ProKeR: A Kernel Perspective on Few-Shot Adaptation of Large Vision-Language Models
Supplementary Material

Yassir Bendou[1*], Amine Ouasfi[2*], Vincent Gripon[1], Adnane Boukhayma[2]
[1]IMT Atlantique, Brest, France
[2]Inria, University Rennes, IRISA, CNRS

A. Detailed derivations

A.1. Nadaraya-Watson estimator

We first derive the solution of the adaptation problem for the Nadaraya-Waston estimator. The adaptation problem writes:

$$\phi(\mathbf{x}) \;=\; \arg\min_{\mathbf{q}} \frac{1}{NK}\sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\|\mathbf{q}-\mathbf{L}_i\|_2^2 + \|\mathbf{q}-f_{\text{clip}}(\mathbf{x})\|_2^2. \quad (1)$$

The derivation of the solution of Eq. 1 is as follows:

$$\mathcal{L} = \frac{1}{NK}\sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\|\mathbf{q}-\mathbf{L}_i\|_2^2 + \|\mathbf{q}-f_{\text{clip}}(\mathbf{x})\|_2^2$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{q}} = 0$$

$$\Rightarrow \frac{1}{NK}\sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\,(\mathbf{q}-\mathbf{L}_i) + \lambda\mathbf{q} - \lambda f_{\text{clip}}(\mathbf{x}) = 0$$

$$\Rightarrow \mathbf{q}\left(\lambda NK + \sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\right) = \lambda NK f_{\text{clip}}(\mathbf{x}) + \sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\mathbf{L}_i$$

$$\Rightarrow \mathbf{q} \;=\; \frac{\lambda NK}{\lambda NK + \sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))} f_{\text{clip}}(\mathbf{x}) + \frac{1}{\lambda NK + \sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))}\sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\mathbf{L}_i$$

A.2. Local Linear Regression

Here, we detail the derivation of the solution of the local linear regression (LLR). Let $\tilde{\mathbf{x}} \;=\; [1 \quad \mathbf{x}]$ and $\mathbf{A} \in \mathbb{R}^{(d+1)c}$ which minimizes the following problem:

$$\min_{\mathbf{A}} \frac{1}{NK}\sum_{i=1}^{NK} k_\beta(d(\mathbf{x},\mathbf{S}_i))\|\tilde{\mathbf{S}}_i\mathbf{A}-\mathbf{L}_i\|_2^2 + \lambda\|\tilde{\mathbf{x}}\mathbf{A} - f_{\text{clip}}(\mathbf{x})\|_2^2. \quad (2)$$

Let $\mathbf{\Omega}$ be the $NK \times NK$ matrix with ith diagonal element as $k_\beta(d(\mathbf{x},\mathbf{S}_i))$. The derivation is as follows:

$$\text{Let} \quad \mathcal{L} = \frac{1}{NK}\mathbf{\Omega}\|\tilde{\mathbf{S}}\mathbf{A}-\mathbf{L}\|_2^2 + \lambda\|\tilde{\mathbf{x}}\mathbf{A} - f_{\text{clip}}(\mathbf{x})\|_2^2$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{A}} = 0$$

$$\Rightarrow \frac{1}{NK}\tilde{\mathbf{S}}^\top\mathbf{\Omega}\left(\tilde{\mathbf{S}}\mathbf{A}-\mathbf{L}\right) + \lambda\tilde{\mathbf{x}}^\top\left(\tilde{\mathbf{x}}\mathbf{A} - f_{\text{clip}}(\mathbf{x})\right) = 0$$

$$\Rightarrow \left(\tilde{\mathbf{S}}\mathbf{\Omega}\tilde{\mathbf{S}} + \lambda NK\tilde{\mathbf{x}}^\top\tilde{\mathbf{x}}\right)\mathbf{A} = \tilde{\mathbf{S}}^\top\mathbf{\Omega}\mathbf{L} + \lambda NK\tilde{\mathbf{x}}^\top f_{\text{clip}}(\mathbf{x})$$

$$\Rightarrow \mathbf{A} = \left(\tilde{\mathbf{S}}\mathbf{\Omega}\tilde{\mathbf{S}} + \lambda NK\tilde{\mathbf{x}}^\top\tilde{\mathbf{x}}\right)^{-1}\left(\tilde{\mathbf{S}}^\top\mathbf{\Omega}\mathbf{L} + \lambda NK\tilde{\mathbf{x}}^\top f_{\text{clip}}(\mathbf{x})\right)$$

B. Comparison per dataset

We provide bellow in Fig. 1 per-dataset comparisons of Training-free Methods on 11 image classification datasets on the CoOp's benchmark.

