BridgeKD: Resolution-Aware Knowledge Fusion for Efficient Image Translation

Shilajit Banerjee[1]
banerjee.shilajit@tcs.com
Hitesh Gupta Kattamuri[2]
hiteshgupta.k@tcs.com
Himanshu Pant[3]
pant.h@tcs.com
Anima Majumder[4]
anima.majumder@tcs.com

[1] TCS Research, Kolkata India
[2] TCS Research, Pune India
[3] TCS Research, New Delhi India
[4] TCS Research, Bangalore India

Abstract

Generative Adversarial Networks (GANs) excel in image-to-image translation with high-resolution images. Such images are computationally expensive and may not suit real-life applications. Real-life scenarios demand a lightweight model that performs translation on low-resolution images, which is challenging. Knowledge distillation (KD) helps to improve the performance of lightweight models. We propose a KD framework with a bridging mechanism. It transfers and fuses high-resolution features from the teacher to the low-resolution student via a bridge module. This module aligns feature maps and retains high-resolution details. It operates only during training, ensuring efficiency in inference. We apply a multi-objective loss with adversarial, cycle consistency, and distillation losses. This combination improves image quality. Experiments on benchmark datasets show the effectiveness of the method. The results confirm that resolution-aware fusion enhances the student's ability to recover fine textures while maintaining computational efficiency. Our approach provides a practical balance between quality and efficiency, making image translation feasible for resource-constrained environments. The code will be released after acceptance.

1 Introduction

Image-to-image translation is an important task in computer vision, enabling the transformation of images across domains. The main objective is to map a source image into a target domain while preserving the essential visual properties of the input. This task has several practical applications, including style transfer, image synthesis, domain adaptation, and data augmentation [2]. Researchers employ a variety of deep learning techniques such as Generative Adversarial Networks (GANs) [10], conditional GANs (cGANs) [6], and Convolutional Neural Networks (CNNs) [9] to learn complex mapping functions. GANs [13] achieve remarkable success in image synthesis, inspiring a range of image-to-image translation frameworks. Despite these advances, existing approaches face critical limitations. Paired training

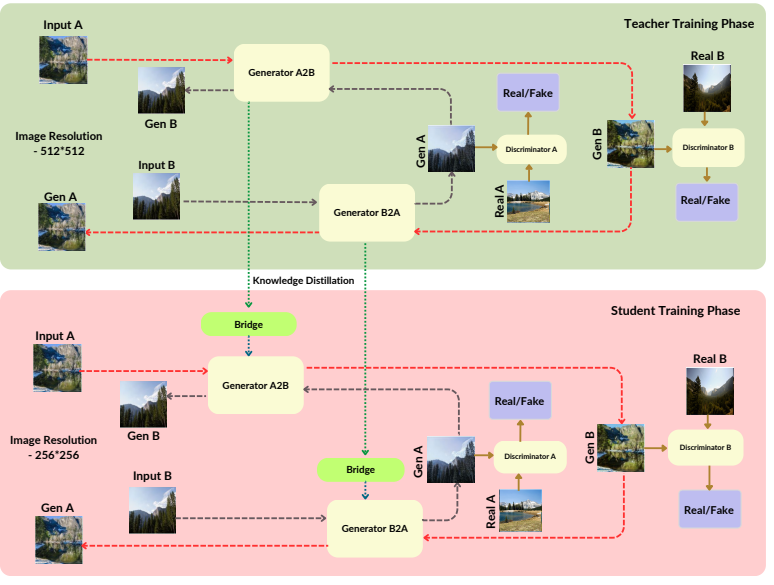

Figure 1: KD framework and model architecture of the proposed BridgeKD. Teacher and student generator details, including feature map dimensions ($C \times H \times W$), are shown in each block. Both models take input from the same domain (Domain-A) but at different resolutions—high for the teacher, low for the student. During training, we distill intermediate features from teacher to student. Since resolutions differ, student features are upsampled via bridge module to match the teacher. The bridge module are used only during training and are removed at inference, adding no runtime cost.

