SINGLE-SHOT DOMAIN ADAPTATION VIA TARGET-AWARE GENERATIVE AUGMENTATIONS

Rakshith Subramanyam[+], *Kowshik Thopalli*[+], *Spring Berman*[+], *Pavan Turaga*[+]
Jayaraman J. Thiagarajan[†]

[+] Arizona State University, [†] Lawrence Livermore National Laboratory

ABSTRACT

The problem of adapting models from a source domain using data from any target domain of interest has gained prominence, thanks to the brittle generalization in deep neural networks. While several test-time adaptation techniques have emerged, they typically rely on synthetic data augmentations in cases of limited target data availability. In this paper, we consider the challenging setting of single-shot adaptation and explore the design of augmentation strategies. We argue that augmentations utilized by existing methods are insufficient to handle large distribution shifts, and hence propose a new approach `SiSTA` (Single-Shot Target Augmentations), which first fine-tunes a generative model from the source domain using a single-shot target, and then employs novel sampling strategies for curating synthetic target data. Using experiments with a state-of-the-art domain adaptation method, we find that `SiSTA` produces improvements as high as 20% over existing baselines under challenging shifts in face attribute detection, and that it performs competitively to oracle models obtained by training on a larger target dataset. Our codes can be accessed at `github.com/kowshikthopalli/SISTA`.

Index Terms— generalization, domain adaptation, augmentation, GANs, single-shot learning

1. INTRODUCTION

Despite producing high accuracies in the *i.i.d.* setting, deep models are known to fail unpredictably under real-world distribution shifts (or domain shifts). Such failures can be potentially mitigated by refining the model weights with data from the target domain of interest. A large class of approaches have been explored in this regard; popular examples include source free domain adaptation (SFDA) [1] and test-time adaptation (TTA) [2]. Not surprisingly, the effectiveness of these approaches can be significantly limited when sufficient target data is not available. In this paper, we consider the extreme scenario where only single-shot target data is accessible.

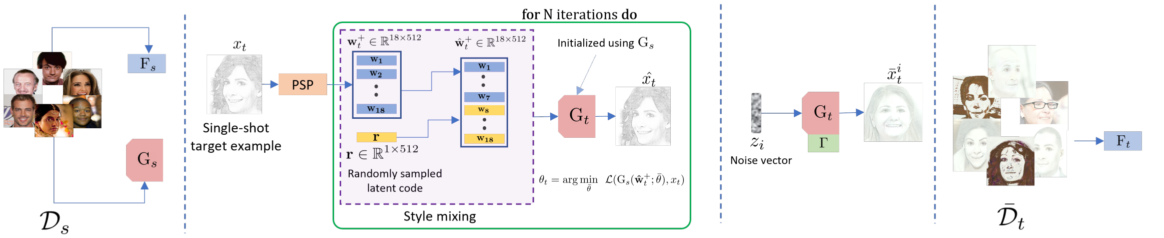
Driven by the data scarcity challenge in practical settings, data augmentation has emerged as a common fix for enabling model adaptation even with limited data. For example, the recently proposed MEMO [3] leverages pre-specified image augmentations (e.g., Augmix [4]) to expand the limited target data and performs test-time adaptation. Note, the success of such approaches directly hinges on how well the chosen augmentation can represent the target data distribution, and hence, in practice, different augmentation techniques may lead to varying degrees of generalization.

With the goal of advancing test-time adaptation with single-shot target data, we propose `SiSTA` a target domain-aware augmentation technique to synthetically generate target data, which can be used with any unsupervised domain adaption method for improving model generalization. At its core, our method relies on deep generative models, in particular StyleGANv2 [5], for data synthesis. To this end, `SiSTA` first adapts the source StyleGAN using a training strategy inspired from [6], and subsequently employs novel activation pruning strategies for sampling the target StyleGAN and curating a synthetic target dataset. Finally, this unlabeled dataset is used in conjunction with any SFDA method [7] to adapt source classifiers. Using empirical studies with multiple face attribute detection tasks and a variety of distribution shifts, we show that `SiSTA` significantly outperforms existing approaches and that it performs competitively to *oracle* models obtained by adapting with large target domain datasets.

