Where Pixels Meet Perception: Understanding CycleGAN Translations Across Aesthetics

Sathammai Sathappan
Independent Research Project
sathammai2005@gmail.com

***Abstract*—This study explores unpaired image to image translation in fashion domain using two datasets: the Fashion MNIST dataset and a curated Indian to Western bridal couture dataset. In the initial phase, experiments were conducted on the Fashion MNIST dataset to explore CycleGAN's behavior in a controlled setting. It was observed that the model's creative translations were restricted due to the low resolution and grayscale nature of the dataset. Consequently, it prioritized structural consistency over creative style translations. It was further hypothesized that a high resolution and high quality dataset would enable richer translations while preserving structure. To validate this, the second phase involved testing CycleGAN on the curated bridal couture dataset. The model successfully achieved a better balance between creative translations and cycle consistency. This suggests that the quality of the dataset has a greater influence on the translation creativity than architecture or hyperparameter choices.**

I. Introduction

- Style and aesthetics play a critical role in everyday life, and fashion is one such domain.This makes it an ideal candidate for image to image translation, specifically to explore style and aesthetic transfer. Applications for this can be found in virtual try-ons, design automations and particularly, cross-cultural translations.
- Most image to image translations algorithms rely on paired dataset, which are expensive or impractical to collect. CycleGAN addresses this limitation by enabling unpaired translations with the help of adversarial training. While prior works have mainly explored CycleGAN's performance on low resolution synthetic dataset or artistic style transfer, fashion translation bridging culture remains underexplored.
- This study investigates CycleGAN's performance on datasets of varying complexity and quality. Experiments were first conducted on a low resolution, grayscale dataset- the Fashion-MNIST dataset. This provided an insight into the model's behavior in a controlled setting. Observations from this phase provided the baseline conditions for the next experiment, which was conducted on a curated high-resolution bridal couture dataset. This enables the examination of how data quality impacts the model's ability to perform detailed style translation while maintaining cycle consistency.

II. Datasets

Two datasets were used for evaluation: the Fashion-MNIST dataset and a curated bridal couture dataset. These datasets were chosen based on their resolution, quality and complexity, to test their impact on unpaired image-to-image translation.

A. Fashion-MNIST

Fashion-MNIST is a standard benchmark dataset used for image classification. It contains low-resolution (28x28) grayscale images of fashion items. For this study, sandals (label 5) were chosen as domain A and sneakers (label 6) were chosen as domain B. To ensure fairness, each domain contained 6,000 images and were taken from the training dataset. To prepare the dataset for CycleGAN, images were resized to 128x128 pixels and converted to RGB to match the model's input requirements. Normalization was applied during model's training by CycleGAN's default preprocessing. This resulted in a controlled setting suitable for analyzing CycleGAN's behavior under the influence of complex textures or significant color variations.

B. Curated Bridal Couture Dataset

- To explore culturally rich fashion translations in a high resolution setup, a custom bridal couture dataset was curated. Indian traditional bridal attire was chosen as domain-A and Western bridal attire as domain-B. Indian bridal dataset mainly includes hand crafted semi-couture pieces with rich texture and intricate detailing. The Western bridal dataset mainly includes high quality, semi-couture gowns and some ready-to-wear, which are less intricate than the Indian dataset, but still suitable enough to explore detailed style translation, especially from a richer domain to a simpler domain.
- Images were obtained using various web-scraping techniques including Selenium and Playwright for dynamic websites. But since most of the websites had lazy-loading or content rendering via Vue or

Angular, full page scrolls were performed to load all images, after which the pages were saved as PDFs. Images were then extracted from these PDFs using PyMuPDF and pdf2image. This process yielded approximately 5,600 images for the Indian bridal Domain and 4,200 images for the Western bridal domain. To supplement the Western Dataset, 800 high quality images were selected from a Kaggle dataset, resulting in total of 5,000 Western bridal images.

- Images with grooms or other unwanted subjects were removed by manually iterating the dataset, ensuring that each image primarily focused on the bridal gown. Backgrounds were blurred using *rembg* to prevent deviation of the model's focus from the dress. Faces were anonymized using *retinaface* to maintain privacy and prevent unintended ethnicity based artifacts in translations.
- To further balance the dataset, 900 semi-couture images from the Western dataset were manually collected as a subset to apply data augmentation. Augmentations used were horizontal flips, shift-scale-rotate, brightness/contrast adjustment, hue-saturation-value modification, perspective/shear transforms and gamma correction. This increased the Western dataset to a total of 5,900 images, slightly above the Indian dataset (5,600 images), deliberately giving the model more exposure to the relatively simpler Western dataset. This strategy aimed to provide **equitable learning opportunities** for both the domains, ensuring the model captures the subtle nuances of the Western gowns and the intricate textures of the Indian bridal couture effectively. To match the CycleGAN's input requirements, images were resized to 256x256 pixels, instead of the lower resolution used for Fashion-MNIST, ensuring better preservation of intricate details for efficient style transfer.