data are often unavailable or expensive to obtain. Models require significant computational resources to learn conditional distributions. High-resolution training further increases cost. These drawbacks restrict their large-scale applicability. Knowledge Distillation (KD) [15] offers a promising strategy to improve lightweight models by transferring knowledge from larger, well-trained models. Research in image-to-image translation evolves rapidly with advances in GAN architectures. Goodfellow et al. [13] introduce GANs as an adversarial framework with a generator and discriminator. This foundation has been extended into several translation models. Pix2Pix [17] provides a general framework for paired image translation, and Pix2PixHD [7] adapts it to high-resolution settings. CycleGAN [36] enables unpaired translation, making it one of the most influential works in the field. Further progress addresses either efficiency or diversity. NICE-GAN [8] and U-GAT-IT-light [18] introduce lightweight designs, but often lose fine-grained details in high-resolution outputs. UNIT [24], MUNIT [16], and DRIT [22] achieve multimodal generation through disentangled representations of content and style, but require heavy computation. CUT [27] and DCLGAN [14] adopt contrastive learning to improve quality, with SimDCL [14] simplifying the process. These developments illustrate the trade-off between efficiency, diversity, and fidelity in translation.

Knowledge distillation is introduced by Hinton et al. [15] as a technique to transfer knowledge from a large teacher model to a smaller student model. It has been widely used in image classification [4, 5], language modeling, and generative tasks. In image translation, KD methods attempt to compress large models without losing translation quality. Prior work explores multiple strategies: Zhang et al. [35] use wavelet-based KD; ReKo [34] applies

region-aware distillation; [25] studies multi-teacher distillation. Other methods rely on intermediate feature matching [12], AutoGAN-Distiller [11] uses AutoML for discovering efficient generators, DMAD [23] applies differentiable masking with co-attention distillation, OMGD [28] introduces multi-granularity distillation, and DCD [4] leverages discriminator cooperation for stable adversarial training. Semantic relation-preserving approaches [5] integrate KD with structural constraints. Co-evolutionary distillation [30] further optimizes convolution filters with regularization. Although effective, most approaches distill teacher and student models at the same resolution. This design ignores the fusion of high-resolution knowledge into low-resolution students, limiting their ability to capture fine details.

There is an increasing demand for efficient image translation models that work on resource-limited devices while maintaining quality. To address this, we propose BridgeKD, a novel KD framework with a bridging mechanism. Unlike direct distillation, our approach fuses high-resolution features from a teacher with low-resolution features from a student. The student's feature maps are upsampled to match the teacher's representations, ensuring effective alignment and feature fusion. This setup allows the student to learn rich semantic and structural details often missing at lower resolutions. High-resolution features guide the student in capturing fine textures and spatial structures, enabling higher-quality outputs. Importantly, the bridge operates only during training and is removed at inference, adding no runtime cost.

We further design a multi-objective loss that integrates adversarial, cycle consistency, and distillation terms. This combination balances visual realism, structural consistency, and effective knowledge fusion. Through this design, the student achieves strong performance while retaining efficiency.

The contributions of this work are summarized as follows:

- We propose BridgeKD, a knowledge distillation framework for image-to-image translation that employs a bridging mechanism to fuse high-resolution features from a teacher into a low-resolution student.
- We design feature alignment at both intermediate and output layers, where student features are upsampled via transpose convolution and matched with teacher representations for effective fusion.
- We introduce a new loss function that combines adversarial, cycle consistency, and distillation losses to guide the student generator and improve translation quality.
- BridgeKD reduces computational cost and memory usage, enabling high-quality translation on resource-constrained devices and low-resolution inputs.
- Extensive experiments on benchmark datasets show that fusion of high-resolution teacher features significantly enhances the student's ability to capture fine details and achieve accurate image-to-image translation.

2 Methodology

We propose a method for knowledge distillation, where a low-resolution student CycleGAN learns from a high-resolution teacher. The student mimics the teacher's high-dimensional features, enabling it to generate high-quality images despite being trained on lower-resolution data.

