Target-Aware Generative Augmentations for Single-Shot Adaptation

Kowshik Thopalli [*1] **Rakshith Subramanyam** [*2] **Pavan Turaga** [2] **Jayaraman J. Thiagarajan** [1]

Abstract

In this paper, we address the problem of adapting models from a source domain to a target domain, a task that has become increasingly important due to the brittle generalization of deep neural networks. While several test-time adaptation techniques have emerged, they typically rely on synthetic toolbox data augmentations in cases of limited target data availability. 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 on a variety of benchmarks, distribution shifts and image corruptions, we find that SiSTA produces significantly improved generalization over existing baselines in face attribute detection and multi-class object recognition. Furthermore, SiSTA performs competitively to models obtained by training on larger target datasets. Our codes can be accessed at https://github.com/Rakshith-2905/SiSTA.

1. Introduction

Deep models tend to suffer a significant drop in their performance when there is a shift between train and test distributions (Torralba & Efros, 2011). A natural solution to improve generalization under such domain shifts is to adapt models using data from the target domain of interest. However, it is infeasible to obtain data from every possible target during source model training itself. Test-time adaptation has emerged as an alternate solution, where a source-trained model is adapted solely using target data without accessing the source data. However, the success of these source-free adaptation (SFDA) methods hinges on sufficient target data availability (Liang et al.; Yang et al., 2021). While there exist online adaptation methods such as TENT (Wang et al., 2021) and MEMO (Zhang et al., 2021), they are are found to be ineffective under complex distribution shifts and when target data is limited, often producing on par or only marginally better results than non-adaptation performance (Thopalli et al., 2022).

In this work, we investigate a practical, yet challenging, scenario where the goal is to adapt models under unknown distribution shifts with minimal target data. Specifically, we focus on the extreme case where only single-shot example is available. In such data scarce settings, it is common to leverage synthetic augmentations; examples range from image manipulations to adversarial corruptions (Gokhale et al., 2023). Despite their wide-spread adoption, the best augmentation strategy can vary for different shifts, and more importantly, their utility diminishes in the single-shot case. Another popular approach is to use generative augmentations (Yue et al., 2022), where data variants are synthesized through generative models. Despite being more expressive than generic augmentations, they require comparatively larger datasets for effective training.

We propose SiSTA, a new target-aware generative augmentation technique for SFDA with single-shot target data (see Figure 1). At its core, SiSTA relaxes the assumption of requiring source data, and instead assumes access to a source-trained generative model. We motivate and justify this assumption using a practical vendor-client implementation in Section 3. In this study, we consider StyelGAN as the choice for generative modeling, motivated by their flexibility in disentangling content and style. Our proposed algorithm has two steps, namely SiSTA-G and SiSTA-S, to fine-tune a source-trained StyleGAN with the target data, and to synthesize diverse augmentations respectively.

Our contributions can be summarized as follows:

1. We propose a new target-aware, generative augmentation technique for single-shot adaptation;

[*]Equal contribution [1]Lawrence Livermore National Laboratory, Livermore, CA, USA [2]Arizona State University, Tempe, AZ, USA. Correspondence to: Kowshik Thopalli <thopalli1@llnl.gov>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

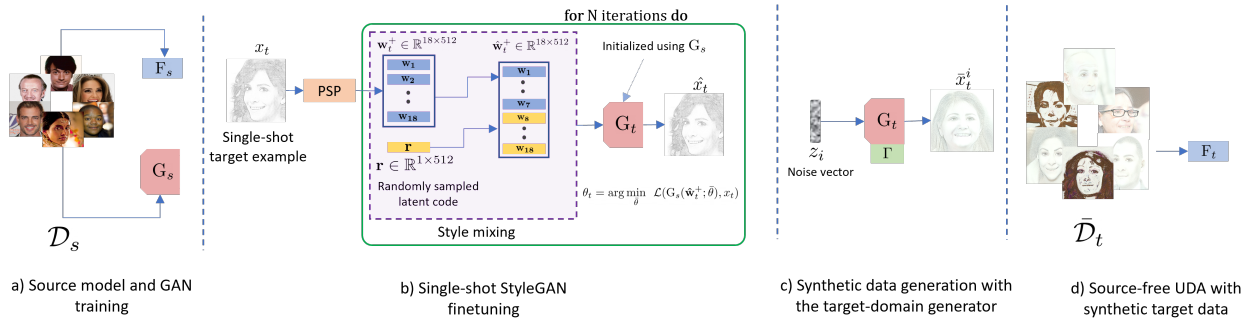


Figure 1. **SiSTA**: Assuming access to both the classifier and a StyleGAN from the source domain, we first adapt the generator to the target domain using a single-shot example. Next, we employ the proposed activation pruning strategies to construct the synthetic target dataset $\bar{\mathcal{D}}_t$.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

2. We introduce two novel sampling strategies based on activation pruning, *prune-zero* and *prune-rewind*, to support domain-invariant feature learning;

3. Using a popular SFDA approach, NRC (Yang et al., 2021), on augmentations from SiSTA, we show significant gains in generalization over SoTA online adaptation;

4. By benchmarking on multiple datasets (CelebA, AFHQ, CIFAR-10, DomainNet) and a wide variety of domain shifts (style variations, natural image corruptions), we establish SiSTA as a SoTA method for 1−shot adaptation;

5. We show the efficacy of SiSTA in multi-class classification using both class-conditional GANs as well as multiple class-specific GANs.

2. Background

Source free domain Adaptation: In the standard setting of SFDA we only have access to the pre-trained source classifier $\mathrm{F}_s : \mathrm{x} \rightarrow \mathrm{y}$ but not to the source dataset $\mathcal{D}_s = \{(\mathrm{x}_s^i, \mathrm{y}_s^i)\}$. Here, $\mathrm{x}_s^i \in \mathcal{X}_s$ and $\mathrm{y}_s^i \in \mathcal{Y}$ denote the ith image and its corresponding label from the source domain $\mathcal{X}_s$. Subsequently, the model needs to be adapted to a target domain $\mathcal{X}_t$ using unlabeled examples $\mathcal{D}_t = \{(\mathrm{x}_t^j\}$, where $\mathrm{x}_t^j \in \mathcal{X}_t$. Note, the set of classes $\mathcal{Y}$ is pre-specified and remains the same across all domains.

A number of approaches to SFDA have been proposed in the literature and can be categorized into two groups: methods which perform adaptation by fine-tuning the source classifier alone, and those that update the feature extractor as well for promoting domain invariance. In the former category, adaptation is typically achieved through unsupervised/self-supervised learning objectives; examples include rotation prediction (Sun et al., 2020), self-supervised knowledge distillation (Liu & Yuan, 2022), contrastive learning (Huang et al., 2021) and batch normalization statistics matching (Wang et al., 2021; Ishii & Sugiyama, 2021). The second category includes state-of-the-art approaches such as SHOT (Liang et al.), NRC (Yang et al., 2021) and N2DCX (Tang et al., 2021), which utilize pseudo-labeling based optimization, and often require sufficient amount of data to update the entire feature extractor meaningfully.

While SHOT is known to be effective under challenging shifts, it relies on global clustering to obtain pseudo-labels for the target data, and in practice, can fail in some cases due to the prediction diversity among samples within a cluster. The more recent NRC (Yang et al., 2021) alleviates this by exploiting the neighborhood structure through the introduction of affinity values that reflect the degree of connectedness between each data point and its neighbors. This inherently encourages prediction consistency between each samples and its most relevant neighbors. Formally, the optimization of NRC involves the following objective:

$$\mathcal{L}_{\text{NRC}} = \mathcal{L}_{\text{neigh}} + \mathcal{L}_{\text{self}} + \mathcal{L}_{\text{exp}} + \mathcal{L}_{\text{div}} \tag{1}$$

where $\mathcal{L}_{\text{neigh}}$ enforces prediction consistency of a sample with respect to its neighbors, while $\mathcal{L}_{\text{self}}$ attempts to reduce the effect of noisy neighbors and $\mathcal{L}_{\text{exp}}$ considers expanded neigbhorhood structure. Finally, $\mathcal{L}_{\text{div}}$ is the widely adopted diversity maximization term implemented as the KL divergence between the distribution of predictions in a batch to a uniform distribution. While SiSTA can admit any SFDA technique, we find NRC to be an appropriate choice, since it updates the feature extractor and utilizes the local semantic context to improve performance. This is particularly important in the context of our rich synthetic augmentations, which exhibit a high degree of diversity.

