Semantic Relation Preserving Knowledge Distillation for Image-to-Image Translation Supplementary Material

Zeqi Li*, Ruowei Jiang*, and Parham Aarabi

ModiFace
{lizeqi,irene,parham}@modiface.com

1 Supplementary

In this supplementary material, we provide more results on the network latency measurement, FID trade off curve and more qualitative results for CycleGAN experiment. We also include objective equation for Pix2Pix setup and its experimental results on Cityscapes dataset.

1.1 Latency Measurement for Different Configurations

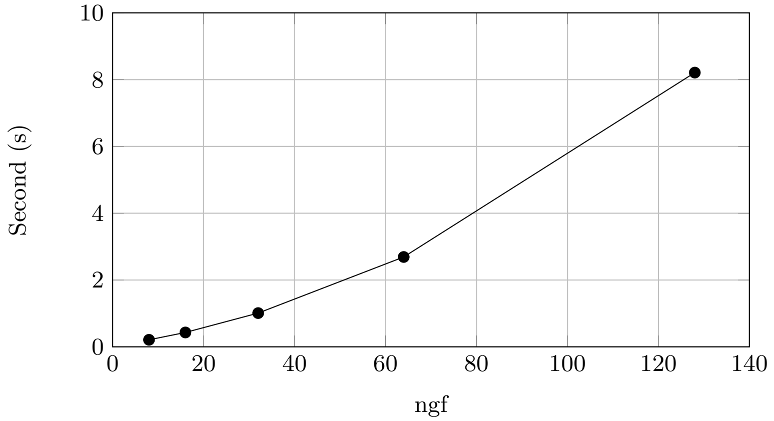

Fig. 1. Latency measurement on a single CPU core of Intel(R) Xeon(R) E5-2686 with different ngfs for Resnet based generators on CycleGAN Experiment