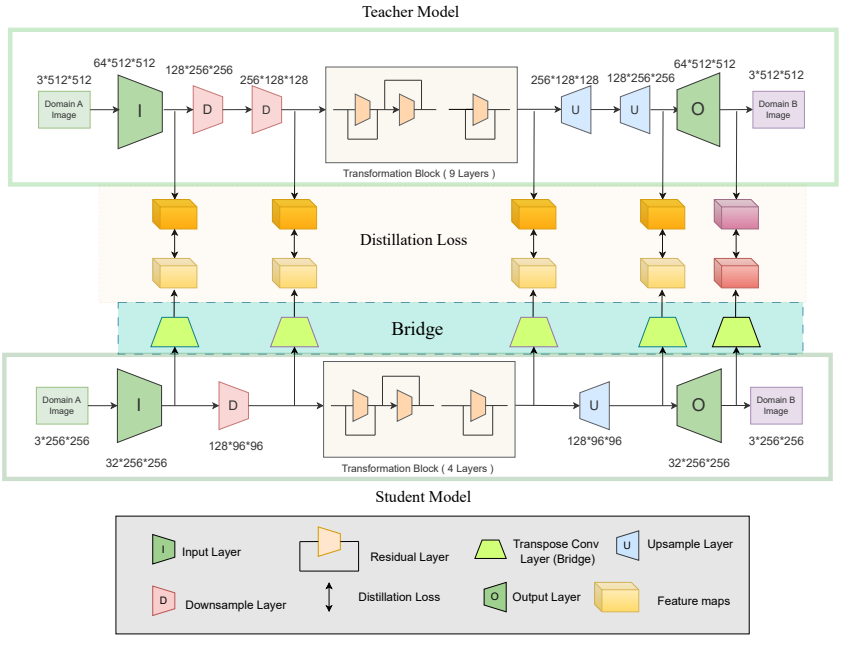

Figure 2: KD framework and model architecture of the proposed BridgeKD. Teacher and student generator details, including feature map dimensions ($C \times H \times W$), are shown in each block. Both models take input from the same domain (Domain-A) but at different resolutions—high for the teacher, low for the student. During training, we distill intermediate features from teacher to student. Since resolutions differ, student features are upsampled via bridge module to match the teacher. The bridge module are used only during training and are removed at inference, adding no runtime cost.

2.1 Teacher Model (High-Resolution CycleGAN)

The teacher model is a standard CycleGAN trained on high-resolution images. It includes two generators and two discriminators. Each generator uses a ResNet-based architecture. It begins with a 7×7 convolution (reflection padding, 64 channels, instance norm, ReLU), followed by two downsampling layers (3×3 conv, 128 and 256 filters, stride 2, padding 1). The core has nine residual blocks with 3×3 convolutions, reflection padding, instance norm, and ReLU. Two upsampling layers use transpose convolutions (4×4 kernel, stride 2, padding 1). The output layer applies a 7×7 conv with reflection padding and tanh activation. See Fig. 2 for the generator architecture. The discriminator processes inputs through four convolutional blocks. Image size reduces from 256×256 to 16×16, and channels increase from 3 to 512. Each block uses 4×4 conv, stride 2, Leaky ReLU (slope 0.2), and instance norm (except the first layer).

2.2 Student Model (Low-Resolution CycleGAN)

The student model uses a lighter architecture for low-resolution image translation. It includes two generators and two discriminators. Extra convolution and transpose convolution layers form a Bridge module to match feature dimensions with the teacher. This module is active only during training and adds no inference cost. The generator follows an encoder-decoder

Table 1: Evaluation of proposed proposed BridgeGAN with Knowledge Distillation (KD) from high resolution teacher and without KD on Various benchmark Datasets[36]. The evolution metric includes : FID (Fréchet Inception Distance), SSIM (Structural Similarity Index), and PSNR (Peak Signal-to-Noise Ratio). The arrows indicate the desired direction of improvement: FID ↓ (lower is better), SSIM ↑ (higher is better), and PSNR ↑ (higher is better). The SSIM and PSNR metrics are calculated between the low-resolution student model and the high-resolution teacher model's output as reference.