2. BACKGROUND

Data augmentation has become an important tool for developing generalizable models, especially when operating in limited data settings. It has been shown that data augmentation can improve both in-distribution and out-of-distribution (OOD) accuracies [8]. Existing augmentations can be broadly viewed in two categories - (i) pixel/geometric corruptions and (ii) generative augmentations. The former category includes strategies such as CutMix [9], Cutout [10], Augmix [4], RandConv [11], mixup [12] and AutoAugment [13]. These domain-agnostic methods are known to be insufficient to achieve OOD generalization, especially under large domain

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344 and supported by the LDRD Program under project 21-ERD-012. This was also partly supported by the Defense Advanced Research Projects Agency (DARPA) under Grant HR00112290073.


F_s
G_s
$\mathcal{D}_s$
x_t
Single-shot target example
PSP
for N iterations do
$\mathbf{w}_t^+ \in \mathbb{R}^{18\times512}$
$\mathbf{w_1}$
$\mathbf{w_2}$
$\mathbf{w_{18}}$
$\hat{\mathbf{w}}_t^+ \in \mathbb{R}^{18\times512}$
$\mathbf{w_7}$
$\mathbf{w_8}$
$\mathbf{r}$
$\mathbf{r} \in \mathbb{R}^{1\times512}$
Randomly sampled latent code
Initialized using G_s
G_t
$\hat{x}_t$
$\theta_t = \arg\min_{\bar{\theta}} \mathcal{L}(\mathrm{G}_s(\hat{\mathbf{w}}_t^+;\bar{\theta}), x_t)$
Style mixing
z_i
Noise vector
Γ
$\bar{x}_t^i$
F_t
$\bar{\mathcal{D}}_t$

a) Source model and GAN training

b) Single-shot StyleGAN finetuning

c) Synthetic data generation with the target-domain generator

d) Source-free UDA with synthetic target data

Fig. 1. SiSTA: Assuming access to both the classifier and a StyleGAN from the source domain, we first adapt the generator to the target domain using a single-shot example. Next, we employ the proposed activation pruning strategies to construct the synthetic target dataset $\bar{\mathcal{D}}_t$. Finally, this dataset is used with any test-time adaptation technique for model refinement.

shifts. To circumvent this, more recent solutions have resorted to generative models (e.g., GANs) for synthesizing plausible augmentations [14]. Specifically, popular methods such as MBDG [15], CyCADA [16], GenToAdapt [17] and [18] have leveraged generative augmentations to better adapt to unlabeled target domains. However, these methods can be ineffective in cases of limited target data availability. In this work, we consider the extreme setting of single-shot target data and assume no access to source data during adaptation. Our goal is to obtain generative augmentations using only single-shot target, which can then be used in conjunction with any existing SFDA technique [7, 1, 19, 2].

3. PROPOSED APPROACH

We investigate the problem of adapting source domain classifiers using a single-shot target example and propose SiSTA, a target domain-aware augmentation technique (see Figure 1).

Setup. Formally, we denote the labeled source data as $\mathcal{D}_s = \{(x_s^i, y_s^i)\}$ with images x_s^i and labels y_s^i and the single-shot target example as x_t. We assume that we have access to both the classifier $\mathrm{F}_s : x \to y$ with parameters Φ_s and the StyleGAN-v2 model (generator $\mathrm{G}_s : z \to x$ with parameters Θ_s and discriminator H_s) trained on the source dataset $\mathcal{D}_s$. Our goal is to generate a synthetic target dataset $\bar{\mathcal{D}}_t = \{\bar{x}_t^i\}$ and refine the source classifier to obtain the target hypothesis $\mathrm{F}_t(.; \Phi_t)$ using any domain adaptation method.

Step 1: Source training. We begin by training the source classifier F_s using labeled data $\mathcal{D}_s$. This is carried out using cross entropy loss and standard training configurations. In addition, we build a generative model for the source data distribution. More specifically, we use the StyleGAN-v2 architecture and infer G_s and H_s respectively.

