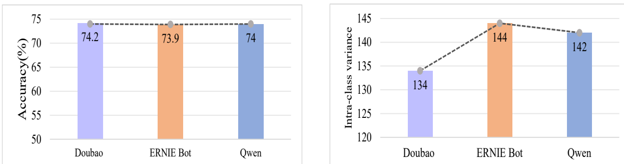
NRFP: A Noise-Robust Feature Plugin for Source-Free Domain Adaptation

Supplementary Material



(a) (b)

Figure 4. Performance comparison of auxiliary datasets generated by different AIGC models on Office-Home using ResNet-50. (a) Classification accuracy. (b) Average intra-class variance of auxiliary samples.

Figure 5. Examples of misleading auxiliary samples generated by ERNIE Bot, where the true class is *drill*.

6. More Experimental Results

6.1. Ablation Analysis on Different AIGC Models

We perform an ablation study using three AIGC models—Doubao, ERNIE Bot, and Qwen—to generate auxiliary datasets on the Office-Home benchmark. As illustrated in Fig. 4(a), the corresponding accuracies are 74.2, 73.9, and 74.0, respectively. Although different AIGC models yield slightly varied results, all consistently contribute to performance improvements over the baseline, confirming the generality of our framework.

To further analyze this difference, we compute the average intra-class variance, measured as the mean variance of the distances between auxiliary samples and their corresponding class centers, as shown in Fig. 4(b). A clear correlation emerges: a smaller average intra-class variance aligns with stronger adaptation performance. Specifically, Doubao produces the lowest variance and achieves the highest accuracy, while ERNIE Bot exhibits the largest variance and the weakest results.

Interestingly, ERNIE Bot presents a more challenging case, as its generated samples occasionally contain misleading semantics, as visualized in Figure 5. Even under such adverse conditions, our plugin consistently improves performance, demonstrating strong robustness to noisy or imperfect auxiliary data.

6.2. Auxiliary Sample Collection Settings

During the auxiliary sample collection for the Office–Home and DomainNet datasets, we recruited six volunteers who were not involved in this research domain. The volunteers were instructed to use the

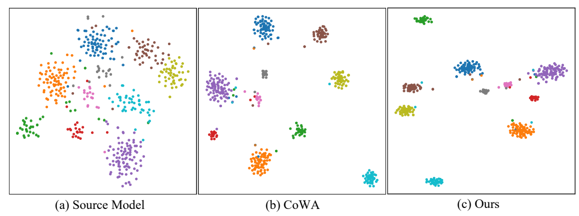


Figure 6. (a) Denotes the performance of the source model before domain adaptation.(b) Denotes the performance after CoWA domain adaptation.(c) Denotes the performance after CoWA domain adaptation with our plugin integrated.The experiments are conducted on 10 randomly selected categories of Office–Home dataset using ResNet–50 as the backbone network.

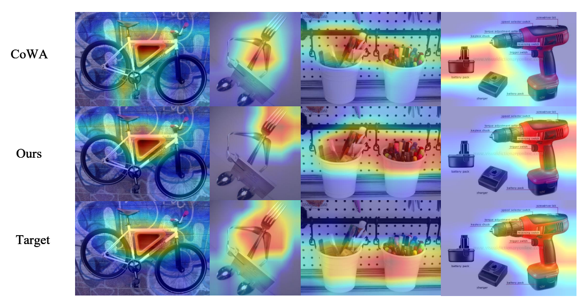


Figure 7. Baseline denotes the performance of CoWA without the plugin, NRFP denotes the performance of CoWA integrated with our plugin, and Target corresponds to the performance under a fully supervised setting. The experiments are conducted on Office-Home using ResNet-50 as the backbone network.