Dataset	FID ↓	SSIM ↑	PSNR ↑	KD
Summer to Winter	**39.80**	**69.50**	**21.22**	Yes
Summer to Winter	62.30	66.00	20.59	No
Winter to Summer	**44.65**	**70.20**	**22.42**	Yes
Winter to Summer	61.09	64.90	21.80	No
Apple to Orange	**52.30**	**74.50**	**22.43**	Yes
Apple to Orange	60.27	73.40	21.70	No
Orange to Apple	**63.56**	**70.20**	**21.72**	Yes
Orange to Apple	71.22	68.10	20.88	No

design. It starts with a 7×7 conv (3→32 channels), followed by one downsampling layer (to 128×128, 128 channels). The core has four residual blocks with 3×3 convs, reflection padding, instance norm, and ReLU. One upsampling layer (transpose conv) restores resolution to 256×256 and reduces channels to 32. A final conv layer outputs a 3-channel image. The Bridge module has five transpose conv layers to align student features with the teacher. The student architecture is shown in Fig. 2. The discriminator shares the same design as the teacher's, using stacked conv layers to classify real and fake images.

2.3 Loss Functions

Adversarial Loss:

The adversarial loss ensures generated images appear realistic by distinguishing between real and fake images. This loss drives the generators to create images that can fool their respective discriminators. The loss for the generator G is:

$$\begin{aligned}\mathcal{L}_{\text{adv}}(G,D,X,Y) = \mathbb{E}_{x_i\sim p_{data}(x_i)}\left[\log D_Y(G(x_i))\right] \\ + \mathbb{E}_{y_i\sim p_{data}(y_i)}\left[\log D_X(F(y_i))\right].\end{aligned} \tag{1}$$

Generator G maps images from domain X to domain Y, while generator F maps from domain Y to domain X. Discriminators D_Y and D_X evaluate the authenticity of generated images $G(x_i)$ and $F(y_i)$. This loss pushes generators to produce increasingly convincing images that discriminators classify as real, enhancing the visual quality of translations.

Cycle Consistency Loss:

Cycle consistency loss preserves the content integrity of translated images. It achieves this by ensuring translations maintain semantic information that allows reverse mapping to recover the original image. This bidirectional consistency is crucial for meaningful translations that preserve important characteristics. The loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{cycle}}(G,F) = \mathbb{E}_{x_i\sim p_{data}(x_i)}\left[\|F(G(x_i)) - x_i\|_1\right] \\ + \mathbb{E}_{y_i\sim p_{data}(y_i)}\left[\|G(F(y_i)) - y_i\|_1\right].\end{aligned} \tag{2}$$

The L1 norm measures the absolute pixel-wise difference between the original and reconstructed images. Without this constraint, generators might create arbitrary outputs that look realistic but lose content correspondence with the input. This mechanism guarantees

meaningful transformations rather than arbitrary style changes, effectively addressing the ill-posed nature of unpaired image translation.

Intermediate Layer Feature Matching:

Knowledge transfer between teacher and student generators requires alignment of feature representations across multiple layers. The student generator operates at lower resolution with reduced channel capacity compared to the teacher. To address these dimensional discrepancies, we introduce a Bridge module in the student generator. This module employs transpose convolution layers when both resolution and depth need enhancement. These components align the student's feature maps with the teacher's more complex representations. The intermediate distillation loss minimizes the L2 distance between corresponding feature maps:

$$\mathcal{L}_{\text{intermediate}} = \frac{1}{L}\sum_{l=1}^{L} \left\| f_l^T(x_i) - \mathcal{F}(f_l^S(x_i)) \right\|_2^2 \tag{3}$$

Here, $f_l^T(x_i)$ represents the teacher's feature map at layer l, $f_l^S(x_i)$ is the student's corresponding feature map, and $\mathcal{F}$ represents the appropriate transformation through the Bridge module. This loss enables the student to capture the rich hierarchical features learned by the teacher, improving representation quality despite the student's compact architecture.

Output Distillation Loss:

While intermediate features capture internal representations, output distillation ensures the final generated images match in quality. This loss directly aligns the end results of both networks:

$$\mathcal{L}_{\text{output}} = \left\| G^T(x_i) - G^S(x_i) \right\|_2^2 \tag{4}$$

Where $G^T(x_i)$ and $G^S(x_i)$ represent the outputs of teacher and student generators. This direct supervision at the output level complements intermediate supervision by focusing on final image quality. It ensures the student's results closely resemble the teacher's high-fidelity outputs, particularly important for preserving fine details and textures that contribute to visual realism.

