Efficient Conditional GAN Transfer with Knowledge Propagation across Classes

Supplementary Material

Mohamad Shahbazi[1], Zhiwu Huang[1], Danda Pani Paudel[1], Ajad Chhatkuli[1], Luc Van Gool[1,2]
[1]Computer Vision Lab, ETH Zürich, Switzerland [2]PSI, KU Leuven, Belgium
{mshahbazi, zhiwu.huang, paudel, ajad.chhatkuli, vangool}@vision.ee.ethz.ch

A. Overview

First, we provide more details on the model architecture and implementation of our experimental setup. Then, we further discuss the quantitative results of CIFAR100 [3] experiments. Moreover, we discuss source-to-target similarities, training curves, and single-class target data. Finally, we provide more visual results obtained using our proposed method.

B. Additional Implementation Details

Architecture: In the main paper, we follow [4, 6] to employ the architecture of BigGAN [1] as our backbone for the GAN transfer tasks. It is worth mentioning that, the BigGAN implementation for CIFAR uses the basic form of BigGAN, which for example, does not use a hierarchical latent variable. In particular, Tables 1 and 2 show the network architecture for CIFAR100 and ImageNet [2] setups. Fig. 1 shows the architecture of the residual blocks used in the generator and the discriminator. The detailed diagram for the conditional batch normalization layer with knowledge propagation across classes has been provided in Fig. 4 of the main paper.

$z \in \mathbb{R}^{128} \sim \mathcal{N}(0, I)$
Linear $4 \times 4 \times 256$
ResBlock up 256
ResBlock up 256
ResBlock up 256
BN, ReLU, Conv 3×3, Tanh

$x \in \mathbb{R}^{32 \times 32 \times 3}$
ResBlock down 64
ResBlock down 128
ResBlock down 256
ResBlock down 512
ResBlock 1024
ReLU, Global Sum Pooling
Embed(y).h + (Linear$\rightarrow$1)

Table 1: The network architecture for CIFAR setup: Left: the generator. Right: the discriminator.

$z \in \mathbb{R}^{120} \sim \mathcal{N}(0, I)$
Linear $4 \times 4 \times 256$
ResBlock up 256
ResBlock up 256
ResBlock up 256
BN, ReLU, Conv 3×3, Tanh

$x \in \mathbb{R}^{128 \times 128 \times 3}$
ResBlock down 96
None-Local Block (64×64)
ResBlock down 192
ResBlock down 384
ResBlock down 768
ResBlock down 1536
ResBlock 1536
ReLU, Global Sum Pooling
Embed(y).h + (Linear$\rightarrow$1)

Table 2: The network architecture for ImageNet setup: Left: the generator. Right: the discriminator.

Baselines: The original study proposing batch normalization adaptation (BSA) [4] uses supervised loss function (L1/Perceptual loss) instead of adversarial loss. In our experiments, whenever we refer to BSA, we mean training the GAN model adversarially, while freezing the filters and learning the BN parameters from the scratch. Therefore our implementation of BSA could be considered as the main baseline for our experiments since it shares the same setup as ours, but without performing any knowledge transfer across the classes.

Ablation Study Experiments: Table 2 in the main paper shows the results for the ablation study on the ImageNet-to-Places365 setup. In the table, "Prior" refers to the BN parameters of the previous classes. The table includes the results for using the prior with and without being further updated via knowledge sharing using target classes. "Shared W" in the fourth experiment refers to using shared similarity wights over all layers of the generator to combine previous BN parameters of each layer. The term "w/o reg" in the fifth experiment refers to not using l1 regularization on the combination weights and l2 regularization on the residuals.

C. Further Discussion on Quantitative Results

In Table 1 of the main paper, which shows the FID scores for different experiments on CIFAR100, the results of the first experiment (20 classes, 600 samples per class) is marginally different from those of the other experiments. It can be seen that, learning from the scratch performs better than all of the transfer learning methods, since the training data is large enough for learning the filters from scratch.

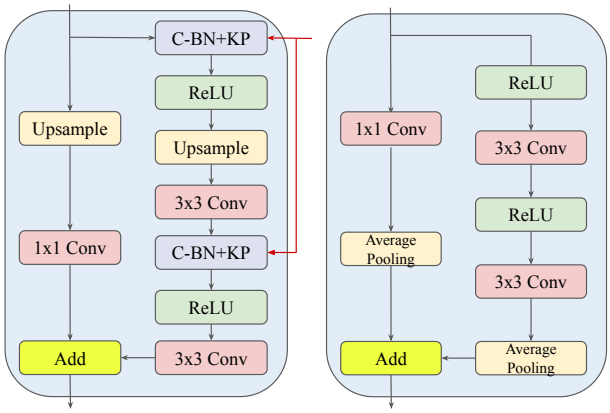

Figure 1: The architecture of the residual blocks used in the network. Left: the generator's ResBlock ("C-BN+KP" layer indicates the conditional batch normalization with knowledge propagation across classes. See Fig. 4 of the main paper for more details). Right: the discriminator's ResBlock.

