JoAPR: Cleaning the Lens of Prompt Learning for Vision-Language Models

Supplementary Material

A. Algorithm

The algorithm for JoAPR is as follows:

Algorithm 1 JoAPR Training Process

Input: Training dataset $\overline{D} : \{(X_i, \overline{Y}_i)\}_{i=1}^{N}$

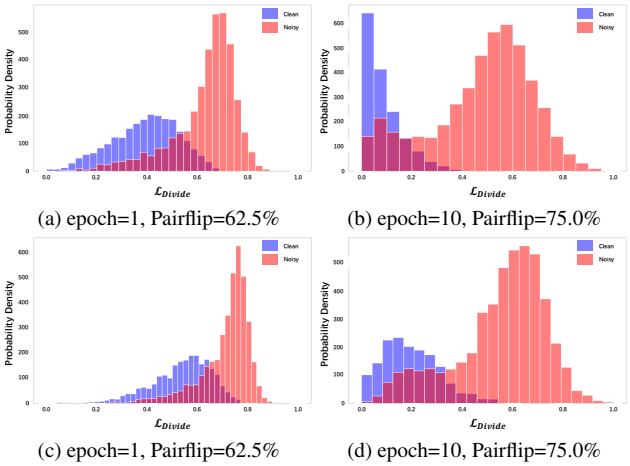
Warmup the model using $\mathcal{L}_{Warmup}$;
for n = 1:Max_epoch **do**
  Fit the $\mathcal{L}_{Divide}$ using GMM;
  Determine thresholds θ_1 and θ_2;
  $D_c : \{(X_i^c, Y_i^c) \mid \mathcal{L} < \theta_1 \; or \; p(g_{clean}|\mathcal{L}) > \theta_2\}_{i=1}^{N_c}$
  $D_n : \{(X_i^n) \mid \mathcal{L} > \theta_1 \; and \; p(g_{clean}|\mathcal{L}) < \theta_2\}_{i=1}^{N_n}$
  for m = 1:num_steps **do**
    Augment data as per Eq. (8);
    Refurbish labels using Eqs. (9) and (10), yielding $\widehat{D}$;
    $\widetilde{D}^{'}$ = MixMatch($\widehat{D}$)
    Retrain model with $\widetilde{D}^{'}$ using $\mathcal{L}_{Retrain}$;
    Update prompt using gradient backpropagation;
  end for
end for
Maximize $P(\overline{Y}|X) \rightarrow$ Maximize $P(\widetilde{Y}^{'}|X)$

B. Potential Issues with the Compensation Term

Figure 3 indicates that in few-shot datasets with a limited number of samples and high noise levels, the penalty term can lead to a highly discrete distribution of loss values. Alternatively, it may result in many noisy samples exhibiting very low loss values, to the extent that the loss values from noisy samples completely overshadow the distribution of clean samples. The inclusion of a compensation term aids in clustering the loss value distribution more tightly, narrowing the gap between the maximum and minimum loss values. This clustering increases the mean loss values for both clean and noisy data, thereby reducing their distribution overlap to a certain degree. However, in few-shot datasets with a larger number of samples, this approach could have negative consequences. As shown in Fig. 6, under conditions of a large sample size and high noise, the compensation term causes the loss value distribution of clean samples to more closely resemble that of noisy samples, leading to increased overlap. This phenomenon explains why JoAPR* tends to perform better than JoAPR in few-shot datasets with a greater number of samples.

C. Experiments on FGVCAircraft

Prompt learning methods often exhibit lower performance on this dataset. Hence, we consider it more essential to improve prompt learning performance on this dataset rather than robustness. Nevertheless, our experiments on FGVCAircraft yield improvements, with a 3.8% and 3.1% increase in accuracy under 50% Symflip and Pairflip noise, respectively. It is crucial to note that even when dealing with a completely clean FGVCAircraft dataset devoid of any label noise, achieving a satisfactory model fit remains challenging. Hence, we implement 150 epochs of Warmup and employ GCE during this phase to mitigate overfitting to noisy data.



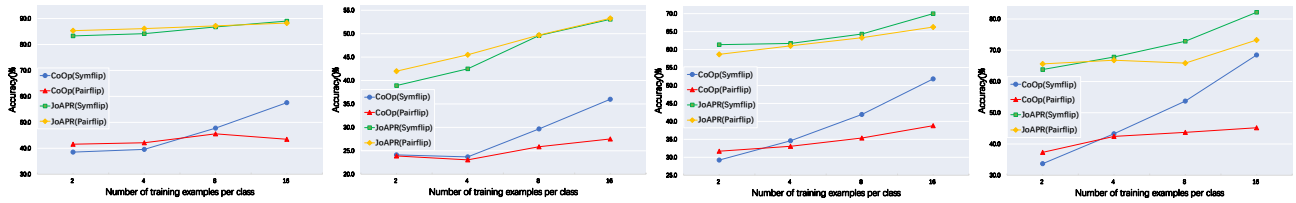
(a) epoch=1, Pairflip=62.5% (b) epoch=10, Pairflip=75.0%

(c) epoch=1, Pairflip=62.5% (d) epoch=10, Pairflip=75.0%

Figure 6. Pairflip noise injection into the SUN397 dataset is depicted with blue indicating clean data and red signifying noisy data. Figures (a) and (b) illustrate the probability density distribution when fitting the $\mathcal{L}_{Divide}$ without incorporating the compensation term. Conversely, figures (c) and (d) display the distribution with the compensation term included in the $\mathcal{L}_{Divide}$ fitting process.

