From Few-way to Many-way: Rethinking Few-shot Fine-grained Image Classification

Supplementary Material

6. Proof of Theorem 3.2

In this section, we provide the supplementary proof for the analysis in Sec. 3.2.

Proof. We first analyze the generalization gap on the training classes. For two source classes i and j with class-conditional distributions $\tilde{P}_i$ and $\tilde{P}_j$, let $\tilde{S}_i \sim \tilde{P}_i^{m_i}$ and $\tilde{S}_j \sim \tilde{P}_j^{m_j}$ be i.i.d. sample sets. According to Proposition 1 of [4], for any $\delta_0 \in (0,1)$, with probability at least $1-\delta_0$,

$$V_f(\tilde{P}_i,\tilde{P}_j) \le \Big(V_f(\tilde{S}_i,\tilde{S}_j) + \frac{1}{2\|\mu_f(\tilde{S}_i)-\mu_f(\tilde{S}_j)\|^2} \cdot \sum_{c\in\{i,j\}} \big(\varepsilon_c^{(2)}(\tfrac{\delta_0}{4}) + 2\|\mu_f(\tilde{P}_c)\|\varepsilon_c^{(1)}(\tfrac{\delta_0}{4}) + (\varepsilon_c^{(1)}(\tfrac{\delta_0}{4}))^2\big)\Big)\Big(1+\frac{\varepsilon_i^{(1)}(\frac{\delta_0}{4})+\varepsilon_j^{(1)}(\frac{\delta_0}{4})}{\|\mu_f(\tilde{P}_i)-\mu_f(\tilde{P}_j)\|}\Big)^2. \tag{24}$$

Next, we relate the training class distributions to novel classes. Let $\mathcal{F}^* \subset \mathcal{F}$ and define the minimum class separation

$$\Delta(\mathcal{F}^*) = \inf_{f\in\mathcal{F}^*} \inf_{c\neq c'} \|\mu_f(P_c)-\mu_f(P_{c'})\| > 0.$$

Using Proposition 2 of [4] and Corollary 3 of [15], for any $\delta_1 \in (0,1)$, with probability at least $1-\delta_1$,

$$\mathbb{E}_{P_c\neq P_{c'}}[V_f(P_c,P_{c'})] \le \frac{2}{l(l-1)}\sum_{i\neq j} V_f(\tilde{P}_i,\tilde{P}_j) + C_1(\mathcal{F}^*)\,\frac{\sqrt{2\pi\log l}\,\mathbb{E}[\mathcal{R}(H_{\mathcal{F}^*})]}{(l-1)\,\Delta(\mathcal{F}^*)^2} + C_2(\mathcal{F}^*)\,\frac{2\sqrt{\log(1/\delta_1)}\cdot\sup_{f,P'}\mathrm{Var}_f(P')}{\sqrt{l}\,\Delta(\mathcal{F}^*)^2}. \tag{25}$$

Finally, we combine Eq. (24) and Eq. (25) via the union bound (Lemma 6.1). Let $\delta > 0$, set $\delta_0 = \delta/2$ and $\delta_1 = \delta/2$. Then, with probability at least $1-\delta$,

$$\mathbb{E}_{P_c\neq P_{c_0}}[V_f(P_c,P_{c_0})] \le \frac{2}{l(l-1)}\sum_{i\neq j}\Big(\Big(V_f(\tilde{S}_i,\tilde{S}_j) + \frac{1}{2\|\mu_f(\tilde{S}_i)-\mu_f(\tilde{S}_j)\|^2}\cdot\sum_{c\in\{i,j\}}\big(\varepsilon_c^{(2)}(\tfrac{\delta}{8}) + 2\|\mu_f(\tilde{P}_c)\|\varepsilon_c^{(1)}(\tfrac{\delta}{8}) + (\varepsilon_c^{(1)}(\tfrac{\delta}{8}))^2\big)\Big)\cdot\Big(1+\frac{\varepsilon_i^{(1)}(\frac{\delta}{8})+\varepsilon_j^{(1)}(\frac{\delta}{8})}{\|\mu_f(\tilde{P}_i)-\mu_f(\tilde{P}_j)\|}\Big)^2\Big) + C_1(\mathcal{F}^*)\,\frac{\sqrt{2\pi\log l}\,\mathbb{E}[\mathcal{R}(H_{\mathcal{F}^*})]}{(l-1)\,\Delta(\mathcal{F}^*)^2} + C_2(\mathcal{F}^*)\,\frac{2\sqrt{\log(2/\delta)}\cdot\sup_{f,P'}\mathrm{Var}_f(P')}{\sqrt{l}\,\Delta(\mathcal{F}^*)^2}. \tag{26}$$

In the following, we present the statement of Lemma 6.1 along with its proof.

Lemma 6.1 *Let $\{E_i\}_{i=1}^n$ be a set of events such that $\mathbb{P}(E_i) \ge 1-\delta_i$, for some $\delta_i \ge 0$, $i = 1,\dots,n$. Then,*

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) \ge 1-\sum_{i=1}^n \delta_i.$$

Proof. We start by expressing the probability of all events occurring simultaneously in terms of their complements:

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) = 1-\mathbb{P}\left(\bigcup_{i=1}^n E_i^c\right), \tag{27}$$

where E_i^c denotes the complement of event E_i.

Next, we make use of the standard union bound

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i^c\right) \le \sum_{i=1}^n \mathbb{P}(E_i^c). \tag{28}$$

By the given condition $\mathbb{P}(E_i) \ge 1-\delta_i$, we have

$$\sum_{i=1}^n (1-\mathbb{P}(E_i)) \le \sum_{i=1}^n \delta_i. \tag{29}$$

Putting everything together, we conclude

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) \ge 1-\sum_{i=1}^n \delta_i. \tag{30}$$

7. Detailed Experimental Setups

7.1. Datasets

We follow the commonly adopted standard dataset splitting protocol to divide the classes into training, validation, and test subsets.

For CUB-200-2011 [19], the dataset contains 11,788 bird images spanning 200 species, which are divided into 100 training, 50 validation, and 50 test categories.

For Stanford Dogs [6], the dataset consists of 20,580 labeled images covering 120 dog breeds worldwide, divided into 70 training, 20 validation, and 30 test categories.

For Stanford Cars [7], it includes 16,185 images across 196 car models, partitioned into 130 training, 17 validation, and 49 test categories.

For Flowers102 [16], the dataset comprises 8,189 images representing 102 flower species, with splits of 51 categories for training, 26 for validation, and 25 for testing.

7.2. Experimental details

Our method is conducted with PyTorch library on a single NVIDIA 3090. We adopt the same data augmentation strategies as prior works, including random cropping, random horizontal flipping, and color jittering. For all benchmark datasets, the model is trained using SGD with a momentum of 0.9. The learning rate is set to 0.1, and the weight decay is set to 5e-4. For the Stanford Cars and Flowers102 datasets, which exhibit more structured and distinctive visual patterns, the collaborative enhancement between lower-level features (i.e., $\hat{\mathbf{F}}^{s,t}$) is not applied. The model achieving the best validation accuracy is retained for testing, and all results are reported as the mean over 2,000 test episodes. The confidence intervals are close to those of recent methods and are omitted in the paper due to space limitations.