Step 2: Single-shot StyleGAN finetuning. Next, we finetune G_s using only the single-shot example x_t, in order to generate images from the target domain. To this end, we first invert x_t onto the style space of G_s using a pre-trained encoder, e.g., Pixel2Style2Pixel or shortly PSP [20], which maps a given image into the style code $\mathbf{w}_t^+ \in \mathbb{R}^{18\times512}$. This latent code corresponds to 18 intermediate layers of StyleGAN-v2. By design, x_t may be out of the training distribution $P_s(x)$ and hence the reconstruction corresponding to $\mathbf{w}_t^+$ is more likely to resemble the source domain. Consequently, we need to refine the generator G_s to synthesize images that are characteristic of the target domain.

Algorithm 1: Single-shot StyleGAN fine-tuning

```
Input: Target sample $x_t$, No. of training iterations $N$,
       Source generator $\mathrm{G}_s$, PSP encoder E.
Output: Target domain StyleGAN $\mathrm{G}_t$.
Invert the target sample to obtain $\mathbf{w}_t^+ = \mathrm{E}(x_t)$;
for $n$ in 1 to $N$ do
    Generate random style latent $\mathbf{r} \in \mathbb{R}^{1\times512}$;
    Perform style-mixing, i.e., replace layers 8-18 of $\mathbf{w}_t^+$
      with $\mathbf{r}$;
    Generate image $\hat{x}_t = \mathrm{G}_s(\hat{\mathbf{w}}_t^+)$;
    Update parameters $\Theta_t = \arg\min_{\bar{\Theta}} \mathcal{L}(\hat{x}_t, x_t; \mathrm{H}_s)$;
end
return $\mathrm{G}_t$ with parameters $\Theta_t$.
```

We take inspiration from JoJoGAN [6], a recent optimization strategy for style transfer in GANs, and update the generator model parameters based on a loss function defined on the activation outputs from the frozen discriminator H_s:

$$\Theta_t = \arg\min_{\bar{\Theta}} \sum_{\ell} \|\mathrm{H}_s^\ell(\mathrm{G}_s(\mathbf{w}_t^+; \bar{\Theta})) - \mathrm{H}_s^\ell(x_t)\|, \qquad (1)$$

where Θ_t refers to the parameters of the updated generator G_t, H_s^ℓ denotes the activations from layer ℓ of the discriminator H_s, and this objective minimizes the discrepancy between the target image and the reconstruction from the generator. Since this optimization can be highly unstable with a single x_t, we construct attribute-shifted versions of x_t through a style-mixing protocol, wherein the latent codes corresponding to a pre-specified subset of layers in $\mathbf{w}_t^+$ are replaced with randomly generated latents obtained by transforming a noise vector $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with the mapping network in StyleGAN-v2. In particular, we replace the layers 8 to 18 of $\mathbf{w}_t^+$, as it

```
Algorithm 2: Generating synthetic target data
Input: Target GAN $\mathrm{G}_t(.;\Theta_t)$, Source GAN $\mathrm{G}_s(.;\Theta_s)$,
       Pruning strategy $\Gamma$, Pruning ratio $p$
Output: Sampled image $\bar{x}_t$
Generate a random latent code $\mathbf{w}^+ \in \mathbb{R}^{18\times512}$;
for $\ell$ in 8 to 18 do
    $\beta \sim \mathrm{RandInt}(0,1)$;
    if $\beta == 1$ then
        Obtain layer $\ell$ activations $h_t^\ell$ from $\mathrm{G}_t(\mathbf{w}^+)$;
        /* Iterate over activation channels $K^\ell$ */
        for $k$ in 1 to $K^\ell$ do
            $\tau_p = p$-th percentile of $h_t^\ell[:,:,k]$;
            if $\Gamma ==$ prune-zero then
                $h_t^\ell[i,j,k] = 0$ if $h_t^\ell[i,j,k] < \tau_p, \forall i,j$;
            else
                Obtain activations $h_s^\ell$ from $\mathrm{G}_s(\mathbf{w}^+)$;
                $h_t^\ell[i,j,k] = h_s^\ell[i,j,k]$ if
                  $h_t^\ell[i,j,k] < \tau_p, \forall i,j$;
            end
        end
    end
end
return Image $\bar{x}_t = \mathrm{G}_t(\mathbf{w}^+;\Gamma)$
```

is known [21] that the initial layers encode the key semantic content, while the later layers contain style characteristics. In each iteration of our optimization, a different style-mixed latent code $\hat{\mathbf{w}}_t^+$ is used with (1). Algorithm 1 lists this procedure.