Generative Augmentations: It is well known that the performance of SFDA methods suffers when the target dataset is sparse. To mitigate this, synthetic augmentations are often leveraged. While it has been found that data augmentation can improve both in-distribution and out-of-distribution (OOD) accuracies (Steiner et al., 2021; Hendrycks et al., 2021), their use in SFDA is more recent. Existing aug-

mentations can be broadly viewed in two categories - (i) pixel/geometric corruptions, and (ii) generative augmentations. The former category includes strategies such as CutMix (Yun et al., 2019), Cutout (DeVries & Taylor, 2017), Augmix (Hendrycks et al., 2020), RandConv (Xu et al., 2021), mixup (Zhang et al., 2018) and AutoAugment (Cubuk et al., 2019). These domain-agnostic methods are known to be insufficient to achieve OOD generalization, especially under complex domain shifts. To circumvent this, generative augmentations based on GANs or Variational Autoencoders (VAEs) have emerged. These methods involve training a generative model to synthesize new samples (Yue et al., 2022). These augmentations have been used in various tasks such as image-to-image translation and improving generalization under shifts. For example, methods such as MBDG (Robey et al., 2021), CyCADA (Hoffman et al., 2018), 3C-GAN (Rahman et al., 2021) and GenToAdapt (Sankaranarayanan et al., 2018) have leveraged generative augmentations to better adapt to unlabeled target domains. However, by design, these methods require large amounts of data from both source and target domains. In contrast, SiSTA focuses on obtaining target-aware generative augmentations by fine-tuning source-trained generative models using only a single-shot target sample.

StyleGAN-v2 Architecture: While significant progress has been made in generative AI, including StyleGANs and denoising diffusion models (Saharia et al., 2022), we utilize StyleGAN-V2 as the base generative model in our work. This choice is motivated by the flexibility that StyleGANs offer in producing images of different styles, which can be attributed to the inherent disentanglement of style and semantic content in their latent space. Existing approaches works (Wu et al., 2021a;b) have studied this disentanglement property and uncovered the StyleGAN's ability to manipulate the style of an image projected onto the latent space by replacing the latent codes corresponding to only style. Another recent study (Chong & Forsyth, 2021) reported that by leveraging such manipulations, one can perform style transfer with a limited number of paired examples. Interestingly, it has also been recently found (Wu et al., 2021b) that, even after transferring a GAN to a different data distribution (faces to cartoons), the latent space of the adapted GAN is point-wise aligned with the source StyleGAN. We take inspiration from these works to develop our single-shot GAN fine-tuning protocol as well as our novel sampling strategies to enable domain-invariant feature learning.

3. Proposed Approach

In this section, we introduce SiSTA, a new target-aware, generative augmentation strategy with the goal of improving domain adaptation of pre-trained classifiers using single-shot target data. While SFDA methods are known to be effective under a variety of distribution shifts, their performance hinges on the availability of a sufficient amount of target data. In this work, we propose to relax SFDA's assumption on source data access by requiring a source-trained generative model (StyleGANs in our study) to synthesize augmentations in the target domain, in order to enable effective adaptation even under limited data. In particular, we consider the extreme, yet practical setting where only $1-$shot target data is available.

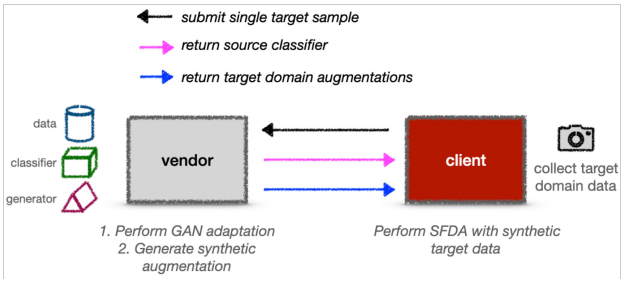


Figure 2. **A high-level illustration of our adaptation approach** SiSTA, which is carried out on the *vendor* side that stores the source classifier and a generative model. Designed to support single-shot adaptation, SiSTA returns target-aware synthetic augmentations. Finally, the *vendor* executes any SFDA technique to update the source classifier using the synthesized augmentations.

Figure 2 illustrates an implementation of such a setup where the source dataset, classifier, and the pre-trained generator are available only on the *vendor* side. A *client* that wants to adapt the classifier to a novel domain submits the one-shot target data and receives both the source classifier as well as the synthetic generative augmentations. Finally, the *client* executes any SFDA approach to update the classifier using only the unlabeled synthetic data. This implementation eliminates the need for the *vendor* to share their generative model, while also minimizing the amount of *client* data that gets shared.

As described earlier, SiSTA is comprised of two key steps that are carried out on the *vendor* side: (i) SiSTA−G: Fine-tune a pre-trained StyleGAN generator G_s using single-shot target data $\{\mathrm{x}_t\}$ under unknown distribution shifts.; and (ii) SiSTA−S: Synthesize diverse samples $\mathcal{D}_t = \{\bar{\mathrm{x}}_t^j\}$ using the fine-tuned generator G_t to support effective classifier adaptation to the target domain. Finally, we leverage the recently proposed NRC method to perform *client*-side adaptation. Now, we describe these steps in detail.

3.1. SiSTA−G: Single-Shot StyleGAN Fine-Tuning

Our goal in this step is to fine-tune G_s using only the single-shot example x_t from the target domain to produce an updated generator G_t. To this end, the proposed approach first inverts x_t onto the style-space of G_s. In practice, this can be

Algorithm 1 `SiSTA-G`

Input: Target sample x_t, No. of training iterations M, Source generator G_s, Inversion module $\mathcal{E}$ Set of style layers $\mathcal{L}_{\text{st}}$.
Output: Fine-tuned generator G_t.
Invert the target sample to obtain $\mathbf{w}_t^+ = \mathcal{E}(\mathrm{x}_t)$
for m in 1 to M **do**
  Generate random style latent $\mathbf{r}^+$
  Perform style-mixing, *i.e.*, replace style layers $\mathcal{L}_{\text{st}}$ of $\mathbf{w}_t^+$ with $\mathbf{r}^+$
  Generate image $\hat{\mathrm{x}}_t = \mathrm{G}_s(\hat{\mathbf{w}}_t^+)$
  Update parameters Θ_t using (2)
end for
return: G_t with parameters Θ_t.

done using one of the following strategies: (i) a pre-trained encoder such as Pixel2Style2Pixel (Richardson et al., 2021) or E4E (Tov et al., 2021), which maps a given image into the style code $\mathbf{w}_t^+ \in \mathbb{R}^{L\times 512}$. This latent code corresponds to L intermediate layers of a StyleGAN model (e.g., $L = 18$ in StyleGAN-v2); (ii) any standard GAN inversion technique to infer an approximate solution in the style space (Xia et al., 2022); (iii) text-guided inversion such as StyleClip (Patashnik et al., 2021) if the label is available for the single-shot target image. Though conventional GAN inversion is known to be expensive, it will not be a significant bottleneck with only a single image.

Without loss of generality, the target domain is expected to contain distribution shifts w.r.t. the source domain, and hence the inverted solution in the style-space is more likely to resemble the source domain. For example, inverting a cartoon into the style-space of a GAN trained on real face images will produce a semantically similar image from the face manifold. Recent evidence (Subramanyam et al., 2022) suggests that one can accurately recover an OOD image using an additional vicinal regularization to the inversion process. However, in our case, we do not want an accurate reconstruction, but rather refine the generator G_s to emulate the characteristics of a target domain.

To this end, we utilize the following loss function defined on the activations from the source-domain discriminator H_s:

$$\Theta_t = \arg\min_{\bar{\Theta}} \sum_{\ell} \|\mathrm{H}_s^{\ell}(\mathrm{G}_s(\mathbf{w}_t^+;\bar{\Theta})) - \mathrm{H}_s^{\ell}(\mathrm{x}_t)\|_1, \quad (2)$$

where $\mathbf{w}_t^+$ is the style-space latent code obtained via GAN inversion, Θ_t refers to the parameters of the updated generator G_t and H_s^{ℓ} denotes the activations from layer ℓ of the discriminator H_s. Intuitively, this objective minimizes the discrepancy between the target image and the reconstruction from the updated generator. Note that, the parameters of the discriminator are not updated during this optimization. While any pre-trained feature extractor can be used for this optimization, the source discriminator provides meaningful gradients by comparing both the content and style aspects of the target image. Upon training, we expect the generator G_t to produce images resembling the target domain for any random latent code in the style-space.

