i-\boldsymbol{w}_k\parallel^2$$
$$+\frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\log\Big(\sum_{j=1}^{K}\exp(-\frac{\tau}{2}\parallel\boldsymbol{f}_i-\boldsymbol{w}_j\parallel^2)\Big)$$
$$+\frac{(\alpha-\beta+1)\tau}{2|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\parallel\boldsymbol{f}_i-\boldsymbol{w}_k\parallel^2$$
$$+\frac{\alpha-\beta+1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\log\Big(\sum_{j=1}^{K}\exp(-\frac{\tau}{2}\parallel\boldsymbol{f}_i-\boldsymbol{w}_j\parallel^2)\Big).$$

$$(21)$$

We now solve the optimization problem using the Alternating Direction Method (ADM). Specifically, the problem is decomposed into two simpler subproblems: (1) updating the weights $\mathbf{W}$ while keeping the auxiliary variables $\boldsymbol{q}$ fixed, and (2) updating $\boldsymbol{q}$ while keeping $\mathbf{W}$ fixed.

- Update $\mathbf{W}$:
  Keep $\boldsymbol{q}$ fixed and omit the terms that do not involve $\mathbf{W}$, Eq. (21) can be simplified as $\mathcal{C}+\bar{\mathcal{C}}$, where

$$\mathcal{C}=\frac{\lambda\tau}{2|\mathcal{S}|}\sum_{i\in\mathcal{S}}\sum_{k=1}^{K}y_{ik}\parallel\boldsymbol{f}_i-\boldsymbol{w}_k\parallel^2$$
$$+\frac{(\alpha-\beta+1)\tau}{2|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\sum_{k=1}^{K}q_{ik}\parallel\boldsymbol{f}_i-\boldsymbol{w}_k\parallel^2$$

$$(22)$$

and

$$\bar{\mathcal{C}}=\frac{\lambda}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\log\Big(\sum_{j=1}^{K}\exp(-\frac{\tau}{2}\parallel\boldsymbol{f}_i-\boldsymbol{w}_j\parallel^2)\Big)$$
$$+\frac{\alpha-\beta+1}{|\mathcal{Q}|}\sum_{i\in\mathcal{Q}}\log\Big(\sum_{j=1}^{K}\exp(-\frac{\tau}{2}\parallel\boldsymbol{f}_i-\boldsymbol{w}_j\parallel^2)\Big).$$

$$(23)$$

Note that α, β, τ, and λ are all positive parameters. Under the assumption that $\alpha - \beta + 1 > 0$, the term $\mathcal{C}$ is convex, while $\bar{\mathcal{C}}$ remains non-convex. However, $\bar{\mathcal{C}}$ can be approximated by a linear term via first-order expansion at the current solution $\boldsymbol{w}_k^{(t)}$, as shown below:

$$
\begin{aligned}
\bar{\mathcal{C}}(\boldsymbol{w}_k) &\approx \bar{\mathcal{C}}(\boldsymbol{w}_k^{(t)}) + \frac{\partial \bar{\mathcal{C}}}{\partial \boldsymbol{w}_k}(\boldsymbol{w}_k^{(t)})^\top (\boldsymbol{w}_k - \boldsymbol{w}_k^{(t)}) \\
&= \frac{\lambda \tau}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} p_{ik}^{(t)} (\boldsymbol{f}_i - \boldsymbol{w}_k^{(t)})^\top \boldsymbol{w}_k \\
&\quad + \frac{(\alpha - \beta + 1)\tau}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} p_{ik}^{(t)} (\boldsymbol{f}_i - \boldsymbol{w}_k^{(t)})^\top \boldsymbol{w}_k + c,
\end{aligned}
$$
(24)

where $c = \bar{\mathcal{C}}(\boldsymbol{w}_k^{(t)}) + \frac{\partial \bar{\mathcal{C}}}{\partial \boldsymbol{w}_k}(\boldsymbol{w}_k^{(t)})^\top \boldsymbol{w}_k^{(t)}$ denotes a constant w.r.t $\boldsymbol{w}_k$. By replacing $\bar{\mathcal{C}}(\boldsymbol{w}_k)$ with its linear approximation, the objective $\mathcal{C} + \bar{\mathcal{C}}$ becomes strictly convex w.r.t $\boldsymbol{w}_k$. Therefore, the optimal $\boldsymbol{w}_k$ for the approximate objective can be obtained in closed form by setting the gradient to zero. Specifically,

$$
\begin{aligned}
&\frac{\mathcal{C} + \bar{\mathcal{C}}}{\partial \boldsymbol{w}_k} \\
&\approx \frac{\lambda \tau}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \left[p_{ik}^{(t)} (\boldsymbol{f}_i - \boldsymbol{w}_k^{(t)}) - y_{ik}(\boldsymbol{f}_i - \boldsymbol{w}_k) \right] \\
&\quad + \frac{(\alpha - \beta + 1)\tau}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \left[p_{ik}^{(t)} (\boldsymbol{f}_i - \boldsymbol{w}_k^{(t)}) - q_{ik}(\boldsymbol{f}_i - \boldsymbol{w}_k) \right].
\end{aligned}
$$
(25)

Setting this gradient to zero, we can get

$$
\boldsymbol{w}_k = \frac{S_1 + S_2}{\frac{\lambda \tau}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} y_{ik} + \frac{(\alpha - \beta + 1)\tau}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} q_{ik}},
$$
(26)

where

$$
\begin{cases}
S1 = \frac{\lambda \tau}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \left[y_{ik} \boldsymbol{f}_i + p_{ik}^{(t)} (\boldsymbol{w}_k^{(t)} - \boldsymbol{f}_i) \right] \\
S_2 = \frac{(\alpha - \beta + 1)\tau}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \left[q_{ik} \boldsymbol{f}_i + p_{ik}^{(t)} (\boldsymbol{w}_k^{(t)} - \boldsymbol{f}_i) \right].
\end{cases}
$$
(27)

