Bridging the Domain Gap: Synthetic Data Generation and Domain Adaptation for Real-World Image Classification

***Abstract*—This work investigates the use of synthetic images generated via a text-to-image model to train a deep neural network for real-world image classification. Using the STL10 dataset and focusing on the classes *airplane*, *bird*, and *car*, we explore the performance gap between a classifier trained solely on synthetic data and its performance on real images. We propose several strategies—including increasing synthetic data diversity, advanced augmentation (including test-time augmentation), and a domain adaptation framework based on Conditional Domain Adversarial Networks (CDAN) with pseudo-labeling, consistency regularization, spectral normalization, and RandAugment—to bridge this gap. Experimental results demonstrate that proper filtering, even on high OOD data, can lead to significant performance improvements.**

***Index Terms*—Synthetic data, domain adaptation, ResNet, CDAN, test-time augmentation, STL10.**

I. Introduction

Limited availability of labeled real-world data motivates the use of synthetic data for training deep learning classifiers. However, a domain shift exists between synthetic and real images, often leading to degraded performance on real data. In this study, synthetic images are generated using a Stable Diffusion model (via Hugging Face Diffusers) for the STL10 classes *airplane*, *bird*, and *car*. A modified ResNet18 is then fine-tuned on this synthetic data and evaluated on the STL10 test set. In addition, we leverage an unlabeled dataset (with $\sim$10% out-of-distribution (OOD) images) through a domain adaptation pipeline to further boost performance.

II. Methodology

A. Synthetic Data Generation

Synthetic images were generated using Stable Diffusion with multiple prompt variations. Since initial generations at 96$\times$96 lacked clarity and semantic relevance, images were first generated at 512$\times$512 and then downsampled to 96$\times$96 using Lanczos resampling. An ImageNet-pretrained ResNet validated the image quality by rejecting samples that did not match the class semantics. The dataset was then split into 80% training and 20% validation.

B. Classifier Training on Synthetic Data

We fine-tuned a modified ResNet18 (with a changed first convolution and removed max-pooling layer) on the synthetic training set. Aggressive data augmentation (random cropping, rotation, color jitter, Gaussian blur), dropout (30%), weight decay, and a cosine annealing learning rate scheduler were employed. Although synthetic validation accuracies approached 100%, initial tests on the STL10 real dataset revealed significant performance gaps.

TABLE I
Performance Summary for Synthetic Data Experiments

Experiment	Train Size/Class	Real Test Acc (%)
Initial	100	78.42
Scaled Data	400	80.58
Scaled Data	1000	87.25
TTA & Advanced Regularization	1000	91.42

C. Domain Adaptation

To further reduce the domain gap, we adopted a domain adaptation framework based on CDAN with the following components:

- **Pseudo-labeling with Temperature Scaling:** Predictions on unlabeled data were sharpened using a temperature parameter (T) and filtered using a high confidence threshold.
- **Adversarial Domain Adaptation:** A domain classifier with a gradient reversal layer encourages learning domain-invariant features.
- **Consistency Regularization:** Weak and strong augmentations (via RandAugment) enforce consistency in predictions.
- **Spectral Normalization:** Applied in the domain classifier to stabilize adversarial training.

Notably, when employing a confidence threshold of 0.99 and $T = 1$ on Unlabeled STL10 data (100,000 images with 70% OOD), the domain adaptation training achieved a synthetic validation accuracy of 99.87% and a real test accuracy of 93.50%, even when filtering out a significant amount of OOD data.

III. Experimental Results and Discussion

Table I summarizes the key experimental findings on synthetic data training.

