Why Is Prompt Tuning for Vision-Language Models Robust to Noisy Labels? Supplementary Material

A. Extended Experimental Results

Robustness Attribution. The findings presented in Table 1 indicate that the reason behind the noise robustness observed in prompt tuning can be attributed to the structured form imposed on class embeddings by CLIP's pre-trained text encoder.

Dataset	Method	Noise rate			
		0	12.5	25	50
OxfordPets	Classifier-R	74.82	64.10	55.96	36.63
	Classifier-C	81.47	70.29	61.87	44.21
	TEnc-FT	84.38	70.73	61.11	41.21
	Prompt Tuning	**87.89**	**84.62**	**81.20**	**73.13**
Caltech101	Classifier-R	88.19	74.48	61.14	42.68
	Classifier-C	89.94	77.04	63.81	45.96
	TEnc-FT	**90.75**	76.67	62.76	46.45
	Prompt Tuning	90.65	**82.51**	**78.70**	**70.13**
Flowers102	Classifier-R	83.40	69.58	60.85	37.74
	Classifier-C	94.11	80.26	69.18	47.55
	TEnc-FT	**94.62**	80.91	70.83	49.54
	Prompt Tuning	91.71	**86.24**	**81.92**	**71.80**
Food101	Classifier-R	63.80	54.66	46.23	28.97
	Classifier-C	69.36	60.46	51.85	34.37
	TEnc-FT	71.30	61.60	52.64	34.74
	Prompt Tuning	**76.99**	**73.63**	**71.07**	**64.30**
FGVCAircr	Classifier-R	30.38	24.84	20.89	13.32
	Classifier-C	34.46	**29.97**	**25.91**	17.36
	TEnc-FT	**35.30**	29.66	25.31	17.42
	Prompt Tuning	27.13	25.07	23.34	**19.05**
DTD	Classifier-R	48.02	44.30	40.32	30.10
	Classifier-C	**63.83**	57.14	50.36	34.86
	TEnc-FT	63.61	55.47	48.21	33.12
	Prompt Tuning	62.86	**58.90**	**53.62**	**46.19**
UCF101	Classifier-R	67.16	58.33	50.34	31.07
	Classifier-C	71.87	64.12	54.79	38.01
	TEnc-FT	**73.74**	64.52	56.10	37.88
	Prompt Tuning	73.12	**68.73**	**67.66**	**60.93**

Table 1: Comparison of transfer performance at incremental noise rates between different variants.

Table 2 validates two observations: (a) the significance of the text encoder in offering robust regularization of the text embeddings to tackle noisy inputs (Prompt Tuning versus classifiers); and (b) the necessity of fixing the text encoder to prevent overfitting (Prompt Tuning versus TEnc-FT).

Dataset	Method	Noise rate			
		0	12.5	25	50
OxfordPets	Full-Prompt-Tuning	85.39	74.00	68.66	50.50
	CLS-Tuning	85.04	77.02	71.03	53.15
	Prompt Tuning	**87.89**	**84.62**	**81.20**	**73.13**
Caltech101	Full-Prompt-Tuning	89.21	74.20	61.26	45.92
	CLS-Tuning	89.13	76.84	62.27	48.64
	Prompt Tuning	**90.65**	**82.51**	**78.70**	**70.13**
Flowers102	Full-Prompt-Tuning	**93.93**	83.58	77.00	59.52
	CLS-Tuning	93.47	84.19	78.74	61.79
	Prompt Tuning	91.71	**86.24**	**81.92**	**71.80**
Food101	Full-Prompt-Tuning	72.36	63.14	55.29	38.69
	CLS-Tuning	72.07	63.91	56.97	41.73
	Prompt Tuning	**76.99**	**73.63**	**71.07**	**64.30**
FGVCAircraft	Full-Prompt-Tuning	**32.28**	**28.16**	**24.67**	16.76
	CLS-Tuning	30.84	27.86	24.51	17.63
	Prompt Tuning	27.13	25.07	23.34	**19.05**
DTD	Full-Prompt-Tuning	62.80	55.50	49.01	34.66
	CLS-Tuning	62.78	56.15	48.46	35.43
	Prompt Tuning	**62.86**	**58.90**	**53.62**	**46.19**
UCF101	Full-Prompt-Tuning	73.02	64.31	57.11	40.42
	CLS-Tuning	72.73	65.64	58.91	44.55
	Prompt Tuning	**73.12**	**68.73**	**67.66**	**60.93**

Table 2: Comparison of transfer performance at incremental noise rates between different prompt designs.

Robustness to Correlated Label Noise. Table 4 shows that transfer learning faces a greater challenge with confusion noise, resulting in a great decline in classification accuracy at higher noise ratios as opposed to random noise. This decline is evident in both prompt tuning and linear probes. The robustness of prompt tuning is evident in its ability to outperform linear probes, even when faced with more challenging noise types.