- Update $\boldsymbol{q}$:
With $\mathbf{W}$ fixed, the objective in Eq. (11) is convex w.r.t the auxiliary variables $\boldsymbol{q}_i$, as it consists of a sum of linear and convex functions. Moreover, the simplex constraints on $\boldsymbol{q}_i$ are affine. Therefore, we minimize this constrained convex problem for each $\boldsymbol{q}_i$ by solving the Karush-Kuhn-Tucker (KKT) conditions. Let $\epsilon_i \in \mathbb{R}$ denote the Lagrange multiplier associated with the equality constraint $\sum_{k=1}^K q_{ik} = 1$. The KKT conditions are obtained by setting the gradient of the Lagrangian to zero while enforc-

ing the simplex constraints:

$$
\begin{aligned}
&\frac{\partial \left[\text{Eq.(11)} + \epsilon_i (\sum_{k=1}^K q_{ik} - 1) \right]}{\partial q_{ik}} \\
&= \frac{1}{|\mathcal{Q}|}(1 + \log \widehat{q}_k) - \frac{\gamma}{|\mathcal{Q}|} \log \hat{y}_{ik} + \frac{1}{|\mathcal{Q}|}(1 + \log q_{ik}) \\
&\quad - \frac{\alpha - \beta + 1}{|\mathcal{Q}|} \log p_{ik} + \epsilon_i \\
&= \frac{1}{|\mathcal{Q}|} \left[2 + \log \frac{\widehat{q}_k \cdot q_{ik}}{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}} \right] + \epsilon_i = 0.
\end{aligned}
$$
(28)

Solving Eq. (28) for q_{ik} yields:

$$
q_{ik} = \frac{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}}{\widehat{q}_k} \exp(-\epsilon_i |\mathcal{Q}| - 2).
$$
(29)

Applying the constraints $\sum_{k=1}^K q_{ik} = 1$ to Eq. (29), we can derive an explicit expression for the Lagrange multiplier ϵ_i:

$$
\exp(-\epsilon_i |\mathcal{Q}| - 2) = 1 \Big/ \Big(\sum_{j=1}^K \frac{\hat{y}_{ij}^\gamma \cdot p_{ij}^{\alpha-\beta+1}}{\widehat{q}_j} \Big).
$$
(30)

Plugging Eq. (30) in Eq. (29), we obtain the closed-form solution for q_{ik}:

$$
q_{ik} = \Big(\frac{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}}{\widehat{q}_k} \Big) \Big/ \Big(\sum_{j=1}^K \frac{\hat{y}_{ij}^\gamma \cdot p_{ij}^{\alpha-\beta+1}}{\widehat{q}_j} \Big).
$$
(31)

We now verify that the solution q_{ik} satisfies the positivity constraint $q_{ik} \geq 0$ required by the original problem. Recall the definition of $\widehat{q}_k$, we have

$$
\begin{aligned}
\widehat{q}_k &= \frac{1}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} q_{ik} = \frac{1}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \frac{\frac{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}}{\widehat{q}_k}}{\sum_{j=1}^K \frac{\hat{y}_{ij}^\gamma \cdot p_{ij}^{\alpha-\beta+1}}{\widehat{q}_j}} \\
&\propto \sum_{i \in \mathcal{Q}} \frac{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}}{\widehat{q}_k},
\end{aligned}
$$
(32)

this implies:

$$
\widehat{q}_k \propto \Big(\sum_{i \in \mathcal{Q}} \hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1} \Big)^{\frac{1}{2}}.
$$
(33)

Substituting back to Eq. (31), the expression for q_{ik} becomes:

$$
q_{ik} \propto \frac{\hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1}}{\Big(\sum_{i \in \mathcal{Q}} \hat{y}_{ik}^\gamma \cdot p_{ik}^{\alpha-\beta+1} \Big)^{\frac{1}{2}}}.
$$
(34)

Apparently, $q_{ik} \geq 0$. Hence the positivity constraint of the original problem is satisfied.

$\square$

Dataset	Prompt template
ImageNet	"a photo of a [ ]."
SUN397	"a photo of a [ ]."
Aircraft	"a photo of a [ ], a type of aircraft."
EuroSAT	"a centered satellite photo of [ ]."
Cars	"a photo of a [ ]."
Food101	"a photo of [ ], a type of food."
Pets	"a photo of [ ], a type of pet."
Flower102	"a photo of a [ ], a type of flower."
Caltech101	"a photo of a [ ]."
DTD	"[ ] texture."
UCF101	"a photo of a person doing [ ]."

Table 5: Prompt templates used for each dataset in the experiments.

Prompt Templates for Each Dataset

We conduct experiments on 11 public datasets. To ensure consistency, fixed prompt templates are used for each dataset throughout all experiments, as summarized in Table 5 below.

Hyperparameter Tuning Process and Final Settings

The objective function of TIM++ involves several hyperparameters, including the temperature parameter τ used to compute the posterior distribution p_{ik}, as well as a set of trade-off coefficients: λ, α, β, and γ. Following the standard TIM setup, we perform $T = 150$ optimization steps in all experiments.

To determine suitable hyperparameter values, we employ a grid-search strategy over the following candidate sets. Note that, as noted in Proposition 2, the optimization procedure depends only on the difference $\alpha - \beta$. Therefore, we directly tune this difference (rather than α and β individually) and ensure the constraint $\alpha - \beta + 1 > 0$ is satisfied throughout.

- Temperature τ: $[5, 10, 15, 30, 60, 100, 120, 150]$;

- CE Coefficient λ: $[0.025, 0.05, 0.1, 0.2, 0.4]$;

- Difference $\alpha - \beta$: $[-0.4, -0.2, -0.1, -0.05, -0.025, 0, 0.025, 0.05, 0.1, 0.2, 0.4]$;

- Relative weight γ for textual guidance: $[0, 0.025, 0.05, 0.1, 0.25, 0.5]$;

For each dataset, we compute validation accuracy across all hyperparameter combinations. Then, we average the Top-1 accuracy across the 11 public datasets to identify the configuration that yields the best overall validation performance. Through this tuning process, we arrive at the final configuration: $\tau = 120, \lambda = 0.4, \gamma = 0.05$, and $\alpha - \beta = 0.1$. This setting is consistently applied across all 11 datasets, all few-shot settings (1, 2, 4, 8, and 16) and all random seeds (1, 2, and 3). Notably, this fixed configuration demonstrates strong robustness and generalizability, demonstrating the effectiveness of TIM++ across diverse few-shot scenarios.

Experimental Results on 11 Public Datasets

We conduct experiments on 11 public datasets. Due to page constraints, the main paper reports only the average Top-1 accuracy across all datasets. Here, we provide the complete per-dataset results for each shot setting and each method, including our TIM++, in Table 8. In addition, Figure 4 complements Figure 2 by visualizing the performance on the remaining 9 datasets that were not shown individually in the main paper, under varying shot settings.