An inherent issue with our objective is that, this optimization can be highly unstable when using a single x_t. To circumvent this, we leverage multiple, style-manipulated versions of x_t through a style-mixing protocol. More specifically, we first generate a random code $\mathbf{r}^+$ in the style-space (using the mapping network in StyleGAN). Next, we perform mixing by replacing the latent codes from a pre-specified subset of layers $\mathcal{L}_{\text{st}}$ in $\mathbf{w}_t^+$ using the corresponding codes from $\mathbf{r}^+$. In effect, this produces a modified image that contains the content from $\mathbf{w}_t^+$ and the style from $\mathbf{r}^+$. We denote this style-manipulated latent using the notation $\hat{\mathbf{w}}_t^+$. In each iteration of our optimization, a different style-mixed latent code $\hat{\mathbf{w}}_t^+$ is generated to compute the loss in (2). Algorithm 1 summarizes the steps of `SiSTA-G`.

Choosing layers for style-mixing. We choose $\mathcal{L}_{\text{st}}$ by exploiting the inherent style and content disentanglement in StyleGANs. Priors works (Wu et al., 2021a; Kafri et al., 2021; Karras et al., 2020) have established that the initial layers typically encode the semantic content, while the later layers capture the style characteristics. Since the exact subset of layers that correspond to style vary as the image resolution changes, following standard practice, we used $\mathcal{L}_{\text{st}} = 8 - 18$ when G_s produces images of size 1024×1024 and $\mathcal{L}_{\text{st}} = 3 - 8$ for images of size 32×32 (CIFAR-10).

3.2. `SiSTA-S`: Target-aware Augmentation Synthesis

Once we obtain the target domain-adapted StyleGAN generator G_t, we next synthesize augmentations by sampling in its latent space. Despite the efficacy of such an approach, the inherent discrepancy between the true target distribution $P_t(\mathrm{x})$ and the approximate $Q_t(\mathrm{x})$ (synthetic data) can limit generalization. Existing works (Kundu et al., 2020) have found that constructing generic representations (using standard augmentations) is useful for test-time adaptation any domain. However, in contrast, our goal is to produce augmentations specific only to a given target domain, thus enabling effective generalization even with single-shot data.

To this end, we propose two novel strategies that perturb the latent representations from different layers of G_t to realize a more diverse set of style variations. Both our sampling strategies are based on activation pruning, *i.e.*, identifying the activations in each style layer that are lower than the pth percentile value of that layer, and replacing them with (i) zero (referred to as *prune-zero*); or (ii) activations from the corresponding layer of the source GAN G_s (*prune-rewind*). The former strategy aims at creating a generic representation by systematically eliminating style information in the

Algorithm 2 SiSTA-S

Input: Target GAN $\mathrm{G}_t(.;\Theta_t)$, Source GAN $\mathrm{G}_s(.;\Theta_s)$, Pruning strategy Γ, Pruning ratio p, Set of style layers $\mathcal{L}_{\text{st}}$;
Output: Sampled image $\bar{\mathrm{x}}_t$
Draw a random latent code $\mathbf{w}^+$ from $\mathrm{G}_t(.;\Theta_t)$
for ℓ in $\mathcal{L}_{\text{st}}$ **do**
  $\beta \sim \text{RandInt}(0,1)$
  if $\beta == 1$ **then**
    Obtain layer ℓ activations h_t^ℓ from $\mathrm{G}_t(\mathbf{w}^+)$
    /* Iterate over activation channels V^ℓ*/
    for v in 1 to V^ℓ **do**
      $\tau_p = p$-th percentile of $\mathrm{h}_t^\ell[:,:,v]$
      if Γ == *prune-zero* **then**
        $\mathrm{h}_t^\ell[i,j,v] = 0$ **if** $\mathrm{h}_t^\ell[i,j,v] < \tau_p, \forall\, i,j$
      else
        Obtain activations h_s^ℓ from $\mathrm{G}_s(\mathbf{w}^+)$
        $\mathrm{h}_t^\ell[i,j,v] = \mathrm{h}_s^\ell[i,j,v]$ **if** $\mathrm{h}_t^\ell[i,j,v] < \tau_p, \forall\, i,j$
      end if
    end for
  end if
end for
return: Image $\bar{\mathrm{x}}_t = \mathrm{G}_t(\mathbf{w}^+;\Gamma)$

image. On the other hand, the latter attempts to create a smooth interpolation between the source and target domains by mixing the activations from the two generators. Note, we perform pruning only in the style layers, so that the semantic content of a sample is not changed. Note, we use the same set of style layers selected for performing SiSTA-G. Algorithm 2 lists the activation pruning step.

3.3. SiSTA-mcG: Extending to class-conditional GANs

When dealing with multi-class problems, it is typical to construct class-conditional GANs, $\mathrm{G}_s(.;\mathrm{c})$, to effectively model the different marginal distributions. In such settings, images from different classes get mapped to disparate sub-manifolds in the StyleGAN latent space. Assuming there are K different classes in $\mathcal{Y}$, we can directly apply SiSTA-G using 1-shot examples from each of the classes. The only difference occurs in the GAN inversion step, wherein we need to identify the conditioning variable c along with the latent code $\mathbf{w}_t^+$. Note, if the labels are available, one can estimate only $\mathbf{w}_t^+$. Finally, the algorithm 1 is repeated with K target images. We refer to this protocol as SiSTA-mcG (multi-class generation).

However, when we perform SiSTA-mcG using only a subset of the classes (say only one out K), there is a risk of not incorporating target-domain characteristics into the images synthesized for all realizations from the latent space. However, as we will show in the results (Figure 5a), even using

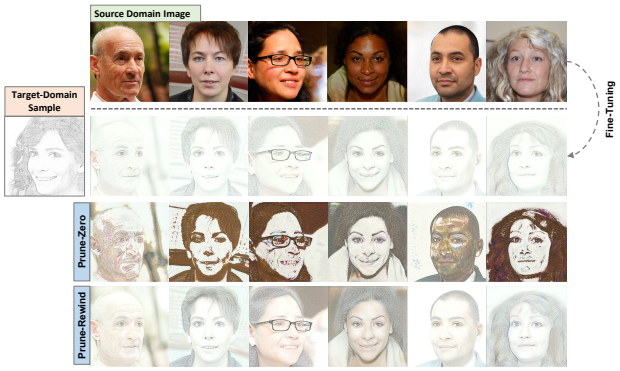


Figure 3. **Synthetic data generated using our proposed approach**. In each case, we show the source domain image and the corresponding reconstructions from the target StyleGAN sampling (base), prune-zero and prune-rewind strategies.

an example from a single class still leads to significantly improved generalization. We hypothesize that this behavior is due to the fact that the synthesized augmentations (random samples from G_t) arise from both $\mathcal{X}_s$ and $\mathcal{X}_t$, thus emulating an implicit mixing between the two data manifolds.

4. Experiments

We perform an extensive evaluation of SiSTA using a suite of classification tasks with multiple benchmark datasets, different StyleGAN architectures and more importantly, a variety of challenging distribution shifts. In all our experiments, we use single-shot target data and utilize publicly available, pre-trained StyleGAN weights.

4.1. Experimental Setup

Datasets: For our empirical study, we consider the following four datasets: (i) CelebA-HQ (Karras et al., 2017) is a high-quality (1024x1024 resolution) large-scale face attribute dataset with 30K images. We split this into a source dataset of 18K images and the remaining was used to design the target domains. We perform attribute detection experiments on a subset of 19 attributes, i.e., each attribute is posed as its own binary classification task; (ii) AFHQ (Choi et al., 2020) is a dataset of animal faces consisting of 15,000 images at 512×512 resolutions with three classes, namely cat, dog and wildlife, each containing 5000 images. For each class, 500 images were used to create the target domains, and the remaining was used as the source data; (iii) CIFAR-10 (Krizhevsky et al., 2009) is also a multiclass classification dataset with 60000 images at 32x32 resolution from 10 different object classes. We use the standard train-test splits for constructing the source and target domain datasets. While we used the StyleGAN-v2 trained on FFHQ faces for our experiments on the CelebA-HQ dataset[1], for AFHQ and