Integration with Noise-Robust Losses. We examine the effectiveness of a robust loss function applied to prompt tuning with noisy labels. In this study, We adopt Generalized Cross Entropy (GCE) [3] as a representative of robust loss functions for noise-robust learning. Specifically, cross-

entropy loss in Eq. 2 is replaced with GCE loss during the training process. Figure 1 shows results of applying GCE loss to prompt tuning and linear probing for CLIP's vision encoder. We observe that both transfer learning methods obtain an improvement of noise robustness by training with GCE loss. In particular, prompt tuning further enhances its inherent noise robustness. This outcome suggests that prompt tuning offers great applicability to couple with existing noise-robust loss functions. In addition to GCE, Figure 2 shows two additional robust loss functions: Symmetric Cross Entropy (SCE) [2] and Normalized Cross Entropy (NCE) combined with a Reverse Cross Entropy (RCE) [1]. Both losses also improve the noise robustness of prompt tuning, but GCE still achieves slightly better performance.

Dataset	Method	0-Shot
OxfordPets	Random Prompt	40.93$\pm$11.19
Caltech101	Random Prompt	59.65$\pm$9.56
Flowers102	Random Prompt	14.86$\pm$8.40
Food101	Random Prompt	34.63$\pm$11.10
FGVCAircraft	Random Prompt	3.43$\pm$1.97
DTD	Random Prompt	21.20$\pm$3.51
UCF101	Random Prompt	32.93$\pm$5.48

Table 3: CLIP zero-shot with random prompts.

Dataset	**Method**	**Random**	**Confusion**
OxfordPets	Linear Probe	46.42$\pm$0.88	41.39$\pm$1.87
	Prompt Tuning	73.13$\pm$3.76	66.55$\pm$2.02
Caltech101	Linear Probe	56.24$\pm$1.96	56.25$\pm$6.92
	Prompt Tuning	70.13$\pm$3.76	70.86$\pm$1.83
Flowers102	Linear Probe	68.92$\pm$0.76	45.94$\pm$0.69
	Prompt Tuning	71.80$\pm$1.00	69.63$\pm$1.31
Food101	Linear Probe	42.63$\pm$0.89	37.71$\pm$0.52
	Prompt Tuning	64.30$\pm$2.58	63.93$\pm$1.45
FGVCAircraft	Linear Probe	21.98$\pm$0.48	15.38$\pm$0.71
	Prompt Tuning	19.05$\pm$1.06	18.04$\pm$1.32
DTD	Linear Probe	42.29$\pm$2.12	37.69$\pm$1.70
	Prompt Tuning	46.19$\pm$2.12	45.76$\pm$1.23
UCF101	Linear Probe	54.05$\pm$1.19	50.90$\pm$1.45
	Prompt Tuning	60.93$\pm$0.94	59.11$\pm$0.70

Table 4: The impact of random and confusion label noise at a 50% noise rate on Linear Probing and Prompt Tuning strategies.

B. Capacity of classifiers with random prompt tokens

Prompt tuning has a limited parameter space given by the length of the prompt tokens. This parameter space is fundamentally different from that of a whole neural network and may present special properties related to the robustness of the model. Instead of updating the learnable prompts with noisy data, we evaluate the classifiers with *random prompts* . Table 3 summarizes the average zero-shot performance over 100 runs. Surprisingly, the results show that CLIP can achieve non-trivial zero-shot performance, even with random prompts. This indicates that as long as the classname token is provided to the pre-trained text encoder, CLIP is capable of computing non-trivial class embeddings for generic image classification.

C. Unsupervised Prompt Tuning Settings

Pseudo-labels are generated by CLIP zero-transfer with ResNet50 image encoder. We follow the prompt engineering used by CLIP. There are three types of hand-crafted prompts: "A photo of a <label name>" for generic object datasets; "A photo of a <label name>, a type of <collective name>" for fine-grained object datasets (e.g., prompts for OxfordPets are appended "a type of dog" or "a type of cat"); and "<label name> texture" for the DTD dataset. K is set to 16 in all experiments.