From Figure 4, it is evident that TIM++ consistently and significantly outperforms the standard TIM across all datasets, with particularly notable improvements in the 1-shot and 2-shot settings—where labeled data is extremely limited. These results underscore the superior ability of TIM++ to effectively leverage textual priors and transductive information under low-supervision conditions.

Furthermore, TIM++ achieves the overall best performance on 8 out of the 11 datasets. In particular, it consistently outperforms all competing state-of-the-art methods across all shot settings on five datasets: Aircraft, EuroSAT, UCF101, StanfordCars, and DTD. On ImageNet, SUN397, and OxfordPets, TIM++ achieves the best results in 3 or 4 out of the 5 shot settings.

In contrast, on Caltech101, OxfordFlowers, and Food101, methods such as TransCLIP and LP++ outperform TIM++ in certain settings. We attribute this performance gap to the use of a unified set of hyperparameters across all datasets: while this configuration yields strong overall performance, it may not be fully optimized for specific datasets like Caltech101, OxfordFlowers, and Food101.

Hyperparameter Analysis

To assess the impact of each hyperparameter, we conducted a comprehensive analysis across the 11 public datasets under various few-shot settings. Fig. 5 presents representative results for the 1-shot, 4-shot, and 16-shot cases. Overall, we observe that the model performance is relatively more sensitive to hyperparameter choices in the low-shot setting.

- **Impact of difference** $\alpha - \beta$**:** we vary $\alpha - \beta$ from -0.4 to 0.8. The overall validation accuracy increases as $\alpha - \beta$ grows, stabilizing once reaching around -0.05. Performance remains relatively consistent within the range $[-0.05, 0.8]$ for most datasets, with the exception of EuroSAT, DTD, Oxford Flowers, and FGVC-Aircraft, which show slightly higher sensitivity.

- **Impact of temperature** τ**:** we vary τ from 5 to 200. The overall accuracy increases steadily with larger τ, reaching a plateau around $\tau = 60$. Performance remains relatively stable in $[60, 200]$, again with EuroSAT, DTD and Oxford Flowers datasets showing slightly less stable behavior.

- **Impact of relative weight** γ**:** we vary γ from 0 to 0.5. The accuracy stays relative stable within the interval $[0.025, 0.1]$, after which it starts to decline as γ increases. Similar to the previous case, most datasets exhibits robut performance within this stable region,

Method	Source	Target										
	ImageNet	SUN397	Aircraft	EuroSAT	StanfordCars	Food101	Pets	Flower102	Caltech101	DTD	UCF101	Average
CoOp	71.74	60.61	17.98	40.68	61.33	83.04	87.55	60.63	89.80	38.28	61.91	60.18
+TransCLIP-ZS	73.67	66.31	19.30	52.89	66.70	85.63	89.16	68.17	92.09	41.75	68.34	65.03
+TIM++-ZS	72.21	66.28	18.04	55.84	68.70	86.45	91.81	68.51	83.66	38.71	69.13	64.71
CoCoOp	70.63	66.57	22.04	38.43	65.07	86.40	90.22	71.09	94.29	46.57	67.26	64.79
+TransCLIP-ZS	67.15	68.21	22.77	57.10	65.51	84.41	88.50	78.10	93.52	50.14	74.62	68.29
+TIM++-ZS	71.02	70.71	24.02	68.69	69.99	87.28	93.75	75.42	88.21	49.41	75.32	70.28
KgCoOp	71.08	66.11	22.61	45.94	65.43	86.33	89.83	69.94	93.70	45.07	67.88	65.29
+TransCLIP-ZS	73.07	70.72	24.19	61.45	69.51	87.18	90.99	76.26	94.31	49.43	73.48	69.75
+TIM++-ZS	71.30	70.41	23.22	66.56	71.22	87.31	93.65	73.39	84.45	48.46	74.69	69.34

Table 6: **Cross-Dataset transferability evaluation.** Prompt-tuning methods are trained on 16-shot ImageNet and evaluated on ten other fine-grained datasets. The reported average excludes ImageNet.

Method	Source	Target				
	ImageNet	ImageNet-Adversarial	ImageNet-V2	ImageNet-Rendition	ImageNet-Sketch	Average OOD
CoOp	71.74	49.12	64.06	75.26	46.51	58.74
+TransCLIP-ZS	73.67	50.95	64.87	76.23	49.77	60.45
+TIM++-ZS	72.21	46.49	63.95	72.42	51.20	58.52
CoCoOp	70.63	51.00	64.04	76.59	48.61	60.06
+TransCLIP-ZS	67.15	45.36	61.45	66.02	47.82	55.16
+TIM++-ZS	71.02	46.61	63.45	72.21	51.62	58.47
KgCoOp	71.08	50.45	64.16	76.65	48.55	59.95
+TransCLIP-ZS	73.07	52.33	64.57	77.34	51.39	61.41
+TIM++-ZS	71.30	46.02	63.94	72.17	51.83	58.49

Table 7: **Domain generalization evaluation.** Prompt-tuning methods are trained on 16-shot ImageNet and then evaluated on four domain-shifted datasets. Here, OOD refers to out-of-domain.

whereas `EuroSAT`, `DTD` and `Oxford Flowers` show slight fluctuations.

- **Impact of CE Coefficient** λ**:** we vary λ from 0.025 to 0.8. TIM++ maintains relative stable performance over this range for nearly all datasets, with `EuroSAT`, `DTD` and `FGVC-Aircraft` datasets showing slightly less stable.

Zero-shot Adaption Results

In this subsection, we further evaluate the zero-shot adaption performance of TIM++ against TransCLIP. Similar to TransCLIP, our TIM++ framework can be applied on top of existing prompt-tuning methods. To better assess its generalization ability, we study both dataset transferability and domain generalization by applying TIM++ on three representative prompt-tuning methods—CoOp(Zhou et al. 2022b), CoCoOp(Zhou et al. 2022a), and KgCoOp(Yao, Zhang, and Xu 2023)—each trained with 16-shot ImageNet supervision.