* Equal contribution

1.2 FID Trade off Curve for Different Configurations for CycleGAN Experiment

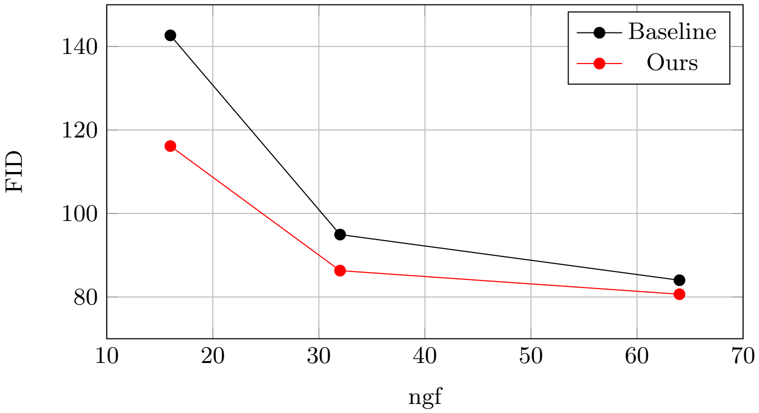

Fig. 2. FID trade off curve on $h \rightarrow z$ for CycleGAN Experiment

1.3 More Qualitative Results for CycleGAN Experiment

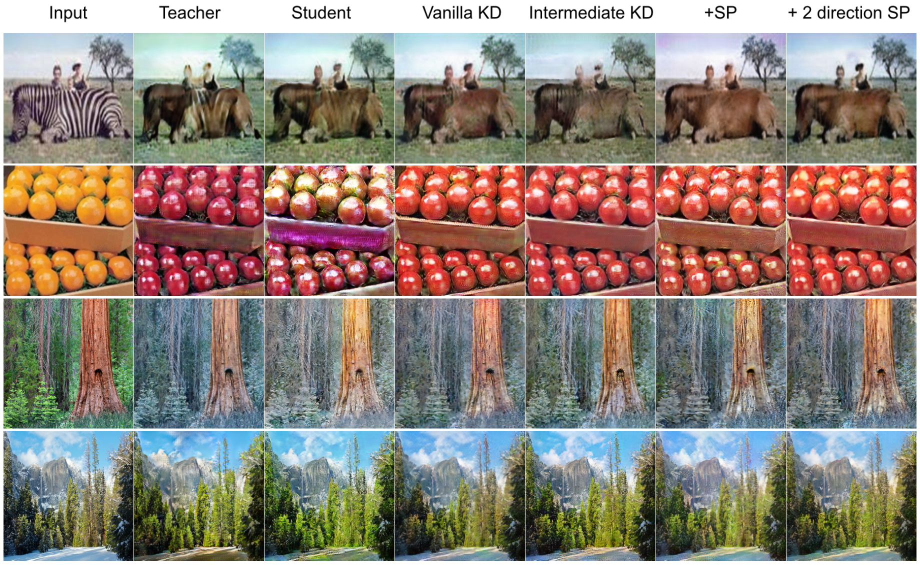

Fig. 3. Ablation study: more examples from multiple datasets comparing results in baseline models and variation of our methods in CycleGAN experiment

1.4 Knowledge Distillation Objective Function for Pix2Pix Experiment

Vanilla Knowledge Distillation. Different from CycleGAN setting, which involves two generators and cycle consistency loss, Pix2Pix only does one direction translation trained with paired data in a supervised way. Analogy to how vanilla knowledge distillation applied on classification task, the objective function has the following form:

$$\mathcal{L}(G_s, D) = \lambda\big(\alpha \cdot \mathcal{L}_{L1}(G_s, X, Y) + (1-\alpha) \cdot \mathcal{L}_{L1}(G_s, X, Y_t)\big) + \mathcal{L}_{adv}(G_s, D, X, Y), \quad (1)$$

where $\mathcal{L}_{L1}$ is an L1 norm loss between the ground-truth labels and the generated images. λ is the balancing coefficient for $\mathcal{L}_{L1}$. α is the hyper-parameter to weigh between the true label and the teacher's label.

Semantic Preserving Knowledge Distillation. Built on vanilla knowledge distillation objective, semantic preserving knowledge distillation loss is directly added to the above objective function:

$$\mathcal{L} = \mathcal{L}_{adv} + \gamma \cdot \mathcal{L}_{SP} + \lambda\big(\alpha \cdot \mathcal{L}_{L1}(G_s, X, Y) + (1-\alpha) \cdot \mathcal{L}_{L1}(G_s, X, Y_t)\big). \quad (2)$$

1.5 Model Size and Computation Results for Pix2Pix Experiment

The teacher and the student models used in Pix2Pix experiments with computation and storage statistics are shown in **Table 1**.

Table 1. Computation and storage results for models on Pix2Pix experiments. The choice is made based on the gap between the teacher and the student baseline performance

Model	Size (MB)	# Params	Memory (MB)	FLOPs
UNet256, ngf 64 (T)	208	54.41M	51.16	2.03G
UNet256, ngf 16 (S)	11(95%↓)	3.40M(94%↓)	13.91(73%↓)	0.14G(93%↓)

1.6 Qualitative Results on Cityscapes

On the Cityscapes dataset, we conducted both paired and unpaired image translation experiments via Pix2Pix and CycleGAN training, respectively. The synthetic street view images translated from their semantic masks along with FCN-8s generated instance segmentation masks are displayed in **Fig 4** and **Fig 5**.