[1] https://github.com/rosinality/stylegan2-pytorch

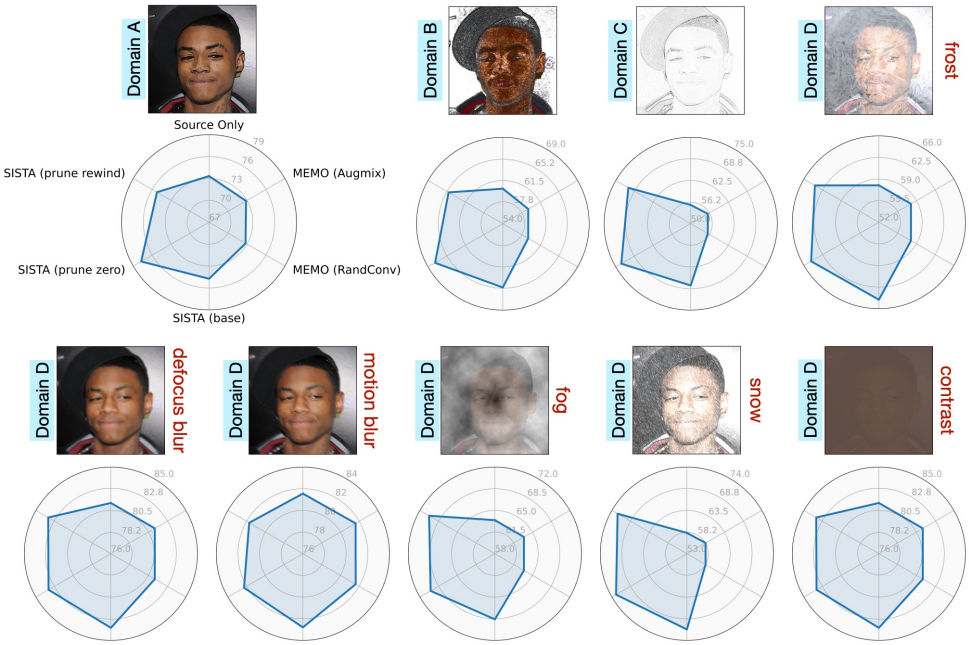


Figure 4. **SiSTA significantly improves generalization of face attribute detectors**. We report the $1-$shot SFDA performance (Accuracy %) averaged across different face attribute detection tasks for different distribution shifts (Domains A, B & C) and a suite of image corruptions (Domain D). SiSTA consistently improves upon the baseline(source-only) and SoTA baseline MEMO in all cases.

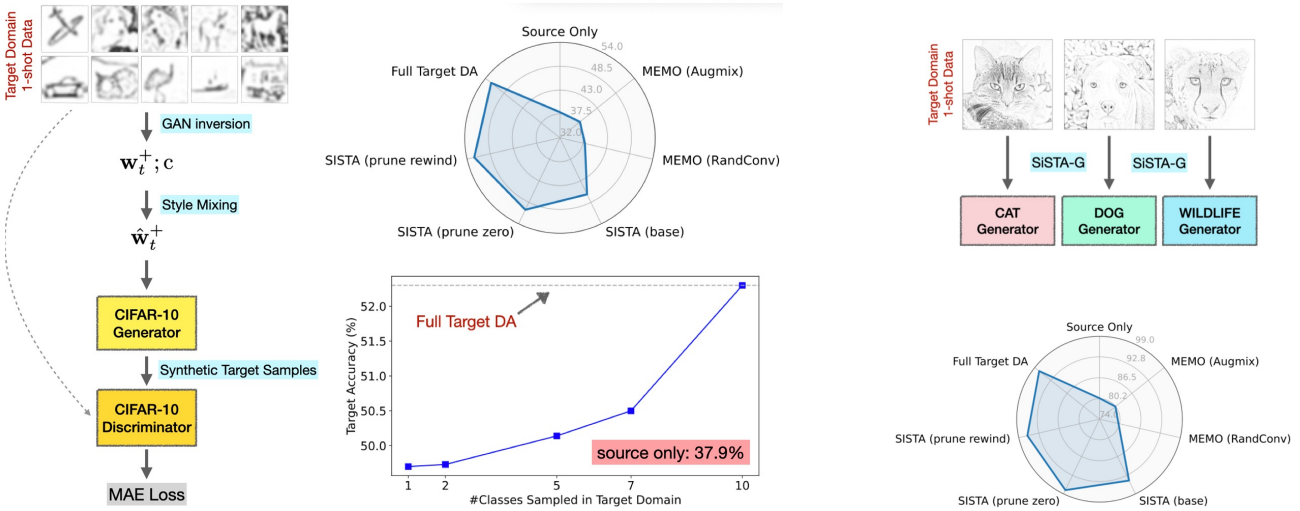
CIFAR-10 we obtained the pre-trained StyleGAN2-ADA models[2] from their respective sources; and (iv) DomainNet (Peng et al., 2019), a large-scale benchmark comprising 6 domains namely Clipart, Painting, Quickdraw, Sketch, Infograph and Real with each domain consisting of images from 340 categories. For this experiment, we used the state-of-the-art StyleGAN-XL model (Sauer et al., 2022) trained on ImageNet (Russakovsky et al., 2015). Note, we used only the subset of categories from DomainNet that directly overlapped with ImageNet classes. To the best of our knowledge, this is the first work to report adaptation performance with a single target image on DomainNet, and to use ImageNet-scale StyleGAN-XL for data augmentation.

Target Domain Design: To emulate a wide-variety of real-world shifts, we employed standard image manipulation techniques (we will release this new benchmark dataset along with our codes) to construct the following target domains: (i) *Domain A*: We used the *Stylization* technique in OpenCV with $\sigma_s = 40$ and $\sigma_r = 0.2$; (ii) *Domain B*: For this shift, we used the *PencilSketch* technique in OpenCV with $\sigma_s = 40$ and $\sigma_r = 0.04$; (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; and (iv) *Domain D*: This shift was created using a different natural image corruptions from ImageNet-C (Hendrycks & Dietterich, 2019) typically used for evaluating model robustness. In particular, we used the *imagecorruptions*[3] package for realizing 6 different shifts, namely *contrast, defocus blur, motion blur, fog, frost and snow*. We report our performance across all the domain shifts for the different attribute detection tasks. Given the inherently challenging nature of Domain C, we used that exclusively to evaluate the multi-class classifiers trained on AFHQ and CIFAR-10 datasets. Finally, for DomainNet evaluations we considered *Real photos* as the source domain and used each of the five remaining domains as the target.

Evaluation methodology: (a) *Source model training*: To obtain the source model F_s we fine-tune an ImageNet pre-trained ResNet-50 (He et al., 2016) with labeled source data. We use a learning rate of $1e-4$, Adam optimizer and train for 30 epochs; (b) *StyleGAN fine-tuning*: We fine-tune G_s for 300 iterations (M in Algorithm 1) using one-target image with learning rate set to $2e-3$ and Adam optimizer with $\beta = 0.99$. These parameters were identified using the CelebA benchmark and we used the same settings for all experiments; (c) *Synthetic data curation*: The size of the synthetic target dataset $\bar{\mathcal{D}}_t$, T, was set to 1000 images in all experiments. Note, in section 4.3, we study the impact of this choice. Another important hyperparameter is the choice of GAN layers for style manipulation: (i) layers $8-18$ in StyleGAN-2; (ii) layers $3-8$ in CIFAR-10 GAN; (iii) layers $10-27$ in StyleGAN-XL.

[2] https://github.com/NVlabs/stylegan2-ada-pytorch

[3] https://github.com/bethgelab/imagecorruptions



(a) CIFAR-10. (b) AFHQ.

Figure 5. **Multi-class classification**: (a)-left illustrates `SiSTA-mcG` with class-conditoned GANs, (a)-right shows the performance of `SiSTA`, while the bottom plot studies the performance of`SiSTA` with exposure to only a subset of classes from the target domain. (b) visualizes our approach for AFHQ dataset where individual class-specific generators are finetuned and bottom plot analyses `SiSTA` along with baselines for this challenging dataset.

This selection was motivated by findings from recent studies on style/content disentanglement in StyleGAN latent spaces (Wu et al., 2021a; Kafri et al., 2021; Karras et al., 2019). (d) *Choice of pruning ratio*: For all experiments, we set $p = 20\%$ for prune-rewind and $p = 50\%$ for prune-zero strategies. Note, in section 4.3, we study the impact of this choice; (e) *SFDA training*: For the NRC algorithm, we set both neighborhood and expanded neighborhood sizes at 5 respectively. Finally, we adapt F_s using SGD with momentum 0.9 and learning rate $1e-3$. All results that we report are computed as an average of 3 independent trials; (f) For evaluation, we report the target accuracy (%) on a held-out test set in each of the target domains.

Baselines: In addition to the vanilla source-only baseline (no adaptation), while there exists a number of test-time adaptation approaches, we perform comparisons to the state-of-the-art online adaptation method, MEMO (Zhang et al., 2021), that enforces prediction consistency between an image and its augmented variants. In particular, we implement MEMO with two popular augmentation strategies namely Augmix and RandConv (Xu et al., 2021). We choose MEMO as the key baseline, since it is already well established that it is superior to other protocols like TENT and TTT. Finally, for comparison, we report the Full Target DA performance as an upper bound, *i.e.*, when the entire target dataset (unlabeled) is used for adaptation.