Total Loss:

The total loss combines all components with weighted importance:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{adv}} + \lambda_2 \mathcal{L}_{\text{cycle}} + \lambda_3 \mathcal{L}_{\text{intermediate}} + \lambda_4 \mathcal{L}_{\text{output}} \tag{5}$$

Each component addresses a different aspect of translation quality. The adversarial loss ensures visual realism, while cycle consistency maintains content integrity. The intermediate and output distillation losses transfer complex feature representations from teacher to student. Hyperparameters λ_1 through λ_4 balance these competing objectives. The combination of adversarial training and knowledge distillation proves particularly effective for resource-constrained applications where high-quality image translation is still required.

Table 2: Comparison of prior methods for unpaired image-to-image translation on benchmark CycleGAN dataset of Summer to Winter (S to W) and Winter to Summer (W to S) transformation, evaluated using the FID (Fréchet Inception Distance) metric. Lower FID values indicate superior translation quality. Additionally, we report the number of parameters for each model to highlight differences in computational complexity.

Method	S to W FID ↓	W to S FID ↓	Parameters (M) ↓
CycleGAN [36]	78.76	79.58	28.29M
NICE-GAN [8]	76.03	76.44	16.20M
U-GAT-IT-light [25]	88.41	80.33	21.20M
UNIT [24]	112.07	95.93	-
MUNIT [23]	114.08	99.14	-
RIT [22]	81.64	78.61	-
Co-Evolution [31]	79.16	78.58	-
AutoGAN-Distiller [21]	78.33	77.73	-
DMAD [20]	78.24	70.97	0.45M
OMGD [26]	73.79	-	**0.14M**
DCD [9]	73.63	-	**0.14M**
CoroNetGAN (85%) [33]	74.70	-	1.67M
CoroNetGAN (75%) [33]	72.30	-	2.69M
SPR [32]	-	70.90	11.38M
SCONE-GAN [1]	51.70	-	-
MSGAN [34]	51.85	46.23	-
Proposed	**39.80**	**44.65**	1.35M

Table 3: Comparison of prior methods for unpaired image-to-image translation on the Orange to Apple (O to A) and Apple to Orange (A to O) transformation using the FID (Fréchet Inception Distance) metric. Lower FID values indicate better translation quality. Additionally, we report the number of parameters (in millions) for each model, where fewer parameters indicate a more lightweight architecture.

Method	O to A ↓	A to O ↓	Parameters (M) ↓
CycleGAN [36]	117.70	174.08	28.29M
CUT [24]	127.00	177.83	14.40M
FastCUT [24]	-	156.77	-
DCLGAN [18]	124.90	-	28.81M
SimDCL [18]	134.40	-	28.85M
SRPKD [8]	132.83	121.17	2.85M
DiscoGAN [15]	345.54	377.58	16.56M
StarGAN [2]	167.68	222.71	53.20M
RfGAN [28]	136.64	173.13	14.14M
AttentionGAN [30]	-	168.71	-
MSPC [35]	-	205.84	-
EUITTS [29]	-	156.76	-
Proposed	**63.56**	**52.30**	**1.35M**

3 Experiments and Results

3.1 Datasets

We use several unpaired image datasets from [36] to train and evaluate our models, designed for upaired image-to-image translation tasks without paired examples. The Summer to Winter dataset [36] consists of outdoor scenes captured in both summer and winter conditions. The Apple to Orange dataset [36] contains images of apples and oranges in an upaired fashion.