Dataset Transferability. As shown in Table 6, applying either TransCLIP or TIM++ consistently improves the cross-dataset performance for all the three prompt-tuning

methods. While TransCLIP yields clear gains over the prompt-tuning baselines, TIM++ achieves comparable or even higher transferability, particularly when combined with CoCoOp, where it achieves the highest average accuracy across the ten fine-grained datasets. These results demonstrate that our proposed TIM++ can effectively enhance generalization to unseen datasets, even without labeled support data, confirming TIM++ as a competitive alternative to TransCLIP for zero-shot transfer.

Domain Generalization. Table 7 summarizes the domain generalization results across four domain-shifted ImageNet variants, showing that the overall performance varies substantially depending on the severity of the domain shift. As shown in Fig. 3, among these datasets, ImageNet-Sketch and ImageNet-Adversarial exhibit larger distribution gap from the original ImageNet data. This explains why all methods, including TIM++ and TransCLIP, show noticeably lower performance on these two datasets.

Within these challenging settings, TIM++ consistently outperforms TransCLIP on ImageNet-Sketch, likely because the sketch domain contains simplified object structures with
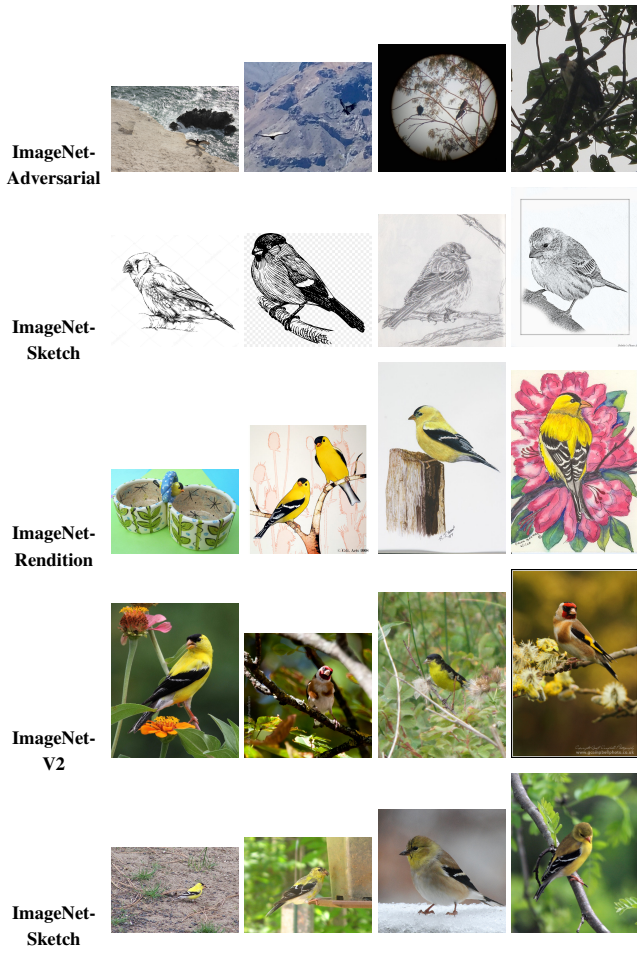
Figure 3: **Visualization of different ImageNet variants.**

minimal background information. Conversely, TransCLIP achieves higher accuracy on ImageNet-Adversarial, where images contain complex and cluttered backgrounds. Its underlying unlabeled clustering mechanism may provide more robust predictions in these difficult conditions compared to the entropy-minimization strategy of TIM++.

For ImageNet-V2 and ImageNet-Rendition, the domain shift is less pronounced and remains more visually consistent with ImageNet. Consequently, both methods perform substantially better on these datasets, and their results are more comparable. This pattern highlights that TIM++ and TransCLIP possess complementary strengths, with each method being better suited to particular types and degrees of domain shift.

Figure 4: **Top-1 Accuracy on 9 public datasets across varying shot settings.** Reported under 1, 2, 4, 8, and 16-shot settings, with results averaged over 3 random seeds (1, 2 and 3). These datasets complement those reported in the main paper and further compare the model's performance with other comparison methods under different shot settings.