4.2. Findings

Figure 3 illustrates the synthetic data generated for a target domain (*pencil sketch*) using vanilla sampling (or base), *prune-zero* and (*prune-rewind*) strategies. More examples can be found in the supplement (Figure 8).

`SiSTA` consistently produces superior performance across different distribution shifts.

In Tables 2-10, the performance of `SiSTA` across different domain shifts (A, B, C, D) on the CelebA-HQ dataset is compared to the baselines for all the 19 attributes. Furthermore, Figure 4 summarizes the average performance (across attributes and multiple trials) for the CelebA-HQ dataset. We see that when compared to the source-only baseline and the state-of-the-art MEMO, `SiSTA` yields average improvements of 4.41%, 7.5%, 17.73% and 5.1% respectively for the four target domains. This improvement can be directly attributed to the efficacy of our proposed augmentations, which enable the SFDA method to learn domain-invariant features when adapting the source classifier.

Additionally, utilizing the proposed activation pruning strategies reveal significant gains under severe shifts over the naïve sampling (base). For example, we see an average improvement of 18% across different attributes in Domain C, when compared to the state-of-the-art MEMO. In particular, we notice that for challenging attributes such as *bangs*, *blond hair*, and *gender*, we obtain striking $26.1\%, 29.6\%, 33.9\%$ improvements over the source-only performance. This illustrates how our pruning strategy can create generic representations that aid in an effective adaptation.

Failure cases: While `SiSTA` is generally very effective, there are a few cases where it does not perform as expected. For example, with the Domain B results in Ta-

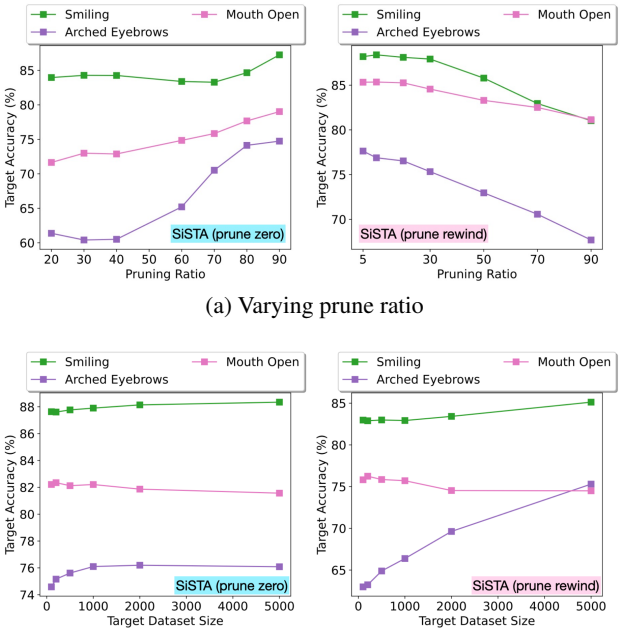
Table 1. Performance of SiSTA on the five different domains of the DomainNet Dataset. SiSTA consistently improves over the Source Only and MEMO baselines even under such complex domain shifts.

	QuickDraw	Painting	ClipArt	InfoGraph	Sketch
Source only	9.23	62.25	58.55	28.45	43.86
MEMO (Augmix)	8.73	62.20	60.15	28.61	43.86
MEMO (RandConv)	8.04	61.91	59.23	28.02	43.52
SiSTA (base)	11.78	63.53	60.98	31.61	47.54
SiSTA (prune-zero)	**13.12**	63.69	60.98	31.65	**48.12**
SiSTA (prune-rewind)	11.86	**64.05**	**61.02**	**31.8**	46.78
Full Target DA	16.27	68.99	69.55	31.77	55.09

ble 3, we notice that for certain attributes (*5'o clock shadow*, *bald*), we fail to improve over the source-only performance (near-random performance), since it becomes challenging to resolve those attributes under that distribution shift. Additionally, in Domain C, we find that the performance of SiSTA (base) is sometimes greater than that of SiSTA (prune zero), likely due to the excessive elimination of style information during pruning. While this can be potentially fixed by adjusting the prune ratio or increasing the number of augmented samples (see 4.3), this reveals some of the failure scenarios for SiSTA.

SiSTA can handle natural image corruption. Natural image corruptions mimic domain shifts that are prevalent in real-world settings. Surprisingly, we find that our proposed SiSTA−S protocol is able to fine-tune the GAN even under such image corruptions and lead to apparent gains in the generalization performance. More specifically, we want the emphasize the two challenging corruptions, namely contrast and fog, where the class discriminative features appear to be muted. Even under these corruptions, as showed in Figure 4, SiSTA achieve average performance improvements of 10.14% and 6.52%, respectively.

SiSTA is effective even with class-conditional GANs. In this experiment, we study how SiSTA performs on CIFAR-10 adaptation, when we are provided with a class-conditional StyleGAN. In this case, we use the SiSTA−mcG procedure to perform GAN fine-tuning, which requires the GAN inversion step to identify both the latent code as well as the conditioning variable. As illustrated in Figure 5a, we use $1-$shot examples from each of the 10 classes and synthesize $T = 1000$ augmentations from SiSTA. Note, during sampling, we draw from the different classes randomly. We find that, for the challenging Domain C target, SiSTA not only outperforms the baselines by a large margin, but also matches the Full Target DA performance, while using only a single-shot example. Furthermore, as argued in Section 3.3, using single-shot examples from even a subset of classes can be beneficial. To demonstrate this, we varied the number of classes from



(a) Varying prune ratio

(b) Varying T

Figure 6. **Analysis** of varying prune ratio p and the amount of synthetic target domain data T used by SiSTA.

which target examples are drawn (1 to 10). We find that, even with a single class example, SiSTA provides a large gain of 12.69% over the source-only baseline. As expected, the generalization performance consistently improves as we expose the model to examples from additional classes.

SiSTA can also be used with multiple class-specific GANs. In this study, we examined the performance of SiSTA in a multi-class classification problem with AFHQ, where we assume access to individual generative models for each class. Given the inherent diversity within classes (different breeds of cats or dogs), it is sometimes challenging to train a single StyleGAN for the entire data distribution. In such cases, a separate generative model can be trained on source images from each of the classes. However, the classifier is trained for a $3-$way classification setting. In this case, we perform SiSTA for each GAN independently using its corresponding example. As shown in Figure 5b, we find that, even our base variant achieves 94.53%, outperforming the source-only and baselines by large margins (14%). Our best performance is achieved by *prune-zero* in this setting and it matches Full Target DA.

Even on large scale benchmarks such as DomainNet, SiSTA provides consistent benefits. To study its performance on large-scale benchmarks, we tested SiSTA on DomainNet that comprises a large number of object types and complex distribution shifts (photo, quickdraw, painting, etc.). Given the diversity of objects in this benchmark, we utilized the state-of-the-art StyleGAN-XL model trained on ImageNet to perform SiSTA and studied the single-shot

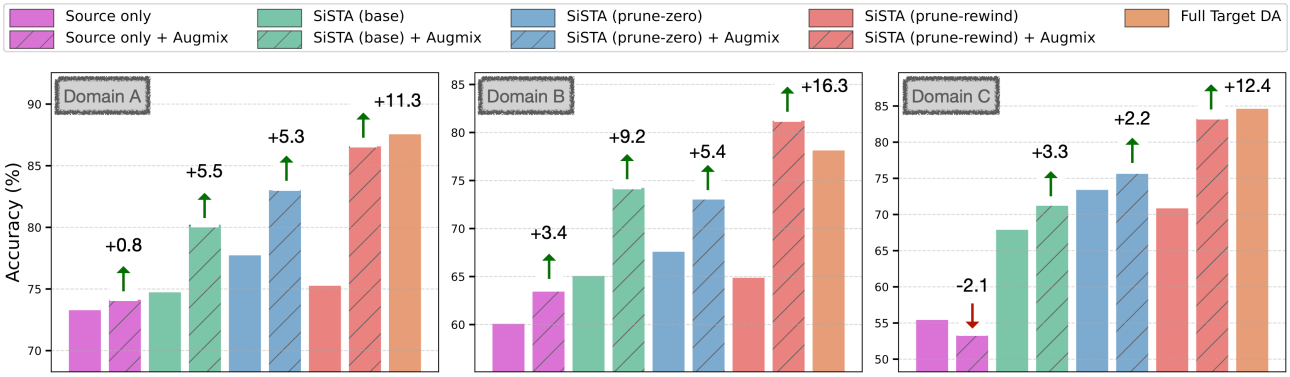


Figure 7. **Effect of Toolbox augmentations on `SiSTA`.** We present the performance of `SiSTA` on Domains A, B, and C of the CelebaHQ dataset when images generated by `SiSTA` are further enhanced with Augmix (Hendrycks et al., 2020). We observe that toolbox augmentations can further improve the performance of `SiSTA`, and in a few cases, `SiSTA` even surpasses the Full Target DA baseline.

adaptation performance for different target domains (real is the source domain). From Table 1, we find that even on this benchmark, `SiSTA` (prune-zero) convincingly improves upon source only baselines. For example, `SiSTA` provides about 4% improvements for Quickdraw and Sketch domains. As with the other benchmarks, `SiSTA` is indeed competitive to the Full Target DA baseline.