D. More Analysis about JoAPR under 100% Noise

Our framework excels under 100% noise due to CLIP's inherent prior knowledge and powerful zero-shot learning ability, which allows us to accurately predict clean labels using model predictions. And that is why we adopt a refurbishment strategy for prompt learning for VL-PTMs in the case of label noise. We experiment on UCF101, where CLIP's zero-shot learning performance is comparatively lower, showing a 53.5% accuracy boost. While significantly improving, it slightly lags behind the two datasets in Tab. 5, validating our analysis.



(a) Caltech101 (b) DTD (c) UCF101 (d) Flowers102

Figure 7. The Few-shot curves for four datasets under 50% noise.

E. Few-shot Learning Analysis

As shown in Fig. 7, JoAPR significantly enhances the robustness of CoOp across various shot and noise types at a noise rate of 50%.

F. Utilization in PLOT

For better generalization validation, we compare it with PLOT [5]. As depicted in Tab. 6, JoAPR enhances the robustness not only of CoOp and CoCoOp but also of PLOT.

G. More discussion

While JoAPR has significantly enhanced the robustness of prompt learning methods for VL-PTMs, it is essential to acknowledge that JoAPR takes approximately two to three times more time than the baseline due to "partition and correction".

H. Additional Details on Training Procedures

The configuration of Warmup epochs and α_1 across nine datasets is detailed in Tab. 7. Specifically for Food101N, the Warmup epoch is configured to 1 and α_1 is set at 0.5, with $-\mathcal{H}$ being utilized in lieu of $\mathcal{L}_R$ during the retraining phase. It's important to note that fitting samples in some datasets can be challenging. Therefore, we opt for an increased number of epochs or lower penalty coefficients in the Warmup phase, particularly when dealing with a low noise ratio.

I. Illustrative Examples of Noisy Samples

In Fig. 8, we display the noisy samples in Food101N as classified by JoAPR, illustrating their original labels (in red) and the refurbished labels post JoAPR's modification (in blue). As Food101N is a dataset for fine-grained classification, it requires a more advanced capability to discern between noisy and clean data. Despite these complexities, the efficacy of JoAPR in accurately identifying and refurbishing these labels is remarkably evident.

Table 6. Comparison with PLOT

Dataset	Noise Type	Symflip			Pairflip		
	Method\Noise Ratio	25.0%	50.0%	75.0%	25.0%	50.0%	75.0%
Caltech101	PLOT	78.10	65.33	41.20	76.57	45.93	14.70
	JoAPR	**88.87**	**87.90**	**83.27**	**91.47**	**91.40**	**89.40**
Flowers102	PLOT	85.30	73.43	44.67	80.47	45.33	9.13
	JoAPR	**90.37**	**85.97**	**76.33**	**89.33**	**76.00**	**60.57**
OxfordPets	PLOT	73.60	58.07	28.67	71.70	42.63	13.20
	JoAPR	**86.90**	**88.10**	**88.30**	**88.53**	**88.77**	**86.33**
DTD	PLOT	54.67	42.13	23.87	51.33	30.50	10.17
	JoAPR	**59.13**	**56.40**	**49.13**	**59.03**	**54.40**	**46.70**

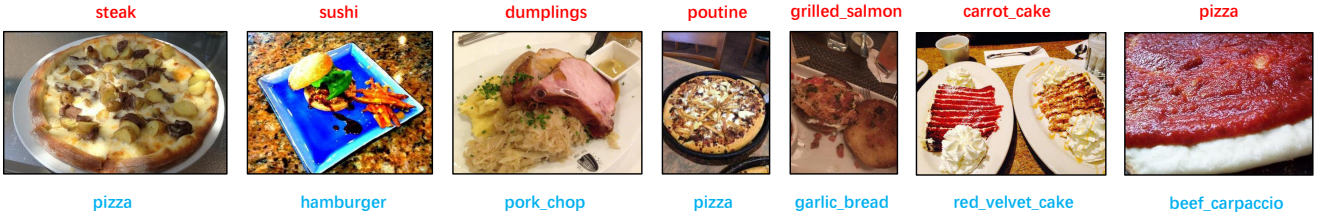


Figure 8. The visualization of label refurbishment by JoAPR in Food101N dataset noisy samples.

Table 7. The settings of Warmup epochs and α_1.

Dataset	Noise Type	Symflip						Pairflip					
	Config\Noise Ratio	12.5%	25.0%	37.5%	50.0%	62.5%	75.0%	12.5%	25.0%	37.5%	50.0%	62.5%	75.0%
ImageNet	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
SUN397	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Caltech101	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
StanfordCars	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
OxfordPets	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
DTD	Warmup epochs	1	1	1	1	1	1	1	1	1	1	1	1
	α_1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
EuroSAT	Warmup epochs	10	10	10	1	1	1	10	10	10	10	1	1
	α_1	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.1
Flowers102	Warmup epochs	20	20	20	20	20	20	20	20	20	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
UCF101	Warmup epochs	20	20	20	20	20	1	20	20	1	1	1	1
	α_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5