Step 3: Synthetic data generation. Once we obtain the adapted StyleGAN generator G_t for the target domain, we can build our synthetic dataset by sampling in its latent space. Despite the efficacy of such an approach, the inherent discrepancy between the true target distribution $P_t(x)$ and the approximate $Q_t(x)$ (synthetic data) can limit the generalization. We propose to address this by perturbing the latent representations from different layers of G_t to realize a more diverse set of style variations. More specifically, we introduce two strategies based on activation pruning, which identify all activations (at the output of each layer) that are lower than the p^{th} percentile value and replace them with zero (referred as *prune-zero*) or with corresponding activations from the source GAN G_s (*prune-rewind*). While the former strategy attenuates the effect of the target generator neurons to synthesize variations, the latter attempts to implicitly sample along the geodesic between the source and target domains by mixing the activations from the two generators. Note, we perform the pruning only in layers 8-18 so that the semantic content of a sample is not changed. Algorithm 2 describes the sampling process and Figure 2 illustrates the synthetic data generated for a target domain (*pencil sketch*) using vanilla sampling (or base), prune-zero and prune-rewind strategies.

Step 4: Source-free UDA: Using the synthetically generated target domain data, we finally perform source-free adaptation of F_s to obtain the target hypothesis F_t. To this end, we employ NRC [7], a state-of-the-art SFDA method[1], which exploits the intrinsic neighborhood of the target data. Formally, NRC uses the following objective:

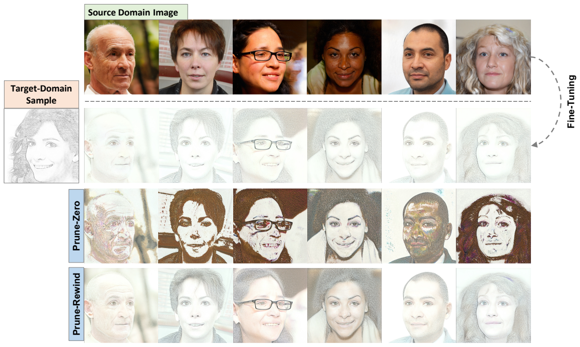


Fig. 2. Synthetic data generated using our proposed approach. In each case, we show the source domain image and the corresponding reconstructions from the target StyleGAN sampling (base), prune-zero and prune-rewind strategies.

$$\mathcal{L}_{\text{NRC}} = \mathcal{L}_{\text{neigh}} + \mathcal{L}_{\text{self}} + \mathcal{L}_{\text{exp}} + \mathcal{L}_{\text{div}}, \tag{2}$$

where $\mathcal{L}_{\text{neigh}}$ enforces prediction consistency of a sample with respect to its neighbors, $\mathcal{L}_{\text{self}}$ attempts to reduce the effect of noisy neighbors and $\mathcal{L}_{\text{exp}}$ considers the expanded neighborhood. Finally, $\mathcal{L}_{\text{div}}$ is the diversity maximization term implemented as the KL divergence between the distribution of predictions to an uniform distribution.

4. EMPIRICAL RESULTS

Dataset. For our empirical study, we consider the task of face attribute detection with images from the CelebA-HQ dataset. We used the pre-trained StyleGAN-v2 from [5] and emulated three different distribution shifts (referred as domains A, B, C in Figure 3). CelebA-HQ is a high-quality large-scale face attribute dataset with 30000 images, which is split into a source dataset with 18K images and the rest was used to construct the target domains. To emulate varying levels of distribution shift, we employed standard image manipulation techniques (we release this new benchmark dataset along with our codes[2]): (i) *Domain A*: We used the Stylization technique in OpenCV with $\sigma_s = 40$ and $\sigma_r = 0.2$; (ii) *Domain B*: For this

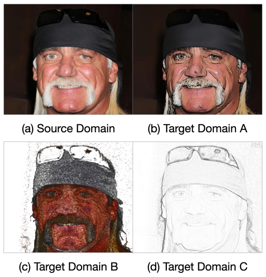

(a) Source Domain
(b) Target Domain A
(c) Target Domain B
(d) Target Domain C

Fig. 3. We emulate real-world shifts with increasing severity.

