Experimenting Domain Adaptation for Synthetic-to-Real Image Classification

Abhishek Goyal
Yardi School of Artificial Intelligence
Indian Institute Of Technology Delhi
New Delhi, India
aib232073@iitd.ac.in

***Abstract*—This paper addresses the performance gap between classifiers trained on synthetic images and their performance on real-world images. We generate synthetic images for classes such as *cat*, *dog*, and *bird* using a text-to-image generation model and train a deep learning classifier solely on the synthetic data. The classifier is then tested on real data, and the performance gap is measured. Furthermore, we investigate whether increasing the synthetic dataset size or using domain adaptation techniques—leveraging an unlabeled dataset consisting 90% as images of interest and 10% out-of-distribution images—can reduce this gap.**

***Index Terms*—Synthetic Data, Domain Adaptation, Text-to-Image Generation, Image Classification.**

I. Introduction

Recent advancements in text-to-image generation have enabled the creation of high-quality synthetic images. However, when classifiers are trained solely on synthetic data, a significant performance gap often arises when the model is tested on real-world images. This paper explores techniques to mitigate this gap by increasing the synthetic dataset size and applying domain adaptation strategies.

II. Problem Statement

Our task is to:

- Generate synthetic images for selected classes.
- Divide the synthetic dataset into training and validation splits.
- Train a classifier using only synthetic images.
- Evaluate the classifier on a real dataset and measure the performance gap.
- Reduce the gap by either increasing synthetic data or using domain adaptation.
- Further improve the model using an unlabeled dataset that includes 10% out-of-distribution images.

III. Dataset

A. Synthetic Data

Synthetic images for the classes *cat*, *dog*, and *bird* were generated using the **CompVis/stable-diffusion-v1-4** text-to-image generation model. The following prompts were randomly employed to introduce diversity in the generated images:

- *in a natural outdoor setting*
- *classic disney style*
- *with vibrant colors*
- *captured in soft lighting*
- *with a shallow depth of field*
- *in a candid moment*
- *during golden hour*
- *with dramatic shadows*
- *with high contrast*
- *with a blurry background*
- *in a whimsical style*
- *with artistic flair*
- *in a studio setting*
- *with surreal colors*
- *in a minimalist composition*
- *with fine details*
- *in an urban environment*
- *with a vintage look*
- *in a lively scene*
- *with dynamic lighting*
- *in a serene mood*

These prompts, applied randomly during generation, resulted in a synthetic dataset that is diverse in style and quality. 1000 images per class were generated. Of these, 300 were kept for testing purpose and 700 were used for training the model.

B. Real Data

To evaluate the synthetic-trained model on real images, we utilized two publicly available datasets:

- **Oxford-IIIT Pet Dataset:** Used for real *cat* and *dog* images.
- **CUB-200-2011 Dataset:** Used for real *bird* images.

The real images were resized to 224×224 to match the input size of the classifier. Additionally, an unlabeled real dataset (which includes approximately 10% out-of-distribution images of elephants from ImageNet dataset) was used for domain adaptation.

IV. Methodology

We first train a classifier RESNET-18 [1] on the synthetic data. The performance gap is quantified by comparing the synthetic validation accuracy with that of same model on a real dataset. To reduce the gap, we adopt a Domain Adversarial

This work was supported in part of Assignment of course AIL862 by Prof. Sudipan Saha.

Fig. 1. Synthetic dog image.

Fig. 2. Real dog image.

Neural Network (DANN [2]) which uses a gradient reversal layer to align feature distributions of synthetic and real domains.

V. Experiments and Results

We conducted a series of experiments by varying the number of synthetic images used for training and by applying domain adaptation with an unlabeled real dataset. Our aim was to quantify and reduce the performance gap between models trained on synthetic data and their performance on real data. Table I summarizes four key experiments:

- **Exp 1:** Training on synthetic data with 350 images per class and testing on 300 real images per class.
- **Exp 2:** Training on synthetic data with 700 images per class and testing on 300 real images per class.
- **Exp 3:** Training on real data with 700 images per class and testing on 300 real images per class.
- **Exp 4:** Fine-tuning the synthetic-trained model using domain adaptation (DANN [2]) on filtered unlabeled real images (700 images in total after removing 10% out-of-distribution samples using trained Exp 1 model) and testing on 300 real images per class.

VI. Conclusion

The study demonstrates that even a classifier trained on synthetic data performs well on a real data set. However, the

TABLE I
Experimental Results for Synthetic and Real Data Training

Experiment	Training Data	Test Data	Accuracy (%)
Exp 1	Synthetic (350 images/class)	Real (300 images/class)	33.33
Exp 2	Synthetic (700 images/class)	Real (300 images/class)	66.56
Exp 3	Real (700 images/class)	Real (300 images/class)	80.00
Exp 4	Synthetic + DANN Fine-Tuning	Real (300 images/class)	61.22

model did not improve using DANN [2], which could be due to the kind of diffusion model selected, which already produces near-real images of cats, dogs, and birds. The second reason we could find out is that in synthetic data, the diversity of data was high, also making it difficult to distinguish between cat and dog breeds, which is not the case when model is trained on a real dataset (Oxford IIIT Pet Dataset).So, ideal results depend on classes and synthetic data given by the diffusion model. We cannot judge the capabilities of DANN in this regard.

Acknowledgment

This work was prepared as part of an assignment for the Special Topics in Computer Vision course at IIT Delhi.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.

[2] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016, pp. 90–99.