1) Domain-specific Segmentation Observations: The segmentation accuracy of the *rembg* preprocessing step was notably higher for the Indian domain. This can be a result of the higher resolution and richer colour contrast of the images, which provided clearer foreground-background boundaries. For images where the attire and the background colours did blend, the dense texture of the lehengas provided a higher edge confidence, compared to the softer texture of the western gowns. This often resulted in ambiguous edges, especially in flowing skirts or veils, leading to partial blur or mis-segmentation of the gowns in the Western domain. Additionally, rembg's base model performs better with high color contrast, which favored the Indian domain over the typically white-beige spectrum of Western gowns. Consequently, approximately 100–150 images ($\approx 2\%$) in the Western domain had partially blurred skirts. These images were still included in training, Because they were limited in number and its nature may even help the model focus better on certain gown components, like corsets.

III. Methodology

A. Overview

CycleGAN was trained on the two datasets: Fashion-MNIST, and a curated bridal couture dataset. This two phase set-up was designed to analyze CycleGAN's behaviour across simple and complex domains.

B. Architecture

1) Generators: ResNet-based generators were used for both datasets:

- **Fashion-MNIST:** ResNet 6 blocks (`--netG resnet_6blocks`)
- **Bridal Couture:** ResNet 9 blocks (`--netG resnet_9blocks`)

2) Structure:

- 2 downsampling convolutional layers
- 6 residual blocks (Fashion-MNIST) / 9 residual blocks (Bridal Couture)
- 2 upsampling layers

3) Input/Output:

- 3-channel RGB images

4) Discriminators:

- PatchGAN (`--netD basic`) was used that classifies 70×70 image patches as real or fake. It focuses on local patterns, edges, and textures, which is critical for fabric details and fine-grained translations.

Rationale: Preserves structural details while allowing style translation. For Fashion-MNIST, this ensures basic design translation between sandals and sneakers. For Bridal Couture, it preserves intricate fabrics, embroidery, and textures.

C. Loss Functions

- Adversarial Loss: Encourages generators to produce realistic images in the target domain
- Cycle Consistency Loss: Ensures $F(G(A)) \approx AF(G(A))$ and $G(F(B)) \approx BG(F(B))$
- Identity Loss: Logged for Fashion-MNIST by default, but not explicitly used for Bridal Couture; it did not influence optimization.

D. Hyperparameters

Parameter	Fashion-MNIST	Bridal Couture	Key Points
Epochs	25	200 (planned 300; training stopped at 161 due to hardware limits)	Fashion-MNIST: constant LR; Bridal Couture: 100 epochs constant LR + 100 epochs linear decay
Learning rate	0.0002	0.0002	Default Adam ($\beta_1 = 0.5, \beta_2 = 0.999$)
Batch size	4	4	Keeps GPU memory manageable
Pool size	50	50	Stores fake images for discriminator stability
Resize / Crop	128×128	Load: 286 $\rightarrow$ Crop: 256	Preserves input resolution; Bridal Couture kept higher resolution for detail

TABLE I: Hyperparameters for different datasets.

This setup enables a comparative analysis of CycleGAN's performance across low-resolution and high-resolution datasets, highlighting the effect of data quality and complexity on translation creativity.

IV. Evaluation Metrics

Both quantitative and qualitative analysis was performed to evaluate CycleGAN's performance on Fashion-MNIST and the curated bridal dataset. A variety of metrics were chosen to capture different aspects of translation quality and cycle reconstruction, including pixel level fidelity, structural consistency, perceptual realism, and color accuracy.

A. Metric Categories and Descriptions

1) Pixel-wise metrics (measures raw differences between generated and reference images):

- MSE (Mean Squared Error): It measure the average square differences between the corresponding pixels. Lower the value, higher is the pixel-wise similarity.
- PSNR (Peak Signal-to-Noise Ratio):It evaluates the ratio of the maximum possible pixel value to the error. Higher value indicates better fidelity.