3.2 Implementation Details

We implement our knowledge-distilled CycleGAN models in PyTorch and optimize the networks using a training schedule for stability and convergence. We use the Adam optimizer for all networks with a learning rate of $\text{lr} = 0.0004$, a first momentum term $\beta_1 = 0.5$, and a second momentum term $\beta_2 = 0.999$. We set the hyperparameters $\lambda_1 = 1.0$, $\lambda_2 = 6.0$, $\lambda_3 = 5.0$, and $\lambda_4 = 5.0$ to balance the loss terms, stabilizing the training process and improving the model's performance. We employ a dynamic learning rate scheduler with a linear decay starting at epoch 50 and continuing until the end of epoch 200. This scheduler is applied to both the generator and discriminator optimizers, ensuring a consistent learning

Table 4: Ablation study on how using a high-resolution teacher generator in Knowledge Distillation (KD) improves image generation quality compared to using a low-resolution teacher. We tested this Summer to Winter, Winter to Summer, Apple to Orange and Orange to Apple generation tasks. The evaluation metrics FID ↓ (lower is better), SSIM ↑ (higher is better), and PSNR ↑ (higher is better).

Dataset Used	FID ↓	SSIM ↑	PSNR ↑	KD Approach
Summer to Winter	**39.80**	**69.40**	**21.22**	**512 × 512**
Summer to Winter	46.04	67.80	21.10	256 × 256
Winter to Summer	**44.65**	**70.20**	**22.42**	**512 × 512**
Winter to Summer	60.70	66.40	22.30	256 × 256
Apple to Orange	**52.30**	**74.50**	**22.43**	**512 × 512**
Apple to Orange	60.60	73.90	22.07	256 × 256
Orange to Apple	**63.56**	**70.20**	**21.72**	**512 × 512**
Orange to Apple	64.04	68.50	21.26	256 × 256

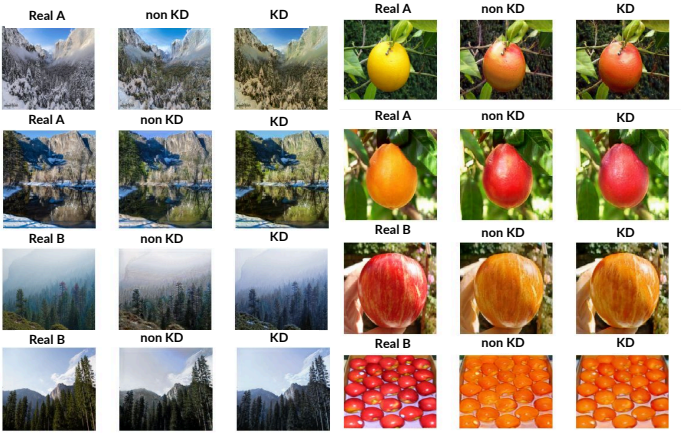

Figure 3: Visual comparison of proposed BridgeKD and Non-KD models on two translation tasks. Left: Summer ↔ Winter. Right: Apple ↔ Orange. Each group shows real input, Non-KD output, and proposed BridgeKD output (left to right). KD improves translation results.

rate adjustment across both components of the model. This approach helps stabilize training by allowing the learning rate to decrease over time, potentially improving the overall performance of the model.

3.3 Comparative Performances

We compare the performance of the proposed method with different state-of-the-art approaches. Next, we present the comparative performances for different datasets.

3.3.1 Summer-Winter Transformation

We train the student model with knowledge distillation (KD) and without it (non-KD) for summer-to-winter and winter-to-summer transformations. We evaluate models using Fréchet Inception Distance (FID) for visual quality and Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for similarity to teacher-generated images. The teacher model trains on higher-resolution images. For unpaired image translation, we generate high-resolution counterparts using the teacher model to evaluate student-generated images. FID scores (lower better) improve with KD for summer-to-winter (39.80 vs 62.30) and winter-to-summer (44.65 vs 61.09). SSIM (higher better) and PSNR (higher better) also increase with

KD across both transformations. Table 1 presents these results, while Table 2 compares with state-of-the-art methods. Image results appear in Fig. 3.

3.3.2 Apple-Orange Transformation

We train models with and without knowledge distillation (KD) for apple-orange transformations. We use FID for visual quality assessment and SSIM and PSNR for similarity measurement. For these metrics, we generate high-resolution counterparts using the teacher model, as they require paired images. FID scores (lower better) improve with KD for apple-to-orange (52.30 vs 60.27) and orange-to-apple (63.56 vs 71.22). SSIM and PSNR (both higher better) also increase with KD across both transformations. Table 1 presents these results, while Table 3 compares with state-of-the-art approaches. Image results appear in Fig. 3.