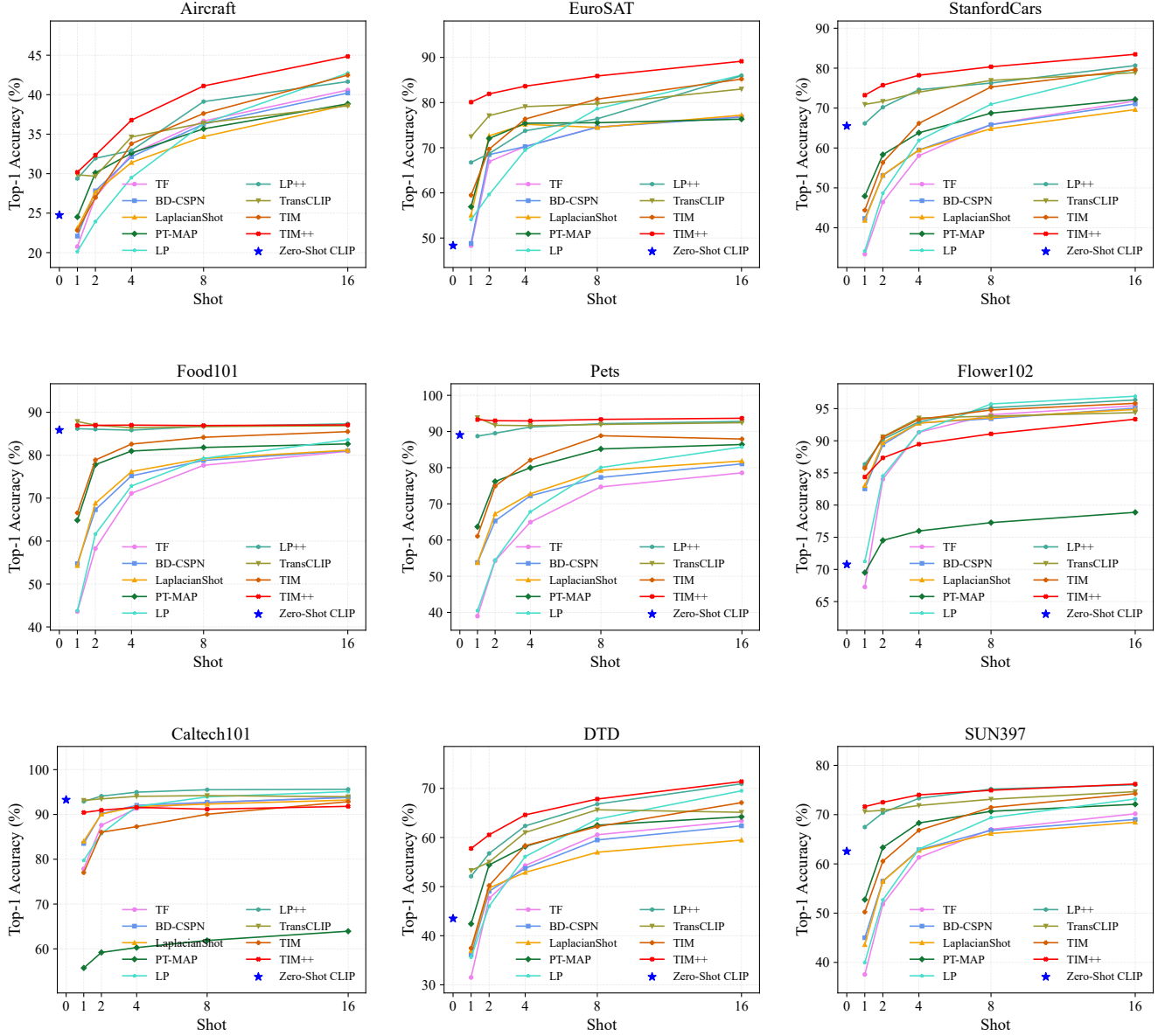| Dataset	Number of shots(S)	1	2	4	8	16
ImageNet	Zero-shot CLIP$_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	29.48 ± 0.54	40.62 ± 0.53	50.99 ± 0.48	57.34 ± 0.34	61.88 ± 0.15
	BD-CSPN $_{\text{ICCV'20}}$	35.72 ± 0.63	46.20 ± 0.32	53.70 ± 0.79	57.80 ± 0.87	61.82 ± 0.25
	LaplacianShot $_{\text{ICML'20}}$	34.73 ± 0.66	46.04 ± 0.27	53.54 ± 0.59	57.58 ± 0.43	60.84 ± 0.11
	PT-MAP $_{\text{ICANN'21}}$	40.35 ± 0.09	50.55 ± 0.75	57.46 ± 0.87	60.99 ± 0.27	63.86 ± 0.20
	LP $_{\text{ICML'21}}$	30.93 ± 0.67	43.18 ± 0.35	53.91 ± 0.45	61.80 ± 0.47	67.26 ± 0.15
	LP++ $_{\text{CVPR'24}}$	68.21 ± 0.22	68.90 ± 0.19	70.43 ± 0.11	71.72 ± 0.11	**72.60** ± 0.12
	TransCLIP $_{\text{NeurIPS'24}}$	**70.07** ± 0.12	70.38 ± 0.04	70.39 ± 0.24	70.26 ± 0.29	71.77 ± 0.15
	TIM $_{\text{NeurIPS'20}}$	38.05 ± 0.83	48.34 ± 0.81	57.42 ± 0.65	62.92 ± 0.54	67.51 ± 0.61
	TIM++	69.94 ± 0.09	**70.56** ± 0.04	**71.25** ± 0.09	**71.86** ± 0.45	72.59 ± 0.16
SUN397	Zero-shot CLIP$_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	37.56 ± 0.71	51.82 ± 0.49	61.31 ± 0.48	67.01 ± 0.30	70.18 ± 0.22
	BD-CSPN $_{\text{ICCV'20}}$	45.01 ± 0.31	56.50 ± 0.92	62.86 ± 0.77	66.80 ± 0.35	69.01 ± 0.53
	LaplacianShot $_{\text{ICML'20}}$	43.61 ± 0.23	56.45 ± 0.88	62.75 ± 0.56	66.17 ± 0.38	68.46 ± 0.08
	PT-MAP $_{\text{ICANN'21}}$	52.74 ± 0.37	63.33 ± 0.50	68.31 ± 0.38	70.67 ± 0.14	72.11 ± 0.12
	LP $_{\text{ICML'21}}$	40.00 ± 0.68	52.67 ± 0.09	63.00 ± 0.08	69.40 ± 0.10	73.18 ± 0.19
	LP++ $_{\text{CVPR'24}}$	67.47 ± 0.08	70.39 ± 0.11	73.31 ± 0.16	**75.19** ± 0.12	76.06 ± 0.06
	TransCLIP $_{\text{NeurIPS'24}}$	70.64 ± 0.16	70.87 ± 0.23	71.86 ± 0.27	73.13 ± 0.49	74.67 ± 0.09
	TIM $_{\text{NeurIPS'20}}$	50.24 ± 0.12	60.54 ± 0.72	66.82 ± 0.14	71.46 ± 0.42	74.25 ± 0.33
	TIM++	**71.64** ± 0.28	**72.51** ± 0.21	**74.01** ± 0.33	74.95 ± 0.13	**76.22** ± 0.22
Aircraft	Zero-shot CLIP$_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	20.76 ± 1.04	27.03 ± 1.94	32.49 ± 0.27	36.66 ± 1.64	40.58 ± 0.35
	BD-CSPN $_{\text{ICCV'20}}$	22.09 ± 1.50	27.83 ± 1.59	32.14 ± 0.78	36.27 ± 0.99	40.21 ± 0.93
	LaplacianShot $_{\text{ICML'20}}$	23.12 ± 1.16	27.64 ± 1.35	31.41 ± 0.77	34.68 ± 0.66	38.72 ± 1.26
	PT-MAP $_{\text{ICANN'21}}$	24.51 ± 1.08	30.10 ± 1.02	32.60 ± 0.99	35.67 ± 1.01	38.84 ± 0.23
	LP $_{\text{ICML'21}}$	20.13 ± 0.60	23.92 ± 0.76	29.50 ± 0.68	36.33 ± 1.01	42.73 ± 0.10
	LP++ $_{\text{CVPR'24}}$	29.37 ± 0.50	31.94 ± 0.81	32.92 ± 1.92	39.12 ± 0.29	41.65 ± 0.74
	TransCLIP $_{\text{NeurIPS'24}}$	29.85 ± 0.83	29.67 ± 1.44	34.64 ± 0.42	36.41 ± 1.51	38.59 ± 0.77
	TIM $_{\text{NeurIPS'20}}$	22.82 ± 1.13	26.98 ± 2.58	33.79 ± 1.05	37.61 ± 0.48	42.47 ± 0.88
	TIM++	**30.18** ± 0.12	**32.34** ± 0.44	**36.77** ± 1.18	**41.10** ± 0.99	**44.84** ± 0.40
EuroSAT	Zero-shot CLIP$_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	48.30 ± 7.37	66.92 ± 7.75	70.27 ± 2.79	74.58 ± 0.87	76.88 ± 1.85
	BD-CSPN $_{\text{ICCV'20}}$	48.81 ± 5.03	68.47 ± 8.06	70.25 ± 5.48	74.53 ± 0.61	76.79 ± 1.54
	LaplacianShot $_{\text{ICML'20}}$	55.07 ± 9.27	72.63 ± 3.71	75.19 ± 6.73	74.49 ± 1.90	77.21 ± 0.94
	PT-MAP $_{\text{ICANN'21}}$	56.92 ± 10.94	72.03 ± 5.51	75.42 ± 0.75	75.54 ± 0.53	76.32 ± 0.86
	LP $_{\text{ICML'21}}$	54.07 ± 2.85	59.62 ± 1.53	69.49 ± 5.58	78.65 ± 0.72	86.07 ± 0.34
	LP++ $_{\text{CVPR'24}}$	66.75 ± 1.97	68.69 ± 0.73	73.75 ± 1.91	76.40 ± 2.87	85.94 ± 0.23
	TransCLIP $_{\text{NeurIPS'24}}$	72.45 ± 7.31	77.10 ± 6.12	79.10 ± 0.49	79.71 ± 0.66	82.99 ± 1.18
	TIM $_{\text{NeurIPS'20}}$	59.49 ± 8.78	69.67 ± 4.85	76.35 ± 3.93	80.77 ± 0.59	85.20 ± 2.97
	TIM++	**80.09** ± 1.15	**81.92** ± 2.34	**83.63** ± 1.62	**85.88** ± 0.78	**89.15** ± 0.61