2) Structure-aware metrics (Helps assess preservations of edges, textures and spatial layout):

- SSIM (Structural Similarity Index): It measures the perceptual similarity by comparing luminance, contrast and structure. A higher value indicates more structural similarity.
- MS-SSIM (Multi-Scale SSIM): It extends SSIM across multiple scales thereby capturing both fine and coarse structural similarities.
- HOG (Histogram of Oriented Gradients): It measures local edge orientations to evaluate structural and texture fidelity.
- Laplacian MSE: It measures differences in second-order derivatives to highlight structural changes like edges.

3) Perceptual metrics (measures similarity in high level feature space, aligns with human perception):

- FID (Fréchet Inception Distance): Compares the distribution of generated images to real images in a feature space. Lower values usually indicate more realistic outputs.
- IS (Inception Score): Evaluates image quality and diversity. Higher the score, more the diversity and realism.
- LPIPS (Learned Perceptual Image Patch Similarity): Measures perceptual similarity using deep feature representations. Lower value indicates closer perceptual match.

4) Color metrics (measures color fidelity and perceptual color differences):

- Delta-E: Gives a quantitative measure of perceptual color difference between generated and reference images.
- Color histogram / HSV histogram: It asses color transfer accuracy by comparing distributions of color.

V. Results

A. Fashion-MNIST

- To evaluate CycleGAN's performance on a simple, low-resolution dataset, the model was trained on the Fashion-MNIST dataset. Sandals were chosen as domain-A and sneakers as domain-B. Quantitative metrics were used to assess the model's ability to preserve structural details and maintain visual fidelity in both identity mappings and cycle reconstruction. The realism and diversity of the translated images were evaluated using FID and IS scores. The results provide insights into the baseline behavior of CycleGAN on a controlled, low-complexity domain.
- Both identity mapping metrics (see Table II) and cycle-consistency metrics (see Table III) produced excellent scores (high SSIM and PSNR, very low MSE). This indicates that the model effectively reconstructed the original image after domain translation while preserving content when the source and target domain are same.
- The observed FID values (see Table IV) for both the domain translations are notably high (108.29). While this many appear counterintuitive, it is an expected behavior due to the low-resolution grayscale

nature of Fashion-MNIST, which differs significantly from the natural RGB images used to train the inception network used in FID computation. Consequently, the distribution of Fashion-MNIST images in the feature space is very far from those learned by the inception model, indicated by the high FID values. The Inception Score is slightly higher for sneakers $\rightarrow$ sandals translation, suggesting a higher diversity in that direction.

Translation Direction	MSE	SSIM	PSNR
Sandals (A)	0.0001	0.994	41.19
Sneakers (B)	0.0002	0.991	40.41

TABLE II: Identity Metrics averaged over 100 images per domain in the Fashion-MNIST dataset

Translation Direction	MSE	SSIM	PSNR
Sandals $\rightarrow$ Sneakers (A$\rightarrow$B)	0.0004	0.975	35.54
Sneakers $\rightarrow$ Sandals (B$\rightarrow$A)	0.0004	0.978	34.51

TABLE III: Cycle-Consistency Metrics averaged over 100 images per domain for in the Fashion-MNIST dataset

Translation Direction	FID	Inception Score
Sandals $\rightarrow$ Sneakers (A$\rightarrow$B)	108.29	2.80 ± 0.08
Sneakers $\rightarrow$ Sandals (B$\rightarrow$A)	108.29	3.42 ± 0.07

TABLE IV: Generative Quality Metrics

1) Qualitative Analysis: Qualitative analysis of the translated images provides an insight into how the model finds a balance between structural preservation and style translations on a low resolution dataset such as the Fashion-MNIST dataset. The model generally applied only subtle modifications for translations. For example, when the model was translating sandals to sneakers, it often added faint wedges to pencil and block heels, while wedge-like heels remain untranslated. Sandals with unusual or complex designs were also mostly unchanged. This was observed in the opposite direction as well: shoes with logos remained mostly untranslated, and converse-like shoes were sometimes converted into unconventional sandals. These observations suggest that on a simple, low-resolution dataset, the generator over-prioritizes structural consistency over style transfer, leaving highly distinctive or unusual elements mostly unaltered.