3.4 Ablation Studies

We perform ablation studies to analyze the role of teacher resolution in knowledge distillation. Two settings are tested: a 256×256 teacher and a 512×512 teacher transferring knowledge to a 256×256 student. Results in Table 4 show that higher-resolution teachers consistently improve student performance. This improvement comes from the fusion of richer high-resolution features, which capture finer textures and details, providing stronger supervision during distillation.

3.5 Comparison Between Teacher and Student Generators

BridgeKD achieves efficiency without sacrificing quality. The teacher model has 11.66M parameters, while the student reduces this to 1.35M. Model size decreases from 44.5 MB to 5.14 MB, and inference speed improves from 26s to 4.74s on 200 images (5.5$\times$ faster). These results demonstrate that fusing high-resolution teacher knowledge allows the student to achieve high-quality translation while remaining suitable for deployment on resource-limited devices.

3.6 Discussion

Our experiments highlight the value of resolution-aware knowledge fusion in image-to-image translation. Students distilled from 512×512 teachers consistently outperform those trained with 256×256 teachers across multiple translation tasks. This shows that teacher resolution directly influences the quality of transferred knowledge. The student benefits from fused high-resolution guidance while maintaining parameter efficiency and fast inference. BridgeKD therefore balances accuracy and efficiency, making high-quality image translation feasible in real-world, resource-constrained environments.

4 Conclusion and Future Work

We present a bridge-based knowledge distillation method for CycleGANs to enable efficient unpaired image translation. A high-resolution teacher guides a compact student model, improving output quality while reducing complexity. Experiments show consistent gains in

FID, SSIM, and PSNR over non-KD models. The high-res teacher improves student performance, highlighting the role of feature quality. Our method achieves 88% fewer parameters, 88% storage savings, and 5.5× faster inference. Future work includes using multiple student models trained from one teacher to boost generalization. We also plan to explore adaptive resolution and attention for better knowledge transfer.

References

[1] Iman Abbasnejad, Fabio Zambetta, Flora Salim, Timothy Wiley, Jeffrey Chan, Russell Gallagher, and Ehsan Abbasnejad. Scone-gan: Semantic contrastive learning-based generative adversarial network for an end-to-end image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. URL https://arxiv.org/abs/2311.03866.

[2] Aziz Alotaibi. Deep generative adversarial networks for image-to-image translation: A review. In *Symmetry*, volume 12, page 1705. MDPI, 2020.

[3] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Dongxu Li, Xue Li, Meng-Hao Wu, and Sarah K. A. O. H. Ibrahim. Review of deep learning: Algorithms and applications in medical image analysis. In *J. Imaging*, volume 7, page 51. MDPI, 2021.

[4] Shilajit Banerjee and Angshuman Paul. An ensemble of well-trained students can perform almost as good as a teacher for chest x-ray diagnosis. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2024. doi: 10.1109/ISBI56570.2024.10635412.

[5] Shilajit Banerjee and Angshuman Paul. Knowledge distillation for an ensemble of students from a pyramid of teachers with diverse perspective. *IEEE Transactions on Artificial Intelligence*, pages 1–10, 2025. doi: 10.1109/TAI.2025.3591588.

[6] Anis Bourou, Valérie Mezger, and Auguste Genovesio. Gans conditioning methods: A survey. In *arXiv:2408.15640*, 2024. URL https://arxiv.org/abs/2408.15640.

[7] H. Chang, H. Yue, A. Doulamis, I. Doulamis, and D. Metaxas. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8798–8807, 2018. URL https://arxiv.org/abs/1711.11585.

[8] R. Chen, W. Huang, B. Huang, F. Sun, and B. Fang. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8165–8174, 2020. URL https://arxiv.org/abs/2003.00273.

[9] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8789–8797, 2018.

[10] M Durgadevi and Others. Generative adversarial network (gan): A general review on different variants of gan and applications. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pages 1–8. IEEE, 2021.

[11] Yonggan Fu, Wuyang Chen, Haotao Wang, Haoran Li, Yingyan Lin, and Zhangyang Wang. Autogan-distiller: Searching to compress generative adversarial networks. In *ICML 2020*, 2020. URL https://arxiv.org/abs/2006.08198.

