Supplementary Material for Linear Calibration Approach to Knowledge-free Group Robust Classification

Ryota Ishizaki
4624501@ed.tus.ac.jp
Shunya Yamagami
4623533@ed.tus.ac.jp
Yuta Goto
4623511@ed.tus.ac.jp
Go Irie
goirie@ieee.org

Tokyo University of Science
Tokyo, JP

1 Implementation Details

1.1 Details of Prompts

Following [1], we use the prompts for a class, a word in the initial vocabulary, and a positive pair described in Table 1.

	CelebA Dataset	Waterbirds Dataset
Class	"A photo of a celebrity with [class]."	"This is a picture of a [class]."
Word	"This is a picture of a [word]."	"This is a picture of a [word]."
Positive Pair	"A photo of a celebrity with [class] with [word]."	"This is a picture of a [class] with [word]."

Table 1: **Prompts for Waterbirds Dataset.**

1.2 Sparse Subspace Clustering Details

We use Sparse Subspace Clustering (SSC) [2] to discover a set of intrinsic subspaces of the text embeddings of the words (nouns) contained in the initial vocabulary. We first obtain the text embeddings of all the words by using CLIP's text encoder and then apply SSC to the set of the embeddings. For SSC, we first solve the following sparse L1 reconstruction problem to find the optimal reconstruction coefficient $\mathbf{w}_i$ for the i-th text embedding $\mathbf{a}_i, i \in \{1, \dots, V\}$:

$$\min_{\mathbf{w}_i} \|\mathbf{a}_i - \sum_{j \neq i} w_{ij} \mathbf{a}_j\|_2^2 + \lambda \|\mathbf{w}_i\|_1, \tag{1}$$

The weight of the L_1 regularization term λ is set to 1.0. The affine constraint $\sum_i w_i = 1$ is not used (the impact of this constraint will be analyzed in Sec. 2.2.). We perform spectral clustering on the matrix of the optimal reconstruction coefficients $W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_V]$ to get the clustering result. The number of clusters is set to 10 by default, and all the other hyperparameters for spectral clustering are set to the default value in sklearn.

Method	CelebA			Waterbirds		
	WG↑	Avg↑	Gap↓	WG↑	Avg↑	Gap↓
Ours (Captioning-based, k-means)	76.0	83.1	7.1	44.2	86.5	42.3
Ours (Captioning-based, GMM)	76.0	83.1	7.1	77.1	86.7	9.6
Ours (Captioning-based, SSC)	**82.2**	**84.2**	**2.0**	**79.4**	**88.5**	**9.1**
Ours (Retrieval-based, k-means)	82.1	83.1	1.0	47.8	**91.2**	43.4
Ours (Retrieval-based, GMM)	72.5	78.7	6.2	53.7	89.5	35.8
Ours (Retrieval-based, SSC)	**85.1**	**85.6**	**0.5**	**69.0**	**91.2**	**22.2**

Table 2: **Comparison of Clustering Methods.**

Method	affine constraint	CelebA			Waterbirds		
		WG↑	Avg↑	Gap↓	WG↑	Avg↑	Gap↓
Ours (Captioning-based)	✓	80.0	83.5	3.5	60.1	**89.0**	28.9
Ours (Captioning-based)		**82.2**	**84.2**	**2.0**	**79.4**	88.5	**9.1**
Ours (Retrieval-based)	✓	71.3	79.2	7.9	37.7	87.3	49.6
Ours (Retrieval-based)		**85.1**	**85.6**	**0.5**	**69.0**	**91.2**	**22.2**

Table 3: **Impact of Affine Constraint.**

2 Additional Results on Clustering

2.1 Impact of Clustering Methods

We use SSC to unveil the potential subspaces inherent in the initial vocabulary. However, it is not evident whether SSC is superior to other alternatives such as k-means. Aiming at validating the superiority of SSC, we compare it with k-means and Gaussian Mixture Models (GMM). The results are shown in Table 2. We found that SSC is the best choice for all the datasets. This is intuitive and reasonable, given that CLIP's text embeddings are learned in an inner product space.

2.2 Impact of Affine Constraint in SSC

We investigate the impact of the affine constraint in SSC, i.e., $\sum_i w_i = 1$. From the results shown in Table 3, we can see that this constraint reduces accuracy. The reason for this may be that the space for the CLIP's text embeddings is not strictly affine, resulting in inconsistency with the constraint.

References

[1] Ching-Yao Chuang, Varun Jampani, Yuanzhen Li, Antonio Torralba, and Stefanie Jegelka. Debiasing vision-language models via biased prompts. *arXiv preprint arXiv:2302.00070*, 2023.

[2] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *TPAMI*, 35(11):2765–2781, 2013.