B. Bridal Couture Dataset

- To evaluate CycleGAN's performance on a complex, high-resolution dataset, the model was trained on the curated bridal couture dataset. It consists of Indian traditional bridal attire as domain-A and Western bridal attire as domain-B. The model's ability to preserve intricate textures, embroidery and structural details in cycle reconstruction and domain translation was evaluated using quantitative metrics.
- Additionally, perceptual and structural fidelity were evaluated using metrics such as LPIPS, color and HSV histograms, edge and HOG-based comparisons, MS-SSIM and ΔE (Delta-E). This evaluation provides insights into how CycleGAN performs detailed style translation for culturally rich, high-resolution fashion images. This highlights the impact of dataset quality and complexity on translation fidelity and creativity.
- Both cycle-consistency metrics (see Table V) and identity metrics (see Table VI) (MSE, PSNR, SSIM, LPIPS) for domain-B (Western bridal) is slightly better than domain-A (Indian). This is highly intuitive, considering Western domain is comparatively simpler than the Indian domain, therefore both reconstruction as well as feature preservation is easier for the Western domain.
- In edge-based metrics (see Table VII), Western $\rightarrow$ Indian translations are slightly better than the reverse. This again aligns with intuition because while minimalism to a maximalism translation is challenging, existing features are mostly preserved. Conversely, Indian $\rightarrow$ Western translation requires removing elaborate visual elements (e.g., jewelry, heavy embroidery), causing a drop in edge and MS-SSIM scores as MS-SSIM penalizes feature loss more heavily than feature addition. Hence the asymmetry.
- For color analysis (see Table VIII), ΔE values are notably higher for Indian $\rightarrow$ Western, because the colourful bridal lehengas are transformed into predominantly white gowns. The Western $\rightarrow$ Indian ΔE values are lower. This may seem counterintuitive at first, but is explainable after visual analyzation of the translated images. Since the model has been trained for only 161 epochs, it's still refining colour mappings, leading to softer, less sharply defined color pixels and hence smaller ΔE differences.
- ΔE for cycle reconstruction is higher in the Indian domain than in the western one, as expected because the wider color range of the Indian bridal attire amplifies even subtle deviations in the color spectrum.

Metric	A→B→A Cycle	B→A→B Cycle
MSE	0.001475	0.000882
PSNR	28.65	30.84
SSIM	0.8983	0.9454
LPIPS	0.13898	0.10481
HOG	0.9293	0.9463
MS-SSIM	0.9168	0.9692

TABLE V: Cycle-Consistency metrics averaged over 100 images for bridal couture dataset dataset

Metric	Identity A (G_B)	Identity B (G_A)
MSE	0.0005	0.00005
PSNR	33.28	34.15
SSIM	0.9641	0.9767
LPIPS	0.0377	0.0466

TABLE VI: Identity mapping metrics averaged over 100 images for bridal couture dataset

Metric	A→B Translation	B→A Translation
Edge MSE	0.001568	0.001462
Edge PSNR	28.68	29.11
Edge SSIM	0.6539	0.6779
Laplacian MSE	469.39	582.83
Laplacian PSNR	27.35	23.77
Laplacian SSIM	0.6616	0.6026
HOG similarity	0.8213	0.8325
MS-SSIM	0.4228	0.6034

TABLE VII: Edge and Structural metrics averaged over 100 images for the bridal couture dataset dataset

Metric	A→B	B→A	A→B→A	B→A→B
Delta-E	19.27	15.98	5.35	3.92

TABLE VIII: Color metrics for domain translations and cycle reconstruction averaged over 100 images for the bridal couture dataset dataset

1) Qualitative Analysis: Visual inspection of the translated images further supports the quantitative analysis. The model effectively maintained overall structural integrity and preserved major textural components across translations in both the directions. The Indian to Western translations mostly involved desaturated and monotonous colors, removal of ornate details, and suppression of intricate embellishments. Western to Indian translations, conversely, introduced more vibrant colors, ornament detailings, and additional texture patterns that resembled embroidery. These observations suggest that both translation directions functioned as near inverses of each other, indicating effective learning by the model. Further, no visual artifacts or distortions were observed. Overall, the translations aligned with the dataset's intended stylistic features, demonstrating a balanced trade-off between content preservation and style translation.