Doubao AIGC model with predefined prompts to generate the required auxiliary images. We only re-

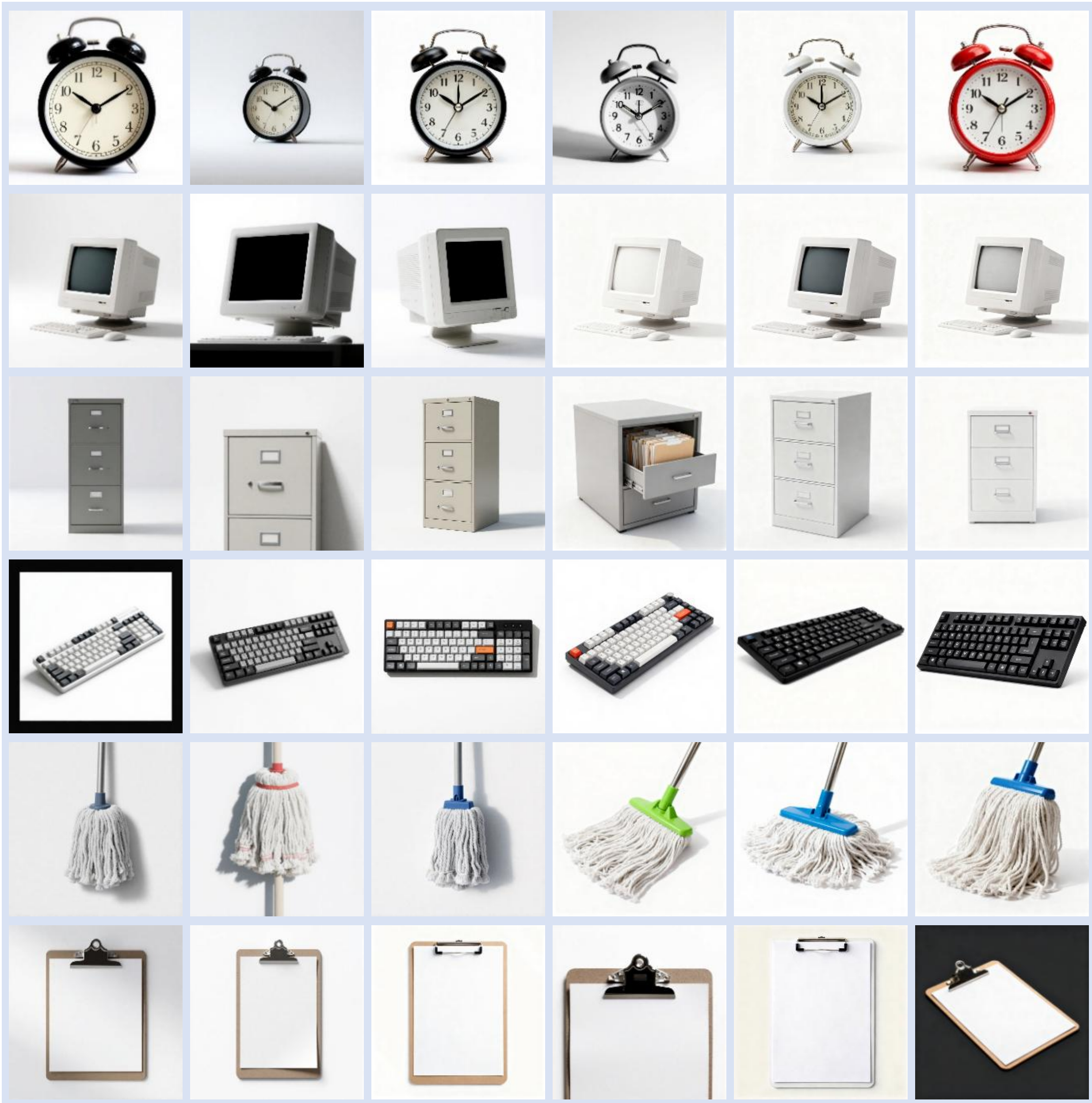


Figure 8. Auxiliary Office-Home was generated using an AIGC model (doubao), containing six categories: Alarm Clock, Computer, File Cabinet, Keyboard, Mop, and Clipboards.

quired that the generated images contain a clear subject and a plain–colored background, without any complex style control. The entire collection process was completed within 15 minutes for Office–Home and 45 minutes for DomainNet. Although the six volunteers inevitably exhibited some degree of subjectivity during the image generation process, the resulting auxiliary datasets maintained good semantic consistency and validity. These observations demonstrate the high efficiency and ease of implementation of our sample generation scheme.