However, by reducing the sample number in the next experiments, the performance of learning from scratch immediately deteriorates, while the transfer learning methods remain more robust. However, after reducing the training data even more (20/100, 10/600, 10/300, 10/100), fine-tuning (TransferGAN [7]) also degrades significantly compared to BSA and our method, and it falls into mode collapse. Comparing the FID scores of BSA and our method on CIFAR, although comparable, it can be observed that our method starts to perform better in the experiments with less amount of data, showing the importance of using prior knowledge from previous classes when training data is small.

D. Discussion on Class Similarities

As explained in the main paper, our method proposes knowledge transfer from previous classes by learning similarity scores over previous BN parameters and combing the BN parameters using those scores to construct the BN parameters of the new classes. Previous classes can contribute to a target class in terms of semantics, shape, texture, or color in a hierarchical manner from bottom to top layers. As an example, Fig. 2 shows the top 3 ImageNet classes of the pre-trained network (planetarium, Bird House, Mountain Tent) contributing to the target class "Arch" in Places365, based on the similarity weights learned for the first layer (the first layer is generally more interpretable in terms of class similarities, since it is responsible for determining the general structure of the output images, as shown in Fig. 9 in the main paper). As it can be seen, these classes contain visual features close to arch structures that can meaningfully be used to generate images from the target class. Fig. 3 shows another example on Animal Face dataset [5] by vi-

Figure 2: Top 3 contributing classes (planetarium, Bird House, Mountain Tent) from the pre-trained network toward the target class "Arch" in places365 dataset [8]. The classes are selected based on the learned similarity scores of the first layer. Each row depicts one class, and the images are generated from the network pre-trained on ImageNet.

Figure 3: Top 3 contributing classes (Buckeye, Football Helmet, Impala) from the pre-trained network toward the target class "Deer" in Animal Face dataset. The classes are selected based on the learned similarity scores of the first layer. Each row depicts one class, and the images are generated from the network pre-trained on ImageNet.

sualizing the top 3 classes (Buckeye, Football Helmet, Impala) contributing to the target class "Deer". In this example, we can see that the third class "Impala" is semantically very close to the target class. However, the contribution of the first two classes is not as clear as the previous example. These classes might be contributing to the background, or this might be due to the fact that the similarity scores are actually learned to combine the pseudo-classes, and this does not always guarantee semantic similarity to the initial pre-training classes.

E. FID and Loss Curves

As an example of how the losses and the FID scores evolve during the training, the curves for on of the CIFAR100 experiments (Exp. 10/600) have been provided in Fig. 4. The convergence speed-up is clearly depicted in the FID curve, whereas the same is difficult to be derived from the loss plots (due to the adversarial training).

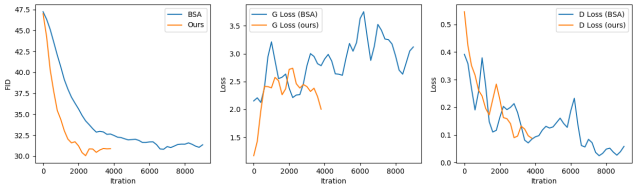

Figure 4: The FID curve (left), the G loss (middle), and the D loss (right) for our method and BSA on CIFAR(10/600).

F. Single-class Target

Although the main focus of our work is multi-class to multi-class knowledge transfer using knowledge propagation and knowledge sharing, the proposed method is not only limited to the multi-class target. As an example, the results of knowledge transfer to the single class "Arch" in Places365 are provided in Table 3.

Method	FID	Iterations
BSA	104	4300
Ours	**78**	**500**

Table 3: FID scores and number of iterations for knowledge transfer from ImageNet to the single target class "Arch" in Places365.

G. Additional Visual Results

Fig. 5 and Fig. 6 show additional visual result obtained from BSA (no knowledge propagation across classes) and our method on ImageNet setup. Regarding CIFAR setup, we visualize the results of the experiments 20/300 and 10/300 for BSA and our method in Fig. 7 and Fig. 8, as examples of CIFAR experiments.

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.

[3] Alex Krizhevsky. Learning multiple layers of features from tiny images. *Tech Report*, 2009.

[4] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2750–2758, 2019.

[5] Zhangzhang Si and Song-Chun Zhu. Learning hybrid image templates (hit) by information projection. *IEEE Transactions on pattern analysis and machine intelligence*, 34(7):1354–1367, 2011.