[1] https://github.com/Albert0147/NRC_SFDA
[2] **SiSTA**: https://github.com/kowshikthopalli/SISTA

Methods	Attribute: Smiling				Attribute: Gender			
	Domain A	Domain B	Domain C	Average	Domain A	Domain B	Domain C	Average
Source only	**89.78**	73.68	62.18	75.21	94.47	83.72	69.43	82.54
MEMO (AugMix)	89.34	71.76	59.43	73.51	94.04	82.45	58.51	78.33
MEMO (RandConv)	89.37	71.80	59.27	73.48	94.05	82.45	58.76	78.42
Ours (base)	84.80	82.53	83.29	83.54	**94.73**	87.52	89.44	90.56
Ours (prune-zero)	88.65	**85.75**	85.89	**86.76**	**94.75**	**89.03**	**93.49**	**92.42**
Ours (prune-rewind)	87.63	83.13	**85.99**	85.58	94.68	86.38	93.18	91.41
Oracle	92.34	87.92	88.80	89.69	96.91	92.13	95.42	94.82

Methods	Attribute: Arched Eyebrows				Attribute: Mouth Slightly Open			
	Domain A	Domain B	Domain C	Average	Domain A	Domain B	Domain C	Average
Source only	72.94	51.29	56.71	60.31	88.24	80.36	60.61	76.40
MEMO (AugMix)	72.72	51.23	56.38	60.11	88.22	80.30	60.60	76.37
MEMO (RandConv)	72.66	51.26	56.44	60.12	88.11	80.27	60.49	76.29
Ours (base)	76.39	73.57	65.37	71.78	91.07	82.49	69.84	81.13
Ours (prune-zero)	**79.23**	**74.41**	63.57	72.40	**92.36**	**84.75**	73.31	83.47
Ours (prune-rewind)	78.26	73.91	**69.68**	**73.95**	91.72	83.11	**77.22**	**84.02**
Oracle	81.85	72.94	80.09	78.29	92.94	88.39	87.78	89.70

Table 1. **Domain-aware augmentation significantly improves generalization.** We report the single-shot SFDA performance (Accuracy %) across different face attribute detection tasks and domain shifts. SiSTA consistently improves upon MEMO while also being competitive to the oracle. Through **bold** and underline formatting, we denote the top two performing methods.

shift, we used the PencilSketch technique in OpenCV with $\sigma_s = 40$ and $\sigma_r = 0.04$; and (iii) *Domain C*: This challenging domain shift was created by converting each color image to grayscale, and then performing pixel-wise division with a smoothed, inverted grayscale image. In our experiments, one randomly chosen example from each target domain was used for performing adaptation, and the performance on the entire target set of $12,000$ images is reported. We consider 4 facial attribute detection tasks: (i) *Smiling* (ii) *Gender* (iii) *Arched Eyebrows* and (iv) *Mouth Slightly Open*.

Experiment Setup. (a) *Source model training*: To obtain the source model F_s we fine-tune a Imagenet pre-trained ResNet-50 with labeled source data. We use a learning rate of $1e-4$, Adam optimizer and train for 30 epochs; (b) *StyleGAN fine-tuning*: For Algorithm 1, we set $N = 300$; (c) *Synthetic data curation*: In Algorithm 2, we set $p = 20\%$ for prune-rewind and $p = 50\%$ for prune-zero strategies, and generated 1000 samples in each case. Note, we experimented by varying the size of $\bar{\mathcal{D}}_t$ (between 100 and $10,000$). We found the performance to improve steadily until 1000 and no significant benefits were observed beyond 1000.; (d) *NRC SFDA training*: For NRC, we set both neighborhood and expanded neighborhood sizes at 5. Finally, we adapt F_s using SGD with momentum of 0.9 and learning rate of $1e-3$.

Baselines. In addition to the vanilla source-only baseline (no adaptation), we perform comparisons to the recent MEMO [3] technique - an online SFDA method which enforces prediction consistency between a image and its augmented variants. In particular, we implement MEMO with two popular augmentation strategies namely Augmix and RandConv [11]. Finally, we report the oracle performance *i.e.*, NRC performance when all 12000 unlabeled target data are available as opposed to our single-shot setting.

