

Region-based Entropy Separation for One-shot Test-Time Adaptation

Kodai Kawamura
4620034@ed.tus.ac.jp

Tokyo University of Science
Tokyo, Japan

Shunya Yamagami
4623533@ed.tus.ac.jp

Go Irie
goirie@ieee.org

A Additional Results

A.1 Local vs. Global Entropy Minimization

A core of our approach is to employ local entropy minimization instead of global (image-level) entropy minimization, which has been explored in many of the existing model adaptation methods, e.g., [8]. To demonstrate the effectiveness of local entropy minimization, we compare the accuracy of the global and local entropy minimization losses when each is used alone. Table 1 shows the results. We can see that local entropy minimization is much better than global entropy minimization for all the datasets, which validate the effectiveness of our idea. Moreover, Ours, i.e., the method jointly performs local entropy minimization and global entropy maximization, shows superior performance, confirming the superiority of our method.

A.2 Other Style-emphasizing Transformation

We employ block shuffle and Gaussian blur to destruct class information to emphasize style information. In Table 2, we analyze the effectiveness of other possible style-emphasizing transformations including vertical flip, random crop and Gaussian noise. After all, we can see that block shuffle and Gaussian blur are significantly superior to the rest and effectively emphasize style information.

A.3 Per-Domain Performance

We show per-domain performance of the three datasets (Office-Home, VLCS, and PACS) in Tables 3, 4, and 5, respectively. Overall, our method is the only method that can improve Zero-shot CLIP in all the domains for all the datasets. In particular, our method outperforms all the existing methods in every domain of VLCS (Table 4) and outperforms the best competitive method (PromptStyler) by a significant margin (1.3%). These results underscore the high stability of our method. One notable advantage of our method would be that it achieves remarkable gains especially in the domains having styles very different from *photo*, such as *clipart* in Office-Home (more than 2% better than Zero-shot CLIP) and *sketch* in PACS

Method	Office-Home	VLCS	PACS
Global entropy minimization	54.18	75.03	75.80
Local entropy minimization	82.30	83.53	96.10
Ours	83.70	84.18	97.23

Table 1: **Local vs. Global Entropy Minimization** Accuracies of the model are listed. We can see that local entropy minimization works much better than global entropy minimization. Ours, i.e., the method incorporating local entropy minimization and global entropy maximization, shows superior performance.

Transformation	Accuracy
None	76.00
Block Shuffle	77.43
Gaussian Blur	79.78
Vertical Flip	75.55
Random Crop	72.39
Gaussian Noise	70.05

Table 2: **Effectiveness of style-emphasizing transformation.** Effectiveness of many kinds of style-emphasizing transformations is shown. we can see that Gaussian blur and vertical flip work well to emphasize style information.

(more than 3% better than Zero-shot CLIP), which proves the strong adaptation ability of our method.

A.4 Qualitative Analysis on Style-emphasized Images

We show the examples of style-emphasized images in Figure 1. We can see that the class information is destructed more by block shuffle as the number of grids increases. However, if we do not apply Gaussian blur, we can still see the remaining class information in some local regions. By applying Gaussian blur, it becomes more difficult to recognize the class information even in local regions. The prior work [9] shows that the style information is captured by abstracted information such as the mean and the standard deviation. Hence, Gaussian blur does not affect the style information. While the class information is disrupted, style information is not affected. Therefore, the transformation successfully emphasizes the style information.

A.5 Qualitative Analysis on Picked Up Patches

We show the visualization of picked up patches when we apply the local entropy minimization in Figure 2. It is clear from the images that our method successfully picks up the patches that contain rich class information in images of various domains: art, cartoon, and photo.

Method	Office-Home				
	Art	Clipart	Product	Real World	Mean
Zero-shot CLIP [14]	82.7	67.7	89.2	89.6	82.3
TAF-Cal [14]	61.5	55.0	74.9	77.0	67.1
Xiao et al. [14]	69.3	58.3	79.3	81.3	72.1
TPT [14]	78.1	63.2	83.2	81.9	76.6
DiffTPT [14]	76.2	61.3	82.1	81.0	75.2
PromptStyler [14]	83.8	68.2	91.6	90.7	<u>83.6</u>
Ours (Bottom-K)	<u>83.9</u>	69.8	<u>90.6</u>	90.5	83.7
Ours (Thresholding)	84.1	<u>69.6</u>	<u>90.6</u>	<u>90.6</u>	83.7

Table 3: **Results on Office-Home.** Accuracies of the model are listed. The highest accuracy is highlighted in bold, and the second highest accuracy is underlined. We test two region selection methods for local entropy minimization: “Bottom-K” picks regions with the K smallest predicted entropy, and “Thresholding” selects regions below a threshold τ .

Method	VLCS				
	Caltech	LabelMe	SUN09	VOC2007	Mean
Zero-shot CLIP [14]	100.0	68.9	74.8	85.9	82.4
TPT [14]	100.0	65.1	72.1	83.7	80.2
DiffTPT [14]	100.0	67.0	<u>75.5</u>	86.4	82.2
PromptStyler [14]	<u>99.9</u>	71.5	73.9	86.3	82.9
Ours (Bottom-K)	100.0	<u>71.9</u>	75.7	89.1	84.2
Ours (Thresholding)	100.0	72.0	<u>75.5</u>	<u>89.0</u>	<u>84.1</u>

Table 4: **Results on VLCS.** Accuracies of the model are listed. The highest accuracy is highlighted in bold, and the second highest accuracy is underlined. We test two region selection methods for local entropy minimization: “Bottom-K” picks regions with the K smallest predicted entropy, and “Thresholding” selects regions below a threshold τ .