6.3. Computational Resource Analysis

We further analyze the computational overhead introduced by the proposed plugin. On the Office–Home and DomainNet datasets, the additional computation time required by the plugin is 36 and 50 seconds, respectively. Although incorporating the plugin increases training time, this overhead remains moderate relative to the full training pipeline. More importantly, the added cost brings substantial performance gains, enhancing the model's generalization ability without modifying its architecture. From the perspective of balancing accuracy improvement and computational cost, the additional time expenditure is reasonable and well justified. It is also worth noting that our plugin does not rely on maintaining a memory queue and thus introduces no additional GPU memory consumption.

7. Visualization

7.1. Visualization of Auxiliary Sample

Figure 8 presents examples from six categories of the Office-Home auxiliary dataset generated by Doubao. Samples within each category exhibit strong semantic and visual consistency, suggesting that the performance gains of our method do not stem from auxiliary-sample diversity but from the effectiveness of the proposed design itself. We also observe a few anomalous cases, such as the first image in the fourth row, the second in the fifth row, and the sixth in the sixth row, which show inconsistent backgrounds or mild deviations from real-world objects. Nevertheless, the presence of such samples does not cause a significant performance drop, highlighting the robustness of our plugin even under slight noise and volunteer-induced subjective variations.

7.2. Visualization of t-SNE

To intuitively illustrate the advantages of our method in cross-domain feature alignment, we visualize the learned features using t-SNE [28] and compare them with those obtained from a baseline model without our method (Figure 6). The visualization includes 10 categories. Although the original CoWA framework improves performance in SFDA, target-domain samples of the same class remain widely dispersed, exhibiting large intra-class variations. In contrast, incorporating our method leads to noticeably more compact and well-organized feature clusters. This demonstrates that our approach effectively promotes the learning of more discriminative representations, thereby yielding clearer and more robust decision boundaries.

7.3. Visualization of Mutually Exclusive Samples

MESGS encourages the model to learn more discriminative features, thereby enhancing its generalization capability. To evaluate its effectiveness, we conducted visualization experiments. As shown in Fig. 9, the first row displays results without MESGS, while the second row shows results obtained after integrating MESGS.

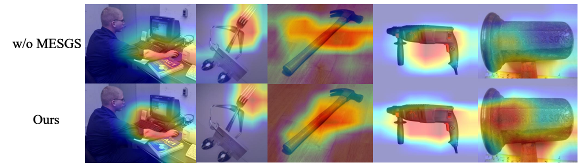


Figure 9. Visualization of the discriminative features learned by the model. The first row shows the results without the MESGS module, while the second row presents the results with MESGS enabled. The experiment is conducted on Office-Home using ResNet-50 as the backbone and integrated within the CoWA framework.

In the first column, without MESGS, the model fails to consistently focus on the main object (i.e., the computer). After applying MESGS, the attention becomes much more concentrated on the target object. In the second column, the image is relatively abstract, and accurate recognition of the "fork" category requires focusing on the top region of the fork—the key discriminative part. The visualizations reveal that MESGS substantially strengthens attention on this region, enabling the model to capture discriminative cues more effectively. These results confirm that MESGS successfully guides the model toward more discriminative features and improves its overall generalization performance.

7.4. Visualization of Attention

To further verify the effectiveness of our method, we visualize the model's attention distribution using Class Activation Mapping (CAM), as shown in Fig. 7. The first row presents the attention maps generated by CoWA without our plugin, the second row shows CoWA augmented with the plugin, and the third row corresponds to the fully supervised setting. Without the plugin (first row), the model's attention is scattered and fails to emphasize key regions. After introducing the plugin (second row), the model focuses more clearly on discriminative and semantically relevant areas. Compared with the fully supervised case (third row), the second row shows the highest degree of overlap in critical regions, demonstrating that the plugin effectively enhances the model's ability to attend to task-relevant features. These visualizations collectively illustrate that our method significantly improves attention localization and strengthens the model's capability to capture meaningful regions.