4.3. Analysis of parameter choices

The choice of prune ratio p. We investigate the effect of the choice for p in *prune-zero* and *prune-rewind* using three face attribute detectors (Figure 6a). This parameter influences the degree of generalizability of the synthetic target representations. For *prune-zero*, higher pruning ratios (severe style attenuation), *i.e.*, p between $80 - 90$, are found to significantly enhance performance when compared to lower ones. In the case of *prune-rewind*, on the other hand, p regulates the amount of source mix-up with the target domain. In this scenario, we see that a smaller p performs better, and we recommend to set p between $5 - 20$.

The choice of synthetic data size T. We study the influence of the number of augmentations T by varying it between $100 - 5000$ and studying the performance of *prune-zero* and *prune-rewind* on three attributes, as illustrated in Figure 6b. While *prune-zero* performs consistently for different values of T, it only makes limited gains on average as the number of samples increases. On the contrary, we see a significant boost in performance in *prune-rewind* in some of attributes. We remark that *prune-rewind* is a sensitive technique due to the mix-up with the source domain; increasing the number of the synthetic augmentations (along with low p) stabilizes the performance and, in a few cases, even matches the performance of *prune-zero*. Finally, we note that the performance variation across the independent trials is around $< 0.5\%$, thus indicating that the performance is consistent and not sensitive to the sampling process.

Toolbox augmentations can further bolster `SiSTA`. In this study, we investigated the benefits of using sophisticated toolbox augmentations such as Augmix for `SiSTA` as well as for the source only baseline. From Figure 7, we observe a consistent boost in performance for all the three variants of `SiSTA`with average improvements of 6%, 4.2% and almost 13.3% respectively. These results highlight the effective complementary nature of `SiSTA` to toolbox augmentations. Furthermore, it is worth noting that applying Augmix to the source-only methods does not lead to the same level of improvements. This observation is consistent with the findings from (Thopalli et al., 2022), which noted that toolbox augmentations alone are insufficient to enhance adaptation performance under real-world distribution shifts.

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. We showed that the proposed approach is effective in multi-class classification using both class-conditioned as well as multiple class-specific GANs. Our future work includes theoretically understanding the behavior of different pruning techniques and extending our approach beyond classifier adaptation.

Acknowledgements

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. Supported by the LDRD Program under project 22-ERD-006. LLNL-CONF-844756.

References

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

Chong, M. J. and Forsyth, D. Jojogan: One shot face stylization. *arXiv preprint arXiv:2112.11641*, 2021.

Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., and Le, Q. V. Autoaugment: Learning augmentation strategies from data. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 113–123, 2019. doi: 10.1109/CVPR.2019.00020.

DeVries, T. and Taylor, G. W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

Gokhale, T., Anirudh, R., Thiagarajan, J. J., Kailkhura, B., Baral, C., and Yang, Y. Improving diversity with adversarially learned transformations for domain generalization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 434–443, 2023.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Hendrycks, D. and Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. *Proceedings of the International Conference on Learning Representations*, 2019.

Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. Augmix: A simple data processing method to improve robustness and uncertainty. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=S1gmrxHFvB.

Hendrycks, D. 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*, pp. 8340–8349, 2021.

Hoffman, J. et al. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pp. 1989–1998. Pmlr, 2018.

Huang, J., Guan, D., Xiao, A., and Lu, S. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34:3635–3649, 2021.

Ishii, M. and Sugiyama, M. Source-free domain adaptation via distributional alignment by matching batch normalization statistics. *arXiv preprint arXiv:2101.10842*, 2021.

Kafri, O., Patashnik, O., Alaluf, Y., and Cohen-Or, D. Stylefusion: A generative model for disentangling spatial segments. *arXiv preprint arXiv:2107.07437*, 2021.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *CoRR*, abs/1710.10196, 2017. URL http://arxiv.org/abs/1710.10196.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8110–8119, 2020.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Kundu, J. N., Venkat, N., Revanur, A., Babu, R. V., et al. Towards inheritable models for open-set domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12376–12385, 2020.

Liang, J., Hu, D., and Feng, J. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*.

Liu, X. and Yuan, Y. A source-free domain adaptive polyp detection framework with style diversification flow. *IEEE Transactions on Medical Imaging*, 41(7):1897–1908, 2022.

Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., and Lischinski, D. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2085–2094, 2021.

Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., and Wang, B. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1406–1415, 2019.

Rahman, A., Rahman, M. S., and Mahdy, M. R. C. 3c-gan: class-consistent cyclegan for malaria domain adaptation model. *Biomedical Physics & Engineering Express*, 7, 2021.

Richardson, E. 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*, pp. 2287–2296, 2021.

Robey, A., Pappas, G. J., and Hassani, H. Model-based domain generalization. *Advances in Neural Information Processing Systems*, 34:20210–20229, 2021.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022.

Sankaranarayanan, S., Balaji, Y., Castillo, C. D., and Chellappa, R. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8503–8512, 2018.

Sauer, A., Schwarz, K., and Geiger, A. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 conference proceedings*, pp. 1–10, 2022.

Steiner, A., Kolesnikov, A., Zhai, X., Wightman, R., Uszkoreit, J., and Beyer, L. How to train your vit? data, augmentation, and regularization in vision transformers. *arXiv preprint arXiv:2106.10270*, 2021.

Subramanyam, R., Narayanaswamy, V., Naufel, M., Spanias, A., and Thiagarajan, J. J. Improved stylegan-v2 based inversion for out-of-distribution images. In *International Conference on Machine Learning*, pp. 20625–20639. PMLR, 2022.

Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., and Hardt, M. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pp. 9229–9248. PMLR, 2020.

Tang, S., Yang, Y., Ma, Z., Hendrich, N., Zeng, F., Ge, S. S., Zhang, C., and Zhang, J. Nearest neighborhood-based deep clustering for source data-absent unsupervised domain adaptation. *arXiv preprint arXiv:2107.12585*, 2021.

Thopalli, K., Turaga, P., and Thiagarajan, J. J. Domain alignment meets fully test-time adaptation. In *Asian Conference on Machine Learning, 2022.*, 2022.

Torralba, A. and Efros, A. A. Unbiased look at dataset bias. In *CVPR 2011*, pp. 1521–1528. IEEE, 2011.

Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O., and Cohen-Or, D. Designing an encoder for stylegan image manipulation. *ACM Transactions on Graphics (TOG)*, 40(4):1–14, 2021.

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., and Darrell, T. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

Wu, Z., Lischinski, D., and Shechtman, E. Stylespace analysis: Disentangled controls for stylegan image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12863–12872, 2021a.

Wu, Z., Nitzan, Y., Shechtman, E., and Lischinski, D. Stylealign: Analysis and applications of aligned stylegan models. *arXiv preprint arXiv:2110.11323*, 2021b.

Xia, W., Zhang, Y., Yang, Y., Xue, J.-H., Zhou, B., and Yang, M.-H. Gan inversion: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Xu, Z., Liu, D., Yang, J., Raffel, C., and Niethammer, M. Robust and generalizable visual representation learning via random convolutions. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=BVSM0x3EDK6.

Yang, S., van de Weijer, J., Herranz, L., Jui, S., et al. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 34:29393–29405, 2021.

Yue, F., Zhang, C., Yuan, M., Xu, C., and Song, Y. Survey of image augmentation based on generative adversarial network. *Journal of Physics: Conference Series*, 2203 (1):012052, feb 2022. doi: 10.1088/1742-6596/2203/1/012052. URL https://dx.doi.org/10.1088/1742-6596/2203/1/012052.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6023–6032, 2019.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

Zhang, M., Levine, S., and Finn, C. Memo: Test time robustness via adaptation and augmentation. *arXiv preprint arXiv:2110.09506*, 2021.