Findings. From Table 1, it can be observed that, SiSTA produces an average improvement of $\sim 10\%$ (across the three domain shifts) compared to the source-only baseline as well as the state-of-the-art MEMO. This improvement can be directly attributed to the efficacy of our generative augmentations, which can more effectively reflect the characteristics of the target domain than the pre-specified augmentations. While MEMO performs comparably to the source-only baseline at mild domain shifts (Domain A), it fairs poorly under severe shifts (Domain C). This clearly evidences the limitation of pixel-level corruptions used by MEMO in handling large domain shifts. Furthermore, with our approach, using the proposed activation pruning strategies leads to consistent improvements over the naïve sampling (base), due to the increased diversity in the curated target dataset. Finally, despite using only single-shot data, SiSTA performs competitively to the oracle model obtained by using the entire target set (12K samples) for adaptation ($2\% - 6\%$ gaps on average).

5. CONCLUSION

In this paper, we explored the use of generative augmentations for test-time adaptation, when only a single-shot target is available. Through a combination of StyleGAN fine-tuning and novel sampling strategies, we were able to curate synthetic target datasets that effectively reflect the characteristics of any target domain. Our future work includes theoretically understanding the behavior of different pruning techniques and extending our approach beyond classifier adaptation.

6. REFERENCES

[1] Jian Liang, Dapeng Hu, and Jiashi Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6028–6039.

[2] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell, "Tent: Fully test-time adaptation by entropy minimization," in *International Conference on Learning Representations*, 2021.

[3] Marvin Zhang, Sergey Levine, and Chelsea Finn, "Memo: Test time robustness via adaptation and augmentation," *arXiv preprint arXiv:2110.09506*, 2021.

[4] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan, "Augmix: A simple data processing method to improve robustness and uncertainty," *arXiv preprint arXiv:1912.02781*, 2019.

[5] Tero Karras, Samuli Laine, and Timo Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[6] Min Jin Chong and David Forsyth, "Jojogan: One shot face stylization," *arXiv preprint arXiv:2112.11641*, 2021.

[7] Shiqi Yang, Joost van de Weijer, Luis Herranz, Shangling Jui, et al., "Exploiting the intrinsic neighborhood structure for source-free domain adaptation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29393–29405, 2021.

[8] Dan Hendrycks et al., "The many faces of robustness: A critical analysis of out-of-distribution generalization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8340–8349.

[9] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[10] Terrance DeVries and Graham W Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[11] Zhenlin Xu, Deyi Liu, Junlin Yang, Colin Raffel, and Marc Niethammer, "Robust and generalizable visual representation learning via random convolutions," in *International Conference on Learning Representations*, 2021.

[12] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.

[13] Ekin D. Cubuk, Barret Zoph, Dandelion Mané, Vijay Vasudevan, and Quoc V. Le, "Autoaugment: Learning augmentation strategies from data," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 113–123.

[14] Fei Yue, Chao Zhang, MingYang Yuan, Chen Xu, and YaLin Song, "Survey of image augmentation based on generative adversarial network," *Journal of Physics: Conference Series*, vol. 2203, no. 1, pp. 012052, feb 2022.

[15] Alexander Robey, George J Pappas, and Hamed Hassani, "Model-based domain generalization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20210–20229, 2021.

[16] Judy Hoffman et al., "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*. Pmlr, 2018, pp. 1989–1998.

[17] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa, "Generate to adapt: Aligning domains using generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8503–8512.

[18] Junxuan Huang, Junsong Yuan, and Chunming Qiao, "Generation for unsupervised domain adaptation: A gan-based approach for object classification with 3d point cloud data," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3753–3757.

[19] Hao-Wei Yeh, Baoyao Yang, Pong C Yuen, and Tatsuya Harada, "Sofa: Source-data-free feature alignment for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 474–483.

[20] Elad Richardson et al., "Encoding in style: a stylegan encoder for image-to-image translation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2287–2296.

[21] Zongze Wu, Dani Lischinski, and Eli Shechtman, "Stylespace analysis: Disentangled controls for stylegan image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12863–12872.