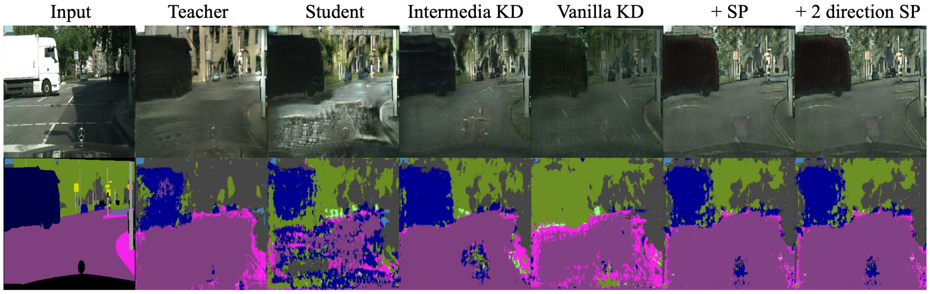

Fig. 4. Ablation study: generated street view images with FCN-8s segmented masks through CycleGAN training on the Cityscapes dataset. The image generated by our method (last column) significantly reduces artifacts compared to the student's generated image. Although the teacher generates a more realistic image, we observe that our model preserves pixels' semantic class with respect to the input mask. For example, in the top right corner, the teacher's generated image only includes buildings in the designated region of trees

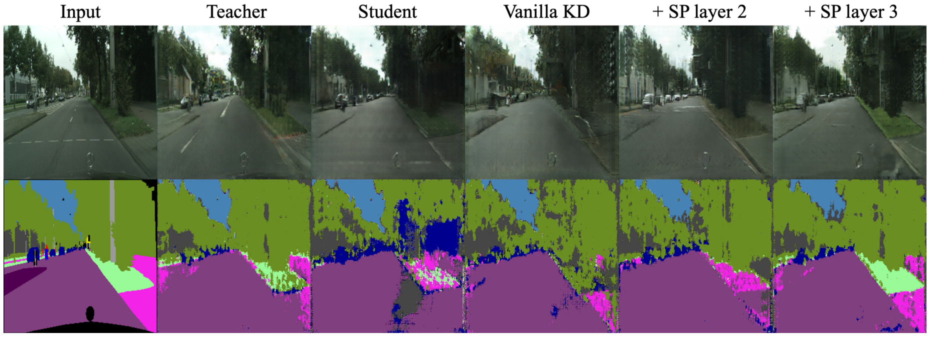

Fig. 5. Ablation study: generated street view images with FCN-8s segmented masks through Pix2Pix training on the Cityscapes dataset. Among all generated masks, our model (last column) shows the most distinct segmentation mask with clear boundaries of each semantic class. For instance, on the right of the segmented masks, we observe a significant improvement at the boundary of the green belt and the side walk.

1.7 Experiment Details

All models are trained on 256x256 input images with a batch size of 1 and optimized using Adam. The other settings for GAN training is the same as CycleGAN and Pix2Pix.

horse ↔ zebra, summer ↔ winter and apple ↔ orange datasets are downloaded using CycleGAN provided script. horse ↔ zebra with segmentation mask sample image, which is used to draw semantic similarity matrices is downloaded from COCO [2]. tiger ↔ leopard dataset is obtained from ImageNet [1] using keyword *tiger* and *leopard*. The Cityscapes dataset is download from the official website [1].

Implementation of FID score is adapted from a PyTorch port version of its official implementation [2]. Calculation of FCN-score is provided in Pix2Pix official Torch implementation [3].

In the vanilla knowledge distillation training, we set $\lambda = 10$ and $\alpha = 0.05$ for all experiments. γ ($\gamma_1 = \gamma_2$) is set to 0.9 in horse ↔ zebra, 0.5 in summer ↔ winter, 0.8 in apple ↔ orange, 0.2 in tiger ↔ leopard and 0.2 in Cityscapes for the unpaired translation experiments. In the paired translation experiments, γ is set to 1 and λ is set to 100.

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
2. Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context. In: Fleet, D.J., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V. Lecture Notes in Computer Science, vol. 8693, pp. 740–755. Springer (2014). https://doi.org/10.1007/978-3-319-10602-1_48, https://doi.org/10.1007/978-3-319-10602-1_48

[1] Official Cityscapes website: https://www.cityscapes-dataset.com/

[2] PyTorch port version of its official implementation: https://github.com/mseitzer/pytorch-fid

[3] Pix2Pix Torch official implementation: https://github.com/phillipi/pix2pix