A. Examples of augmentations from SiSTA

In Figure 8, we show the augmentations synthesized by SiSTA for different domain shifts and StyleGAN models.

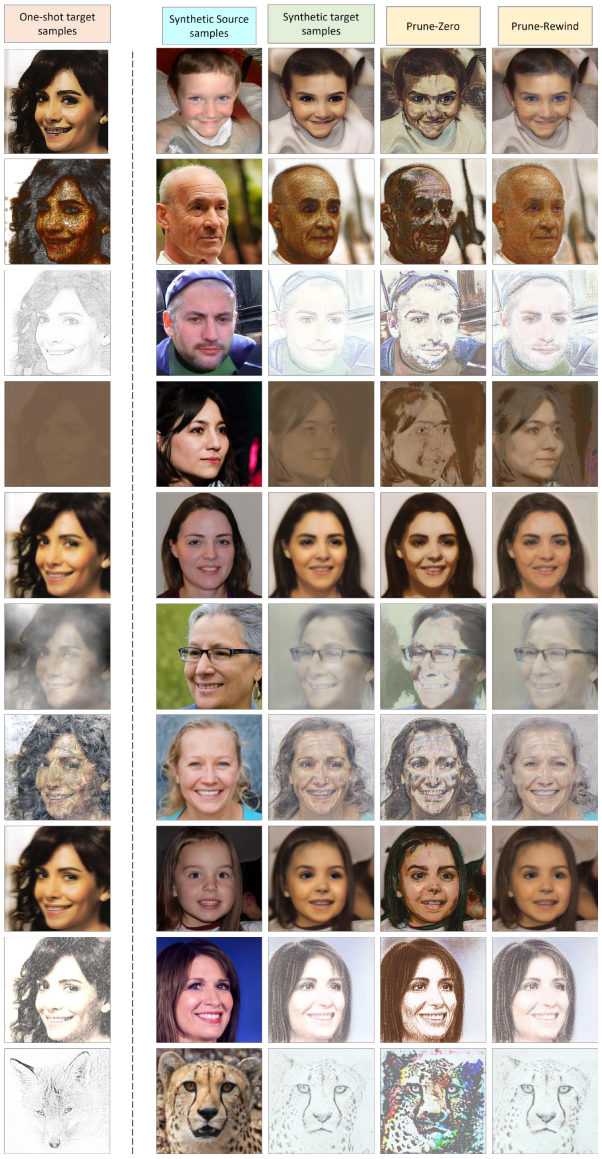


Figure 8. **SiSTA generated augmentations** on random samples drawn from the style space of StyleGAN; The rows 1 to 9 correspond to different domain shifts in CelebA-HQ and row 10 corresponds to AFHQ.

B. Detailed results for our CelebA experiments

We provide comprehensive tables for the results discussed in Section 4. Tables 2-10 illustrate the performance of source-only, MEMO, and all the three variants of SiSTA along with Full target performance.

	5' o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	53.7	69.9	63.4	83.2	55.3	**89.5**	80.7	80.4	93.2	88.2	58.1	**82**	60.2	**89.5**	53.4	68.3	70.9	88.5	64.8
MEMO (Augmix)	53.6	69.9	64.5	81.1	53.8	89.1	79.7	78.6	93.8	87.6	57.9	80.8	59.6	89.4	52.5	70.5	68.6	88.1	65.1
MEMO (Randconv)	53.7	69.6	64.5	81	53.7	89.1	79.5	78.4	93.9	87.6	57.9	80.8	59.5	89.3	52.5	70.2	68.6	88	65
SiSTA (base)	52.8	74.6	**77**	80	85.2	69.8	87.2	72.8	95.1	91.2	55.2	69.8	58.3	84.4	57	79.1	71.3	**90.1**	**69.1**
SiSTA (prune-zero)	**55.2**	**78.2**	76.3	**87.1**	**87.6**	81.5	**88.1**	**81.2**	**95.5**	**91.7**	**60.4**	70.8	**61.1**	89.2	**59.3**	**79.5**	**76.2**	89.6	68.6
SiSTA (prune-rewind)	53.1	76.6	70.1	85.6	83	78.2	87.1	76	95.2	91.6	57.8	67.5	58.5	87.3	59.2	78.6	74.2	89.3	60.6
Full target DA	87	81.9	92.3	93.5	90.1	97.3	89.3	87.1	97.4	92.7	72.5	91.5	93	92.6	74.5	80.6	82.5	92.3	75.2

Table 2. Performance of `SiSTA` on Domain A of the CelebA dataset.

	5' o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	50	51	50.5	67.2	50	74.2	54.2	54.6	80.2	78.6	52.1	**63.9**	**54**	76.9	50.1	65	50.4	63.3	55.5
MEMO (Augmix)	50	51.2	50.5	64.5	50	74.1	52.1	52.4	81.1	79	51.2	63	50.8	73.2	50	65.5	50.2	58.6	55.6
MEMO (Randconv)	50	51.2	50.5	64.5	50	73.9	52.1	52.3	81.2	79	51.2	62.9	50.8	73.1	50	65.6	50.2	58.5	55.7
SiSTA (base)	50	73	50.2	83.3	50.5	67.8	77.6	56.3	86.5	82.5	56.7	56.1	50.1	77	51.7	72.6	**56.3**	**80**	58.1
SiSTA (prune-zero)	50.1	**73.9**	51.1	**86.7**	51.4	**75.8**	**79.9**	**67.2**	**88.7**	**84.4**	**58.3**	58.1	50.2	**85.4**	**53.8**	**74**	54.8	79.8	**60.5**
SiSTA (prune-rewind)	50	73.4	50	84.7	50.2	75.2	75.5	57.1	85.9	82.9	54	54.5	50.1	78	52.7	72.8	**56.3**	73	56.3
Full target DA	71.6	71.7	72.6	89.9	58.4	94.2	81.9	78.5	92.2	88	63.9	84.3	83	88.4	68.6	71	68.6	86.7	71.2

Table 3. Performance of `SiSTA` on Domain B of the CelebA dataset.

	5' o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	50	52.8	50.1	58.2	50.5	63.8	56.5	50.2	58.3	58.9	50	51.3	50.5	64	52	59.9	51.8	71.6	52.7
MEMO (Augmix)	50	53.6	50.2	61.6	50.5	66.6	55.5	50.1	56.1	60.4	50	50.8	50.4	65.8	52	59.2	51.7	72.3	52
MEMO (Randconv)	50	53.6	50.2	61.6	50.5	66.6	55.4	50.1	56	60.4	50	50.7	50.4	65.5	52	59.1	51.7	72.4	52
SiSTA (base)	53.2	65.3	**64.7**	80	77.9	69.4	54.5	71.2	91.8	71.4	**59.1**	66.6	53.2	79.2	54.7	77.3	57.8	78.8	63.7
SiSTA (prune-zero)	**58**	**74.7**	64.1	82.6	77.1	**82.7**	**80.7**	**77.2**	88.3	**78.2**	56.3	**68.2**	**55.3**	**86.7**	**68.5**	74.3	**62.8**	**86.5**	67.6
SiSTA (prune-rewind)	53.1	69.7	63.5	**84.3**	**80.1**	79.9	62.1	69.7	**92.2**	**78.2**	54.4	65	53.7	84.4	57.3	**78.5**	58.2	**86.5**	**74.5**
Full target DA	83.1	80.5	92	93	84.2	96.7	83.8	80.8	95.7	87.6	66.9	90	93.2	89.2	69.9	77.5	76.6	89.5	77.5

Table 4. Performance of `SiSTA` on Domain C of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	64.5	79	82.5	90	87.4	91	90.4	87.8	97.2	92	64.5	79.7	63.4	93	68.8	79.9	65.7	92.6	74.9
MEMO (Augmix)	63.2	78.1	87.5	88	87.1	**91.3**	90.6	**89.8**	**97.8**	90.8	65.3	77.4	62	**92.9**	70.6	80.8	63.9	91	75.5
MEMO (Randconv)	63.2	78.1	87.5	87.5	87.1	**91.3**	90.6	**89.8**	**97.8**	90.8	65.3	77.4	62	**92.9**	70.6	81	63.7	91	75.3
SiSTA (base)	**85.6**	80	**88.9**	88.9	91.2	76.9	89.8	79	95.3	91.5	**65.6**	**91.4**	**89.3**	87.5	65.2	**82.4**	68.2	91.9	**81.9**
SiSTA (prune-zero)	85.1	79.5	85.1	90.3	**92.8**	83.3	**90.7**	82.4	96.4	90.7	63.8	89.7	76.7	89.9	**73.7**	81.5	69.3	**92.2**	73.3
SiSTA (prune-rewind)	78.2	**81.5**	85.3	**92.3**	92.5	83.4	90.5	81.7	97.2	**92.7**	64.2	87.7	77.3	90.7	71.1	82.1	**71.1**	**92.2**	75.5
Full target DA	89.4	83	96.1	94	92.9	97.1	90.7	88	97.8	93.7	74.4	93.3	94.1	93.3	76.9	82.4	84.5	92.6	83.1