Continued on next page

Table 8 – *Continued from previous page*

Dataset	Number of shots(S)	1	2	4	8	16
	Zero-shot CLIP$_{ICML'21}$			65.25		
StanfordCars	TF $_{Arxiv'19}$	33.39 ± 1.61	46.45 ± 1.23	58.07 ± 0.39	65.88 ± 0.65	71.78 ± 0.20
	BD-CSPN $_{ICCV'20}$	42.33 ± 1.95	53.09 ± 1.39	59.50 ± 0.47	65.82 ± 0.86	71.04 ± 0.12
	LaplacianShot $_{ICML'20}$	41.86 ± 1.98	53.21 ± 1.49	59.49 ± 0.23	64.83 ± 0.67	69.61 ± 0.29
	PT-MAP $_{ICANN'21}$	47.93 ± 2.37	58.33 ± 1.43	63.80 ± 0.18	68.72 ± 0.90	72.18 ± 0.58
	LP $_{ICML'21}$	34.21 ± 0.09	48.65 ± 0.64	61.89 ± 0.47	70.96 ± 0.32	79.77 ± 0.12
	LP++ $_{CVPR'24}$	66.15 ± 0.96	70.20 ± 0.41	74.60 ± 0.48	76.27 ± 0.11	80.65 ± 0.19
	TransCLIP $_{NeurIPS'24}$	70.92 ± 0.56	71.67 ± 0.95	73.99 ± 0.54	76.93 ± 0.18	78.87 ± 1.60
	TIM $_{NeurIPS'20}$	44.41 ± 2.55	56.35 ± 0.62	66.16 ± 0.67	75.27 ± 0.32	79.63 ± 0.35
	TIM++	$\mathbf{73.23} \pm 1.09$	$\mathbf{75.73} \pm 0.66$	$\mathbf{78.22} \pm 0.67$	$\mathbf{80.35} \pm 0.91$	$\mathbf{83.47} \pm 0.33$
	Zero-shot CLIP$_{ICML'21}$			65.25		
Food101	TF $_{Arxiv'19}$	43.59 ± 1.66	58.28 ± 1.21	71.12 ± 1.11	77.65 ± 0.38	80.91 ± 0.38
	BD-CSPN $_{ICCV'20}$	54.73 ± 1.62	67.31 ± 0.38	75.19 ± 0.38	78.79 ± 0.19	81.06 ± 0.49
	LaplacianShot $_{ICML'20}$	54.29 ± 1.58	68.81 ± 0.67	76.19 ± 0.51	79.24 ± 0.54	81.16 ± 0.38
	PT-MAP $_{ICANN'21}$	64.86 ± 1.58	77.82 ± 0.95	80.97 ± 0.05	81.81 ± 0.30	82.64 ± 0.19
	LP $_{ICML'21}$	43.79 ± 0.55	61.63 ± 0.81	72.83 ± 0.54	79.26 ± 0.40	83.60 ± 0.18
	LP++ $_{CVPR'24}$	86.19 ± 0.06	86.05 ± 0.14	85.84 ± 0.67	86.72 ± 0.20	$\mathbf{87.30} \pm 0.03$
	TransCLIP $_{NeurIPS'24}$	$\mathbf{87.87} \pm 0.14$	$\mathbf{87.00} \pm 0.59$	86.41 ± 0.51	86.69 ± 0.06	86.87 ± 0.12
	TIM $_{NeurIPS'20}$	66.57 ± 2.15	78.89 ± 1.16	82.59 ± 1.02	84.18 ± 0.50	85.49 ± 0.21
	TIM++	86.94 ± 0.20	86.97 ± 0.14	$\mathbf{87.02} \pm 0.07$	$\mathbf{86.92} \pm 0.17$	87.06 ± 0.11
	Zero-shot CLIP$_{ICML'21}$			65.25		
Pets	TF $_{Arxiv'19}$	38.98 ± 2.43	54.28 ± 1.49	64.93 ± 0.65	74.68 ± 1.82	78.57 ± 0.86
	BD-CSPN $_{ICCV'20}$	53.79 ± 1.50	65.28 ± 1.96	72.23 ± 1.15	77.31 ± 2.96	81.04 ± 0.70
	LaplacianShot $_{ICML'20}$	53.77 ± 1.46	67.24 ± 2.46	72.79 ± 0.60	79.25 ± 1.20	81.84 ± 0.75
	PT-MAP $_{ICANN'21}$	63.67 ± 2.21	76.18 ± 3.38	79.98 ± 1.31	85.17 ± 0.33	86.40 ± 0.73
	LP $_{ICML'21}$	40.50 ± 4.48	54.44 ± 2.67	67.80 ± 3.50	80.04 ± 0.10	85.71 ± 0.71
	LP++ $_{CVPR'24}$	88.72 ± 1.29	89.52 ± 0.37	91.23 ± 0.59	92.16 ± 0.73	92.79 ± 0.04
	TransCLIP $_{NeurIPS'24}$	$\mathbf{93.81} \pm 0.70$	91.72 ± 0.93	91.61 ± 1.26	91.94 ± 0.37	92.43 ± 0.30
	TIM $_{NeurIPS'20}$	61.07 ± 4.11	74.91 ± 1.75	82.08 ± 0.19	88.83 ± 1.94	87.92 ± 2.47
	TIM++	93.27 ± 0.44	$\mathbf{92.98} \pm 0.16$	$\mathbf{92.92} \pm 0.21$	$\mathbf{93.35} \pm 0.38$	$\mathbf{93.66} \pm 0.29$
	Zero-shot CLIP$_{ICML'21}$			65.25		
Flower102	TF $_{Arxiv'19}$	67.28 ± 0.91	84.03 ± 2.85	91.32 ± 0.57	94.06 ± 0.78	95.45 ± 0.83
	BD-CSPN $_{ICCV'20}$	82.53 ± 2.03	89.39 ± 1.21	92.77 ± 0.72	93.44 ± 0.66	95.07 ± 0.48
	LaplacianShot $_{ICML'20}$	83.07 ± 2.22	89.71 ± 1.68	92.73 ± 0.71	93.63 ± 0.50	94.87 ± 0.61
	PT-MAP $_{ICANN'21}$	69.51 ± 1.86	74.52 ± 1.60	75.98 ± 0.41	77.28 ± 0.51	78.87 ± 0.24
	LP $_{ICML'21}$	71.25 ± 2.48	84.50 ± 1.29	91.35 ± 0.30	$\mathbf{95.72} \pm 0.14$	$\mathbf{96.93} \pm 0.53$
	LP++ $_{CVPR'24}$	$\mathbf{86.34} \pm 0.18$	90.24 ± 0.53	92.94 ± 0.23	95.14 ± 0.53	96.35 ± 0.16
	TransCLIP $_{NeurIPS'24}$	85.74 ± 1.15	$\mathbf{90.62} \pm 1.06$	$\mathbf{93.54} \pm 0.63$	93.83 ± 0.70	94.38 ± 0.24
	TIM $_{NeurIPS'20}$	85.75 ± 3.53	90.54 ± 1.44	93.31 ± 0.70	94.79 ± 0.29	95.82 ± 0.46
	TIM++	84.37 ± 0.65	87.35 ± 0.61	89.47 ± 0.92	91.07 ± 0.32	93.35 ± 0.15