VI. Discussion

- The qualitative along with the quantitative analysis reveal a clear relationship between dataset quality and CycleGAN behavior.
- For Fashion-MNIST, although the numerical metrics (MSE, PSNR, SSIM) for identity mapping and cycle reconstruction were very strong, the visual translations were minimal or sometimes unconventional. This indicates that the generator prioritized cycle consistency over meaningful style transfer. Since Fashion-MNIST is a low-resolution, low-diversity dataset, the discriminator struggled to learn domain variety. Consequently, the generator learnt that minimal transformations were sufficient to fool the discriminator while maintaining low cycle loss. This outcome demonstrates a critical limitation: when the dataset is overly simplistic, the cycle-consistency constraint dominates, preventing creative or semantically rich translations. In such cases, the model prioritizes stability over style expression. Thus, cycle-consistency, which was originally designed to enforce uniqueness, paradoxically becomes a liability in such cases.
- In contrast, the Bridal Couture dataset comprised of complex, high-resolution images that has distinct stylistic features. These characteristics fundamentally changed the model's behavior. The model mainly had two sources of difficulties: the high textural complexity of the Indian domain, and a small percentage of blurred skirts ($\approx$ 2%) in the Western domain caused by imperfect segmentation. These imperfections slightly hindered the Indian→Western translation, but they also tested the model's robustness. Despite early termination at 161 epochs (instead of the planned 300), the generator still produced meaningful translations, even for blurred gowns. Notably, only 10% of the translations were visually exceptional. The remaining 90%, while it was less striking, it accurately reflected the data's natural variability and segmentation noise (unlike for Fashion-MNIST where the CycleGAN sometimes produced unconventional sneakers/sandals). This suggested that the model was faithfully learning the underlying domain distributions, instead of memorizing idealized features.
- Further, for Western→Indian translations, the model produced visually rich translations: it introduced vibrant colors, embroidery-like patterns and ornament detailings. This direction ideally posed a higher risk

of model collapse, given the translation from a minimalistic to a maximalistic domain, yet, the model retained diversity and visual realism. This indicates that the discriminator learned complex high-quality patterns effectively. Therefore, to deceive such a strong discriminator, the generator was compelled to produce creative translations, even if it was at the cost of a perfect cycle reconstruction. Finding the right amount of balance between preserving content and introducing stylistic variations marks a significant step towards stable and expressive unpaired translation. Further, even though the model wasn't optimized for identity loss, the metrics for identity mappings were excellent, indicating that the model learnt meaningful semantic features for both the domains.

- Overall, these findings confirm that CycleGAN's success is directly proportional to the dataset's richness and diversity. With a high quality, well pre-processed dataset, the model achieved stable training, intuitive metrics and visually coherent translations, despite the presence of minor noise. This robust performance of the model, even under imperfect conditions, highlights the importance of thoughtful preprocessing and balanced domain complexity. In essence, the performance of CycleGAN on bridal couture demonstrates the theoretical ideal in practice: when given a strong sandbox, the model learns mappings that are not only accurate but also meaningfully creative.

VII. Conclusion

- This work explored the intricate task of unpaired image-to-image translations using CycleGAN on datasets of varying quality. The results confirm one core truth: the success of a CycleGAN is intrinsically tied to the quality of the dataset. The rigorous preprocessing and careful domain balancing ensured that the model achieved remarkable robustness, despite the presence of subtle noise like the blurred skirts in the Western dataset. The model maintained stable, intuitive metric behavior and produced structurally consistent, visually creative translations across the domains.
- The qualitative analysis further revealed that the model's natural tendency is to prioritize cycle consistency. This behavior is exaggerated in low-quality datasets due to the weaker learning of the discriminator. Whereas the availability of high-quality, detailed, and varied images forces the model to find the ideal balance between reconstruction and translation. In doing so, it validates the theoretical expectations of CycleGAN behavior under near-optimal conditions.
- Ultimately, this project demonstrates the significance of a carefully curated and well processed dataset. Provision of such a dataset enables the CycleGAN to perform at its theoretical best even without architectural modifications. The preprocessing pipeline developed here effectively created a perfect sandbox that allows the model to not only translate with clarity and detail, but to also remain resilient to minor inconsistencies and noises.