Table 5. Performance of `SiSTA` on Domain D (Defocus blur) of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	71.4	**79.7**	79.5	88.3	88.9	91.5	89.7	87.1	97.6	91.6	69.6	80.5	65.7	92.9	72.5	73.9	62.2	92.2	74.8
MEMO (Augmix)	73	78.6	73.7	88.3	88.8	**91.8**	91.9	**88.5**	**97.5**	92	**70.7**	80.8	63	**93.1**	**73.5**	75	62.2	92.6	75.5
MEMO (Randconv)	73	78.6	73.7	88.3	88.8	**91.8**	**92**	**88.5**	**97.5**	92.1	**70.7**	80.8	63	**93.1**	**73.5**	75	62.2	**92.7**	75.5
SiSTA (base)	**79.8**	74.7	**89.8**	89.3	**93.6**	78.2	89.6	79.5	94.4	92.2	67.4	**87.8**	**73.1**	87.9	69.7	**81.5**	**71**	92	**82**
SiSTA (prune-zero)	74	75.4	87.1	92.1	**93.6**	86.9	90.6	83.7	96.5	91.4	66.3	78.6	63	90.8	72.9	81.2	70.9	92.4	76.3
SiSTA (prune-rewind)	70.7	76.1	85.9	**92.5**	**93.6**	85.5	90	81.2	96.2	**92.8**	65.9	79.7	64.9	89.9	72.5	80.4	68.9	92.2	73.9
Full target DA	90.1	82.8	96.7	93.8	93.2	98.1	90.8	88.2	97.9	93.7	72	94.9	94.6	93.2	75.7	82.6	85.4	92.9	84.2

Table 6. Performance of `SiSTA` on Domain D (Motion blur) of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	59.5	52.5	**71.9**	59.9	51.9	88.1	50	57.6	78.3	79.6	50.6	82	63	77.2	**53.8**	63.6	52.5	50.6	62.5
MEMO (Augmix)	**62.6**	52.5	60.9	61.4	52.4	83	50	57.9	78.2	78.9	50.6	81.3	**64.1**	77.1	52	62.3	53.1	50.5	61.1
MEMO (Randconv)	**62.6**	52.4	60.9	61.1	52.4	83	50	57.9	78.3	78.9	50.6	81.3	63.5	77	51.6	62.3	53.1	50.5	61.1
SiSTA (base)	57.3	56.3	**71.9**	77.9	57.4	80.6	60.7	68.2	75.2	84.2	**57.1**	**84.9**	63	76.8	51.8	74.8	62.3	73.5	69.6
SiSTA (prune-zero)	54.1	57.3	70.8	80.6	**58.9**	**89**	63.6	**77**	**81.1**	82.8	56.4	73.9	55.4	**85.3**	53.7	**76.2**	**63.8**	78.2	**71.7**
SiSTA (prune-rewind)	54.3	**58.5**	68.8	**84.3**	53.6	87.3	**69.4**	75.9	78.4	**85.8**	56.1	80.9	60	81.5	52.1	74.8	62.2	**80.8**	70.6
Full target DA	86.9	78.8	80	90.6	90	97.8	85.9	82.9	94.6	92.9	72.6	92.6	92.5	89	70.6	78.3	84.2	89.6	76

Table 7. Performance of `SiSTA` on Domain D (Fog) of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	51.1	51.5	54.6	55.8	51.3	70.5	50.1	53.5	75.5	72.9	50.8	68.6	56.3	66.7	50.2	64.6	51.7	51.5	54.7
MEMO (Augmix)	50.3	51.4	55.6	55.8	51.4	**72**	50.2	53.9	75.9	74.3	51.1	68.3	56.4	67.2	50	64.3	51.2	51.1	53.9
MEMO (Randconv)	50.3	51.4	55.6	55.8	51.4	71	50.2	53.9	75.9	74.4	51.1	68.4	56.4	67.2	50	64	51.2	51.1	54
SiSTA (base)	51	65.2	58.7	60.4	59.8	62.2	76	58.5	**82.6**	79.5	**57.8**	**69.2**	51.9	74.4	54.8	**66.4**	**62.1**	**77.9**	**58.2**
SiSTA (prune-zero)	50.5	**66.3**	**59.3**	59.4	**70.9**	65.3	75.6	**66.8**	80.4	76.4	57.7	64.1	50.3	**78.7**	**56.3**	58.7	61.1	73.6	57.1
SiSTA (prune-rewind)	50.3	65.6	55.6	**61.2**	61	65.4	**76.5**	60.4	82.3	**80.2**	56.2	64.6	50.4	76	54.9	61.7	61.8	77	53.9
Full target DA	67.6	68.8	65.5	75.6	75.9	90.7	78.6	76.7	88	89.3	61.9	74.2	61.2	86.2	61.3	60.2	67.2	84.9	62.2

Table 8. Performance of `SiSTA` on Domain D (Frost) of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	50.1	52.5	50.9	59.6	50.2	73.9	51.9	50.1	76.9	66.4	50.1	70.6	57.7	62.9	50.8	61.6	51.2	54	61.8
MEMO (Augmix)	50	52.7	50	60	50	73.5	51.5	50	77	69.6	50	70.2	58.7	64.4	50.7	61	51.3	54.3	61.7
MEMO (Randconv)	50	52.7	50	60	50	73.5	51.3	50	77	69.7	50	70	58.8	64.7	50.8	60.9	51.3	54.3	61.6
SiSTA (base)	62	64.2	60.9	67.2	79.1	82.6	79.4	65	83.4	77.4	**68.7**	**82.1**	**60.3**	75.4	54.7	75.6	**67.1**	84.3	66.6
SiSTA (prune-zero)	**62.4**	63.2	**61.4**	70.7	**87.2**	**87.8**	79.5	**68.8**	86.2	75.7	67.8	81.4	59.7	79.4	**57.9**	**76.1**	64.9	86.4	**67.4**
SiSTA (prune-rewind)	59.4	**67.5**	57.5	**73.6**	79.4	86.2	**83.3**	65.9	**86.7**	**79.3**	66.7	81.2	58.7	**79.7**	55.8	74.9	67	**87.1**	65.9
Full target DA	76.32	77.68	66.79	82.68	85.69	94.96	84.32	77.87	89.45	83	70.09	85.44	83.71	85.55	62.11	72.93	79.14	84.89	65.89

Table 9. Performance of `SiSTA` on Domain D (Snow) of the CelebA dataset.

	5'o clock shadow	Arched eyebrows	Bald	Bangs	Blond hair	Eyeglasses	Makeup	Cheekbones	Gender	Mouth open	Eyes closed	Beard	Sideburns	Smiling	Straight hair	Wavy hair	Earrings	Lipstick	Young
Source only	50	50	50.4	53	53.4	51.5	50	51.2	69.7	54.5	50	58.9	50	59.4	50.5	50.8	50	56.5	61.6
MEMO (Augmix)	50	50	52.6	51.8	51.9	52.1	50	51.1	68.9	54.5	50	**58.8**	50	58.3	49.9	50.6	50	55.9	57.3
MEMO (Randconv)	50	50	52.6	51.8	51.4	52.1	50	51.1	69	54.6	50	**58.8**	50	58.3	49.9	50.6	50	55.7	58.1
SiSTA (base)	50	60.1	54.1	70.3	66.7	50.3	72	65.5	**83.5**	75.3	50.8	52.5	50.6	74.4	51.7	**70.6**	51.2	**77.3**	**62.5**
SiSTA (prune-zero)	50	**65.7**	**58.7**	**76.4**	**75.6**	51.1	**80.1**	**73.7**	74.2	73.2	50.3	51.1	50.2	**82.9**	54.9	67.1	52.2	72	57.8
SiSTA (prune-rewind)	50	63.4	55	72.8	72	51	76	67	81.6	**76.9**	50.5	52.4	50.2	78.5	**55**	69.8	50.5	76.4	61.5
Full target DA	63.2	76.7	65.56	69.31	76.7	92.1	76	76.3	86.4	89.2	58.8	73.1	71.8	87.1	57.1	68.5	73.7	80.5	62.3

Table 10. Performance of `SiSTA` on Domain D (Contrast) of the CelebA dataset.