Table 8 – *Continued from previous page*

Dataset	Number of shots(S)	1	2	4	8	16
Caltech101	Zero-shot CLIP $_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	77.86 ± 2.30	87.60 ± 0.39	91.32 ± 0.49	92.64 ± 0.17	93.83 ± 0.41
	BD-CSPN $_{\text{ICCV'20}}$	83.52 ± 1.66	90.14 ± 0.32	92.04 ± 0.73	92.70 ± 0.15	93.77 ± 0.38
	LaplacianShot $_{\text{ICML'20}}$	84.04 ± 2.16	90.07 ± 0.39	91.75 ± 0.75	92.31 ± 0.24	93.21 ± 0.30
	PT-MAP $_{\text{ICANN'21}}$	55.74 ± 1.59	59.20 ± 0.66	60.28 ± 0.53	61.88 ± 0.19	63.94 ± 0.39
	LP $_{\text{ICML'21}}$	79.76 ± 1.21	85.71 ± 0.28	91.76 ± 0.69	93.91 ± 0.23	95.09 ± 0.26
	LP++ $_{\text{CVPR'24}}$	92.90 ± 0.93	$\mathbf{94.09} \pm 0.78$	$\mathbf{94.97} \pm 0.29$	$\mathbf{95.51} \pm 0.34$	$\mathbf{95.61} \pm 0.11$
	TransCLIP $_{\text{NeurIPS'24}}$	$\mathbf{93.14} \pm 0.16$	93.47 ± 0.32	94.04 ± 0.95	94.20 ± 0.77	93.96 ± 0.42
	TIM $_{\text{NeurIPS'20}}$	77.03 ± 5.01	86.03 ± 1.75	87.28 ± 6.97	90.03 ± 1.86	92.85 ± 0.65
	TIM++	90.43 ± 0.08	90.95 ± 0.64	91.58 ± 0.76	91.17 ± 0.35	91.81 ± 0.81
DTD	Zero-shot CLIP $_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	31.50 ± 3.64	47.62 ± 0.29	54.29 ± 1.12	60.56 ± 0.69	63.38 ± 0.68
	BD-CSPN $_{\text{ICCV'20}}$	36.09 ± 1.78	49.05 ± 1.38	53.72 ± 1.09	59.50 ± 1.28	62.37 ± 1.21
	LaplacianShot $_{\text{ICML'20}}$	36.94 ± 2.58	49.70 ± 0.88	52.90 ± 1.13	56.99 ± 0.92	59.48 ± 0.42
	PT-MAP $_{\text{ICANN'21}}$	42.42 ± 1.37	54.41 ± 1.03	58.18 ± 0.29	62.51 ± 0.12	64.22 ± 0.75
	LP $_{\text{ICML'21}}$	35.64 ± 0.28	46.00 ± 1.24	56.11 ± 1.26	63.75 ± 0.88	69.50 ± 1.41
	LP++ $_{\text{CVPR'24}}$	52.09 ± 0.88	56.76 ± 1.50	62.35 ± 1.40	66.80 ± 0.19	70.92 ± 0.52
	TransCLIP $_{\text{NeurIPS'24}}$	53.29 ± 1.78	54.98 ± 2.37	61.01 ± 1.08	65.64 ± 1.50	65.13 ± 1.78
	TIM $_{\text{NeurIPS'20}}$	37.47 ± 0.70	50.20 ± 3.64	58.33 ± 1.85	62.21 ± 5.72	67.10 ± 3.56
	TIM++	$\mathbf{57.76} \pm 2.04$	$\mathbf{60.54} \pm 1.38$	$\mathbf{64.60} \pm 1.23$	$\mathbf{67.83} \pm 0.78$	$\mathbf{71.39} \pm 0.31$
UCF101	Zero-shot CLIP $_{\text{ICML'21}}$			65.25		
	TF $_{\text{Arxiv'19}}$	49.81 ± 1.23	62.23 ± 0.36	71.28 ± 0.92	76.61 ± 0.63	77.66 ± 0.35
	BD-CSPN $_{\text{ICCV'20}}$	56.63 ± 2.15	67.21 ± 1.09	72.74 ± 1.12	75.95 ± 0.88	77.34 ± 0.56
	LaplacianShot $_{\text{ICML'20}}$	56.30 ± 1.02	67.36 ± 0.27	73.24 ± 1.07	76.32 ± 0.34	76.93 ± 0.22
	PT-MAP $_{\text{ICANN'21}}$	64.56 ± 1.08	69.87 ± 1.17	74.40 ± 0.36	75.40 ± 0.46	76.49 ± 0.62
	LP $_{\text{ICML'21}}$	48.57 ± 0.71	61.91 ± 1.10	70.30 ± 0.71	77.13 ± 0.53	80.70 ± 0.69
	LP++ $_{\text{CVPR'24}}$	71.46 ± 0.39	75.23 ± 0.26	79.05 ± 0.12	81.81 ± 0.23	83.72 ± 0.02
	TransCLIP $_{\text{NeurIPS'24}}$	78.40 ± 0.34	78.46 ± 0.96	79.09 ± 1.01	81.46 ± 0.34	82.08 ± 0.28
	TIM $_{\text{NeurIPS'20}}$	62.90 ± 2.19	70.52 ± 1.06	77.07 ± 1.71	78.83 ± 2.05	81.44 ± 0.90
	TIM++	$\mathbf{79.65} \pm 0.62$	$\mathbf{80.80} \pm 0.44$	$\mathbf{82.54} \pm 0.86$	$\mathbf{84.69} \pm 0.06$	$\mathbf{85.75} \pm 0.39$