[6] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: effective knowledge transfer from gans to target domains with few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9332–9341, 2020.

[7] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data. In *Proceedings of the European Conference on Computer Vision*, pages 218–234, 2018.

[8] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.

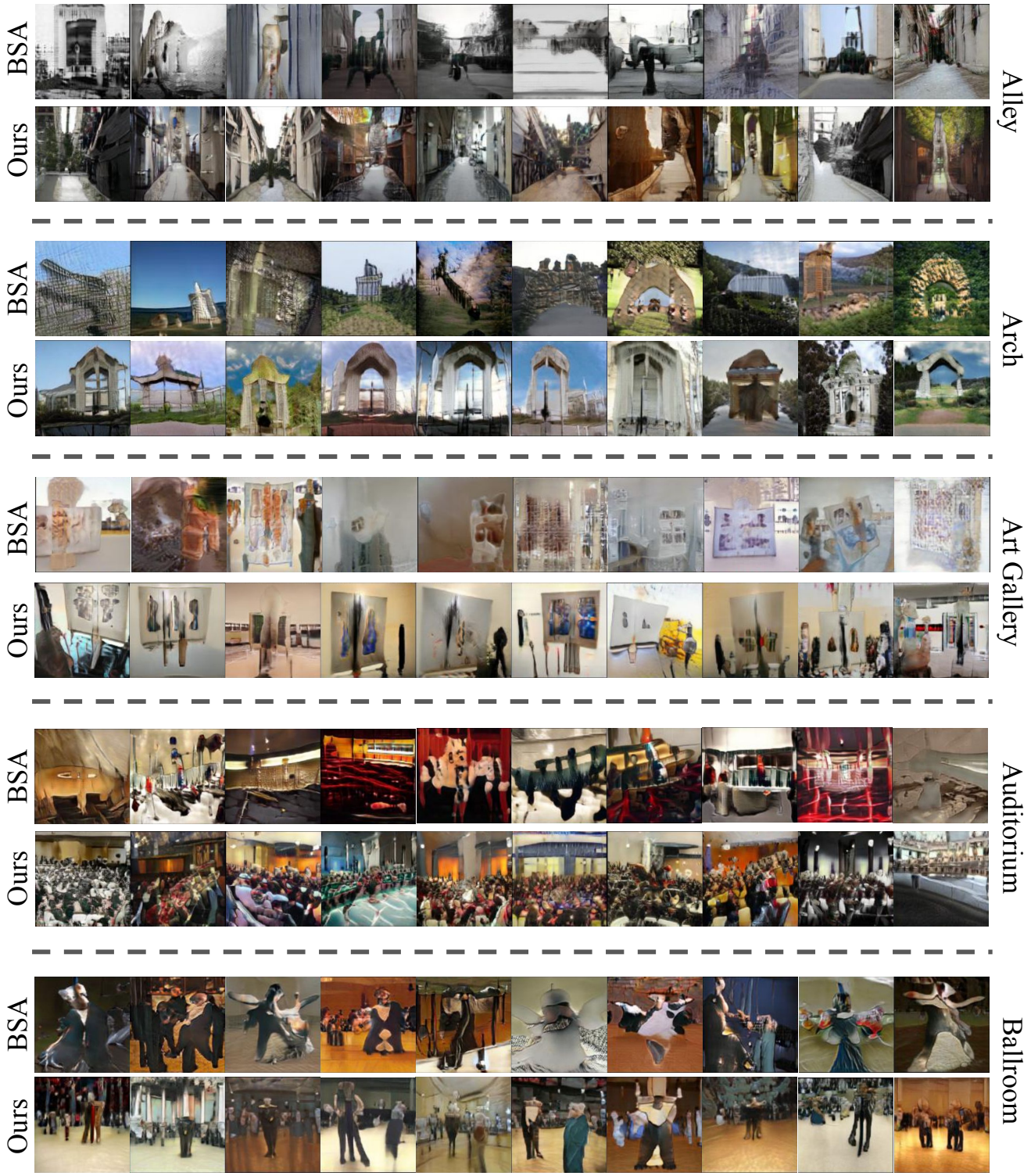

Figure 5: Visual comparison between the images obtained from BSA (no knowledge transfer across classes) and our method on 5 classes of Places365.

Figure 6: Visual comparison between the images obtained from BSA (no knowledge transfer across classes) and our method on some of the classes of Animal Face (for each class, the first row is from BSA, and the second row from our method).

Figure 7: Visual comparison between the images obtained from BSA (left) and our method (right) for transferring from 80 classes of CIFAR100 to 20 classes each containing 300 samples.

Figure 8: Visual comparison between the images obtained from BSA (left) and our method (right) for transferring from 80 classes of CIFAR100 to 10 classes each containing 300 samples.