*These authors contributed equally to this work

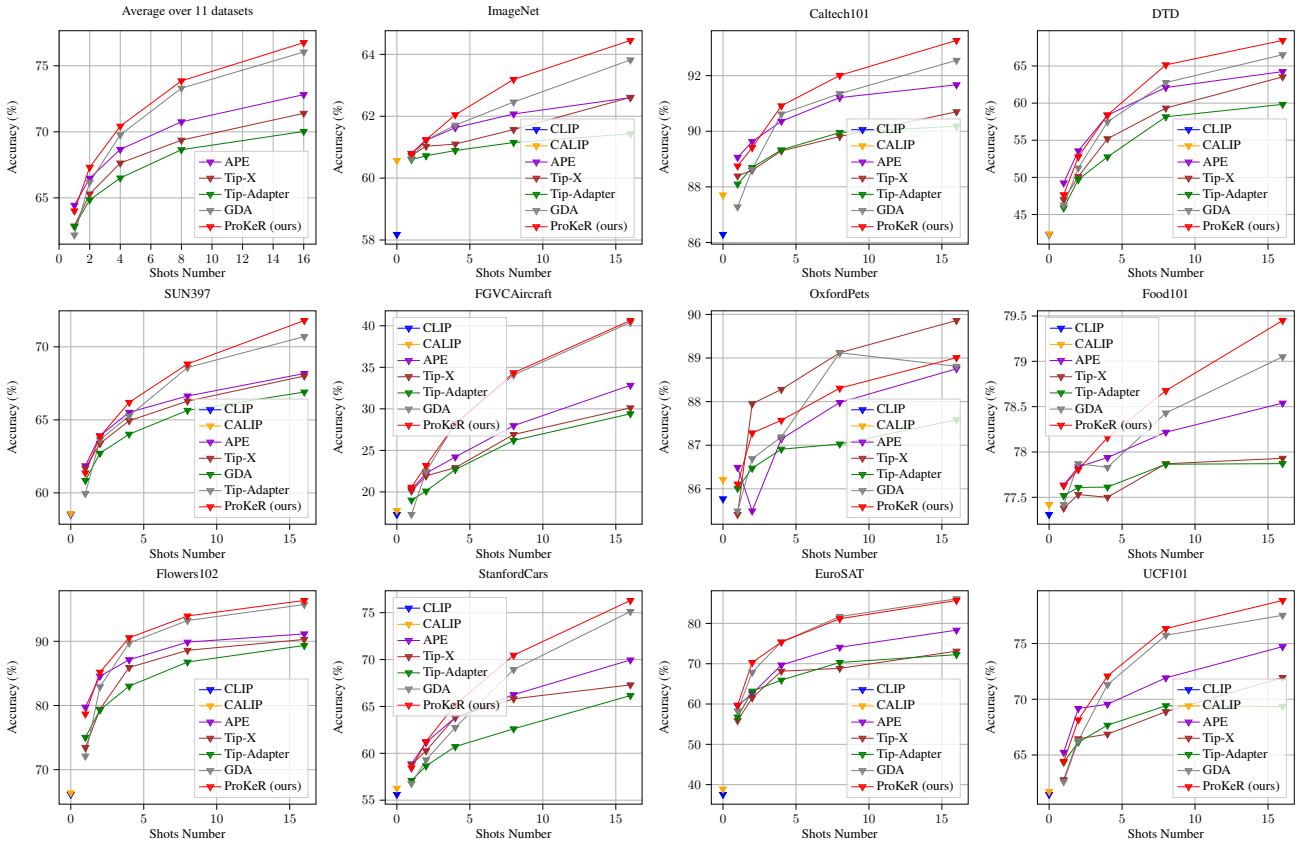

Figure 1. Few-shot Performance of Training-free Methods on 11 image classification datasets (CoOp's benchmark).

C. Sensitivity Analysis of λ

We analyze the sensitivity of λ in our method ProKeR in Tab. 1. We compute the average value for each dataset and study the effect of deviating from this value. Overall, performance is relatively stable in a range between 1/3 and up to 3 times this value, with only a drop of 1.2% in accuracy. Varying lambda up to a fifth or 5 times this value still only leads to a drop of 3%.

	Average
$\lambda \times 5$	73.17
$\lambda \times 4$	74.23
$\lambda \times 3$	75.24
$\lambda \times 2$	76.27
λ	76.58
$\lambda \times 2$	75.85
$\lambda \times 3$	75.11
$\lambda \times 4$	74.45
$\lambda \times 5$	73.84
ProKeR	76.75

Table 1. Sensitivity Analysis of λ on 11 datasets for 16-shots.