References

[1] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels. In *ICML*, 2020. 2

[2] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *ICCV*, 2019. 2

[3] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *NeurIPS*, 2018. 1, 3

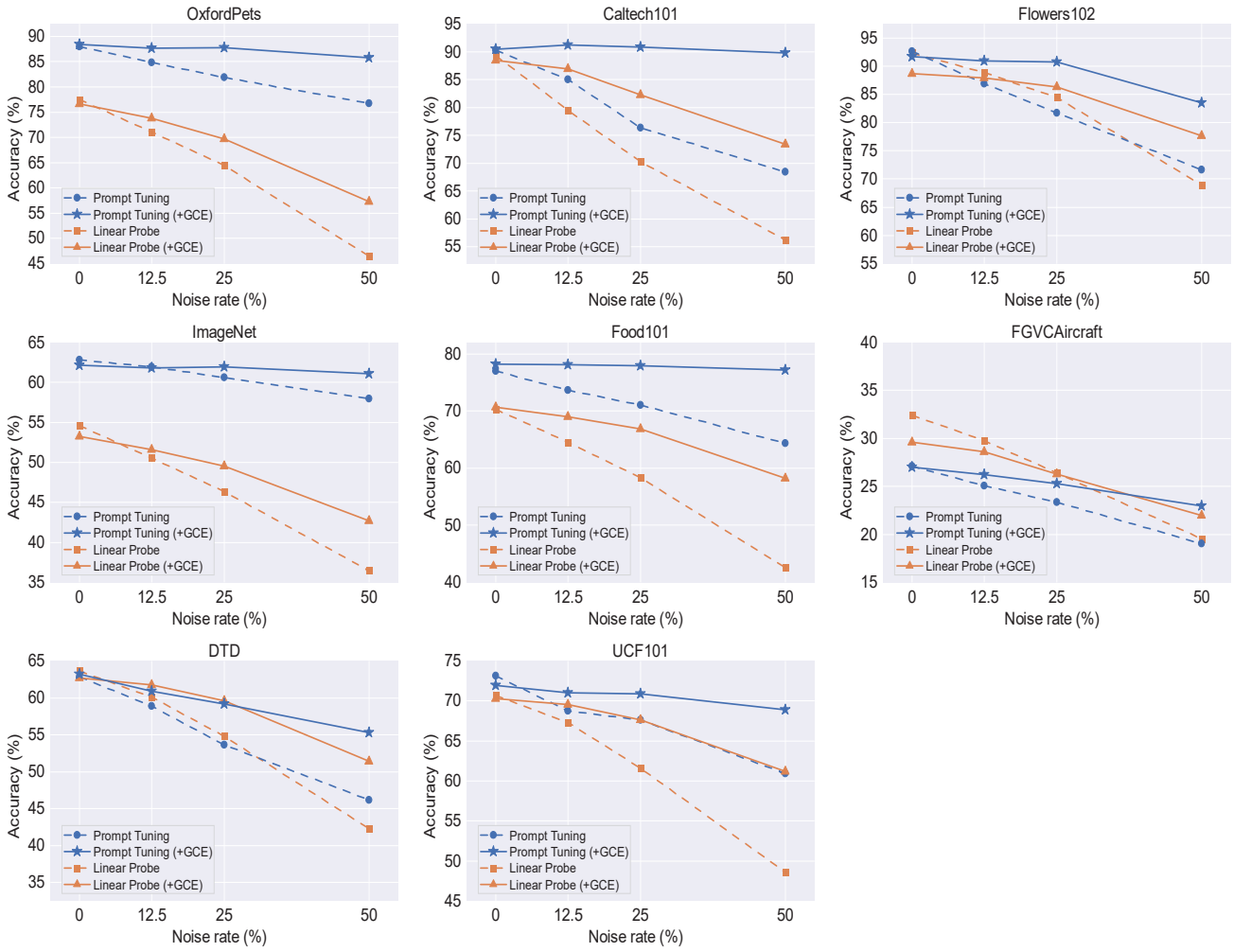

Figure 1: Incorporating the generalized cross-entropy (GCE) [3] loss with Prompt Tuning and Linear Probe methods, originally trained using cross-entropy, can enhance their noise robustness. At high noise rates. PromptTuning(+GCE) consistently and significantly outperforms other approaches on all datasets.

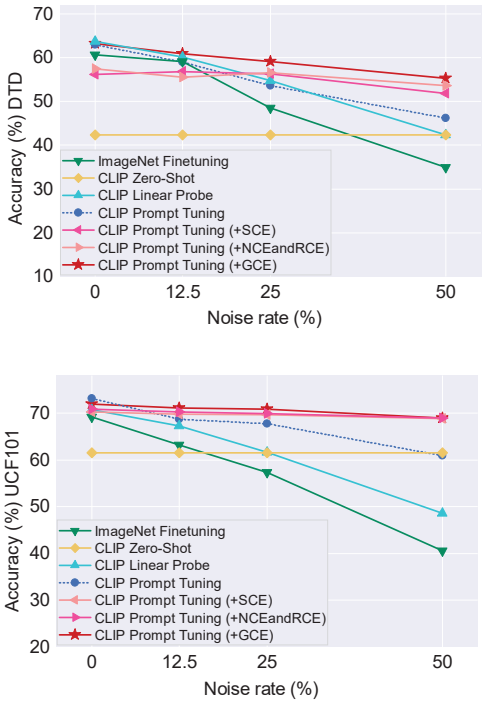

Figure 2: Combination of traditional transfer learning and prompt tuning approaches, with three robust loss functions.