Figure 1 displays the training curves for the synthetic data experiments (loss and accuracy on training and validation sets). Figure 2 illustrates the domain adaptation training curves, highlighting the high synthetic validation accuracy (99.87%) and the corresponding real test accuracy (93.50%). These

TABLE II
PERFORMANCE ON REAL TEST SET (UNLABELED DATASET 1667 IMAGES: 1500 IN-CLASS + 167 OOD)

Model	**Real Test Accuracy (%)**
Zero-Shot Inference on Pretrained ResNet18 *(manual mapping from ImageNet class index)*	31.29
Model Finetuned on Synthetic Images	91.42
Finetuned + Adversarial DA *(Confidence Threshold = 0.8)*	93.88
Finetuned + Adversarial DA CDAN *(Threshold = 0.95, $T = 0.3$)*	91.96
Finetuned + Adversarial DA CDAN, Consistency Regularization, Spectral Normalization, RandAugment	95.58

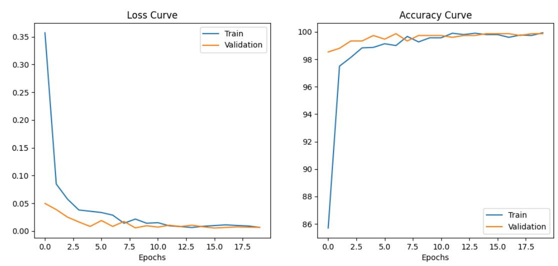

Fig. 1. Training curves for synthetic data experiments showing loss and accuracy on training and validation sets.

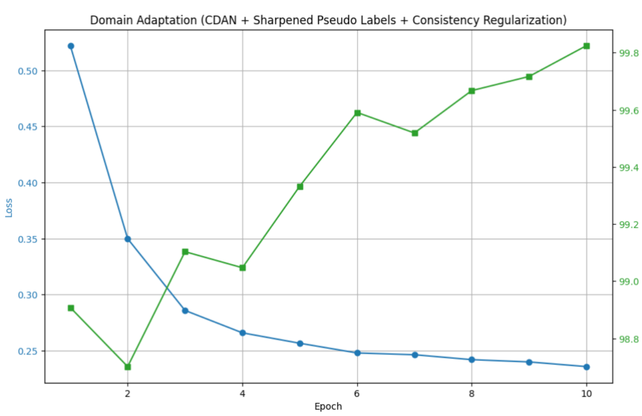

Fig. 2. Domain adaptation training curves (Adversarial DA CDAN with pseudo-label temperature scaling $T = 1$ and confidence threshold = 0.99) with a synthetic validation accuracy of 99.87% and real test accuracy of 93.50%.

results demonstrate that even in the presence of high OOD data, rigorous filtering and adversarial domain adaptation can substantially improve real-world performance.

Additionally, Table II presents performance results on a real test set(domain adaptation done on unlabeled dataset comprising 1667 images - 1500 images for the 3 classes and 167 images as 10% OOD data). These results compare zero-shot inference using a pretrained ResNet18 with various fine-tuning and domain adaptation strategies.

IV. CONCLUSION

The experiments confirm that while synthetic data training yields high in-domain accuracy, a significant domain gap remains when transitioning to real-world data. By increasing the diversity of synthetic images, applying advanced augmentation techniques, and integrating a domain adaptation framework, we substantially reduce this gap. In particular, the domain adaptation strategy employing CDAN with stringent pseudo-labeling (confidence threshold of 0.99 and $T = 1$) effectively handles high OOD presence, achieving a real test accuracy of 93.50%. Additionally, performance on a real test set shows that further adaptation techniques, including consistency regularization, spectral normalization, and RandAugment, can push the accuracy up to 95.58%. Future work will explore further improvements in generative modeling and semi-supervised learning to better bridge the domain gap.

V. FUTURE WORK

While our approach effectively narrows the domain gap, several avenues remain for exploration. First, we plan to expand the pipeline to handle more complex, multi-class scenarios with a larger number of classes. Second, exploring more advanced generative models (e.g., Latent Diffusion Models with additional control mechanisms) could yield higher-quality and more diverse synthetic data. Third, we aim to incorporate semi-supervised learning and self-supervised pretraining to better leverage unlabeled data, even when OOD content is high. Lastly, further investigation into curriculum-based domain adaptation, where the model progressively adapts from synthetic to real distributions, may provide additional gains in performance and robustness.

REFERENCES

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 10684–10695.

[2] M. Long, Z. Cao, J. Wang, and M. Jordan, “Conditional adversarial domain adaptation,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 1640–1650.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[4] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.
in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 10687–10698.

[5] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” in *International Conference on Learning Representations (ICLR)*, 2018.

[6] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, “Randaugment: Practical automated data augmentation with a reduced search space,” in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2020, pp. 3008–3017.