[12] T. Gao and R. Long. Accumulation knowledge distillation for conditional gan compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1302–1311, 2023.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014. URL https://arxiv.org/abs/1406.2661.

[14] Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. Dual contrastive learning for unsupervised image-to-image translation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 746–755, 2021.

[15] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. In *Proceedings of the Neural Information Processing Systems (NIPS)*, 2015. URL https://arxiv.org/abs/1503.02531.

[16] X. Huang, MY. Liu, S. Belongie, and J. Kautz. Multimodal unsupervised image-to-image translation. In *Computer Vision – ECCV 2018*, volume 11207 of *Lecture Notes in Computer Science*. Springer, Cham, 2018. URL https://doi.org/10.1007/978-3-030-01219-9_11.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, 2017. URL https://arxiv.org/abs/1611.07004.

[18] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. 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=BJlZ5ySKPH.

[19] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17)*, pages 1857–1865. JMLR.org, 2017.

[20] Aman Kumar, Khushboo Anand, Shubham Mandloi, Ashutosh Mishra, Avinash Thakur, Neeraj Kasera, and Prathosh A P. Coronetgan: Controlled pruning of gans via hypernetworks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2024. URL https://arxiv.org/abs/2403.08261.

[21] Ali Köksal and Shijian Lu. Rf-gan: A light and reconfigurable network for unpaired image-to-image translation. In *Computer Vision – ACCV 2020*, pages 542–559. Springer, Cham, 2021.

[22] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[23] S. Li, M. Lin, Y. Wang, F. Chao, L. Shao, and R. Ji. Learning efficient gans for image translation via differentiable masks and co-attention distillation. In *IEEE Transactions on Multimedia*, volume 25, pages 3180–3189, 2023. URL https://arxiv.org/abs/2011.08382.

[24] M.-Y. Liu, T. M. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. In *CVPR 2018*, 2017. URL http://arxiv.org/abs/1703.00848.

[25] X. Liu, L. Lv, J. Liu, Y. Han, M. Liang, and X. Jiang. More teachers make greater students: Compression of cyclegan. In *International Conference on Intelligent Information Processing*, pages 125–139. Springer, 2024.

[26] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1429–1437, 2019.

[27] T. Park, A.A. Efros, R. Zhang, and JY. Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision – ECCV 2020*, volume 12354 of *Lecture Notes in Computer Science*. Springer, Cham, 2020. URL https://doi.org/10.1007/978-3-030-58545-7_19.

[28] Y. Ren, J. Wu, X. Xiao, and J. Yang. Online multi-granularity distillation for gan compression. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6773–6783, 2021. URL https://arxiv.org/abs/2108.06908.

[29] K Shibasaki and M Ikehara. Enhanced unpaired image-to-image translation via transformation in saliency domain. In *IEEE Access*, volume 11, pages 137495–137505, 2023. URL https://doi.org/10.1109/ACCESS.2023.3338629.

[30] Han Shu, Yunhe Wang, Xu Jia, Kai Han, Hanting Chen, Chunjing Xu, Qi Tian, and Chang Xu. Co-evolutionary compression for unpaired image translation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3234–3243, 2019.

[31] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe. Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks. In *IEEE Transactions on Neural Networks and Learning Systems*, volume 34, pages 1972–1987, April 2023.

[32] S. Xie, Y. Xu, M. Gong, and K. Zhang. Unpaired image-to-image translation with shortest path regularization. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10177–10187, 2023.

[33] Y. Xu et al. Maximum spatial perturbation consistency for unpaired image-to-image translation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18290–18299, 2022.

[34] L. Zhang, X. Chen, R. Dong, and K. Ma. Region-aware knowledge distillation for efficient image-to-image translation. In *BMVC 2023*, 2022. URL https://arxiv.org/abs/2205.12451.

[35] L. Zhang, X. Chen, X. Tu, P. Wan, N. Xu, and K. Ma. Wavelet knowledge distillation: Towards efficient image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12464–12474, 2022.

[36] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017. URL https://arxiv.org/abs/1703.10593.