Method	PACS				
	Art Painting	Cartoon	Photo	Sketch	Mean
Zero-shot CLIP [14]	97.2	99.1	99.9	88.2	96.1
TAF-Cal [14]	85.7	82.6	<u>96.1</u>	83.0	86.9
Xiao et al. [14]	82.0	79.7	95.9	79.0	84.2
TPT [14]	<u>97.8</u>	<u>99.2</u>	99.9	89.1	<u>96.5</u>
DiffTPT [14]	97.3	99.1	99.9	88.8	96.3
PromptStyler [14]	97.6	99.1	99.9	92.3	97.2
Ours (Bottom-K)	97.9	<u>99.2</u>	99.9	<u>91.9</u>	97.2
Ours (Thresholding)	97.9	99.3	99.9	91.8	97.2

Table 5: **Results on PACS.** Accuracies of the model are listed. The highest accuracy is highlighted in bold, and the second highest accuracy is underlined. We test two region selection methods for local entropy minimization: “Bottom-K” picks regions with the K smallest predicted entropy, and “Thresholding” selects regions below a threshold τ .

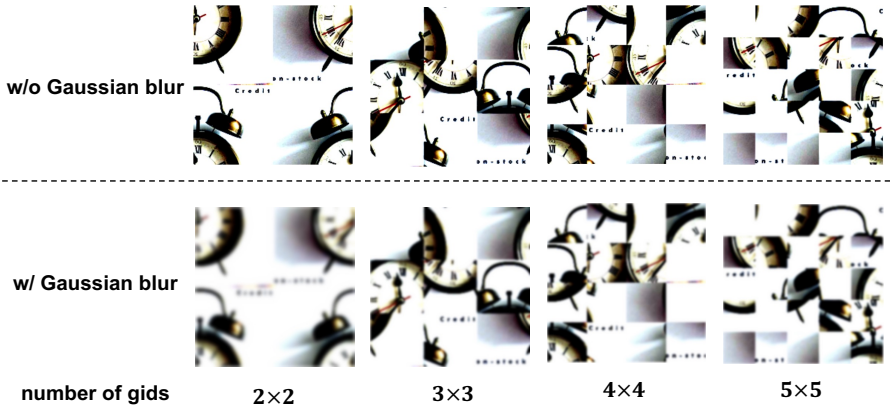


Figure 1: **Examples of transformed images.** We can see that the class information is destroyed more by block shuffle as the number of grids increases. Moreover, it becomes more difficult to recognize the class information by applying Gaussian blur.

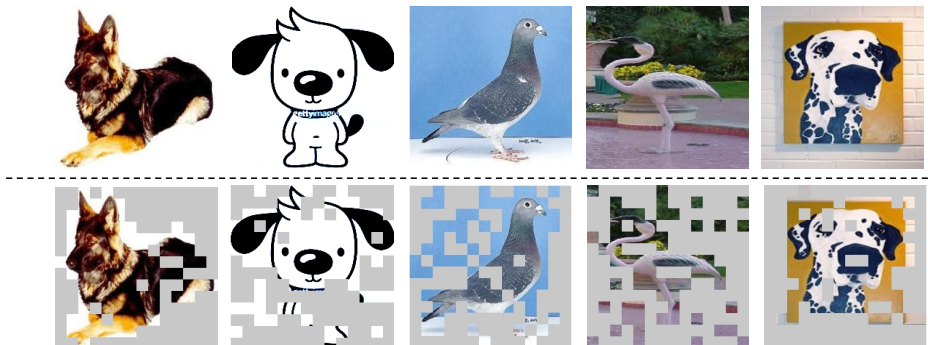


Figure 2: **Visualization of picked up patches.** We can see that our method successfully picks up the patches with strong class information.

References

- [1] Junhyeong Cho, Gilhyun Nam, Sungyeon Kim, Hunmin Yang, and Suha Kwak. Prompt-styler: Prompt-driven style generation for source-free domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15702–15712, 2023.
- [2] Chun-Mei Feng, Kai Yu, Yong Liu, Salman Khan, and Wangmeng Zuo. Diverse data augmentation with diffusions for effective test-time prompt tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2704–2714, 2023.
- [3] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.
- [4] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [5] Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, and Chaowei Xiao. Test-time prompt tuning for zero-shot generalization in vision-language models. *Advances in Neural Information Processing Systems*, 35:14274–14289, 2022.
- [6] Zehao Xiao, Xiantong Zhen, Ling Shao, and Cees GM Snoek. Learning to generalize across domains on single test samples. *arXiv preprint arXiv:2202.08045*, 2022.
- [7] Xingchen Zhao, Chang Liu, Anthony Sicilia, Seong Jae Hwang, and Yun Fu. Test-time fourier style calibration for domain generalization. *arXiv preprint arXiv:2205.06427*, 2022.