Table 8: **Top-1 accuracy across different shot settings for each dataset.** Reported are the average classification accuracy (%) and standard deviation over three random seeds (1, 2, and 3). The best performance among all methods is highlighted in **bold**.
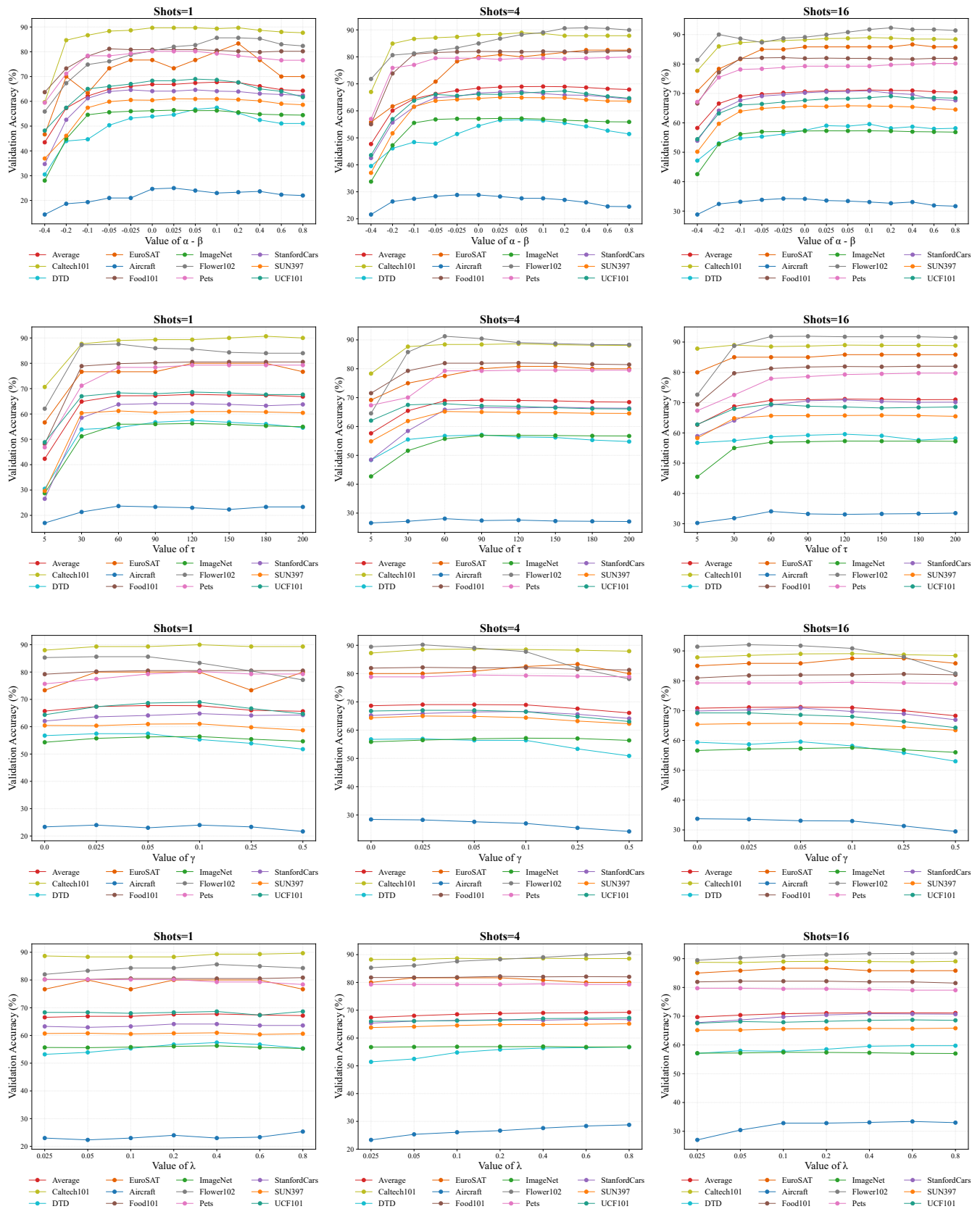
Figure 5: **Hyperparameter sensitivity analysis.** Validation performance of TIM++ across 11 public datasets under 1-shot and 16-shot settings.