VIII. Future Work and Limitations

- While this project achieved stable and deeply insightful results, few limitations arose from the practical constraint of the setup. The first was segmentation accuracy. Although the *rembg* preprocessing pipeline performed surprisingly well given its general purpose design, it still struggled for some images in the Western bridal dataset. This caused a small number of blurred skirts and soft edges in the dataset. Including these images had repercussions on the translations, especially for the earlier epochs, where the noise was exaggerated in the translations. Therefore, a better segmentation or matting tool, such a MODNet or BackgroundMattingV2, could have offered a higher segmentation accuracy, further refining the dataset's visual quality. These tools weren't explored for this project due to hardware constraints.
- This leads us to the second limitation - GPU constraints, which restricted training to only 161 epochs. The model had just begun converging, therefore running it for around 300 epochs would've likely resulted in sharper color mapping, especially for the Western to Indian translations, would've increased the sharpness in the patterns, and reduced the reflection of the noise of the dataset in the translations.
- For future work, the next steps could involve exploring architectures that are naturally more resistant to dataset noise, like CUT (Contrastive Unpaired Translation) model or UGATIT model, which use perceptual guidance and attention mechanisms to balance structure and creativity. Another approach could be to combine GANs with diffusion based models. This can help create hybrid systems that are capable of creating detailed, texture-rich translations without relying heavily on a perfect preprocessing.
- It's important to highlight that the limitations here are not reflections of oversight, but of realism. Despite all the limitations or boundaries, whether computational or architectural, the model still performed at its best. Therefore, this work stands as a proof that even with limited resources, a CycleGAN

can be pushed to perform at its theoretical best when guided by precision, clarity and thoughtful data design.

REFERENCES

[1] Isola, P., Zhu, J. Y., Zhou, T., Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Pix2Pix paper)

[2] Zhu, J. Y., Park, T., Isola, P., Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). (CycleGAN paper)

[3] Xiao, H., Rasul, K., Vollgraf, R. (2017). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. arXiv preprint arXiv:1708.07747. (Fashion MNIST dataset)

[4] Deng, J., Guo, J., Ververas, E., Kotsia, I., Zafeiriou, S. (2020). RetinaFace: Single-stage Dense Face Localisation in the Wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (Used for facial bounding boxes and alignment)

[5] rembg: Image Background Removal Tool. (n.d.). Retrieved from https://github.com/danielgatis/rembg(Used for background removal preprocessing)

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004. (SSIM)

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-Scale Structural Similarity for Image Quality Assessment,” IEEE Asilomar Conference on Signals, Systems Computers, 2003. (MS-SSIM)

[8] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005. (HOG)

[9] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” Advances in Neural Information Processing Systems (NeurIPS), 2017. (FID)

[10] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved Techniques for Training GANs,” Advances in Neural Information Processing Systems (NeurIPS), 2016. (Inception Score)

[11] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. (LPIPS)

[12] G. Sharma, W. Wu, and E. N. Dalal, “The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations,” Color Research Application, vol. 30, no. 1, pp. 21–30, 2005. (E)

[13] A. Hore and D. Ziou, “Image Quality Metrics: PSNR vs. SSIM,” 20th International Conference on Pattern Recognition (ICPR), 2010. (PSNR comparison)

[14] R. Gonzalez and R. Woods, Digital Image Processing, 4th ed., Pearson, 2018. (General pixel-wise metrics Laplacian MSE)

[15] E. Azizi, “Bridal Dress,” Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/elhamazizi/bridal-dress

APPENDIX

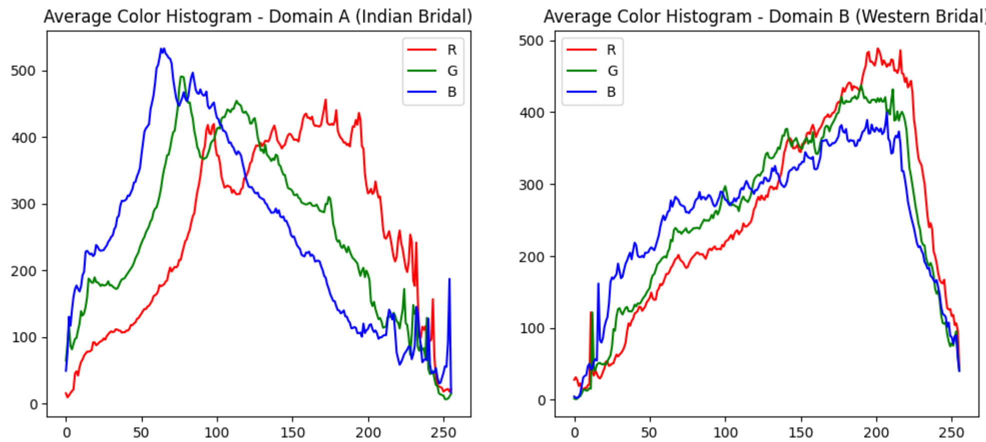

Fig. 1: Color histograms of Indian vs. Western bridal domains. The Indian dataset shows distinct RGB channel peaks, reflecting vibrant color diversity, while the Western dataset exhibits overlapping peaks, indicating a more uniform tonal palette

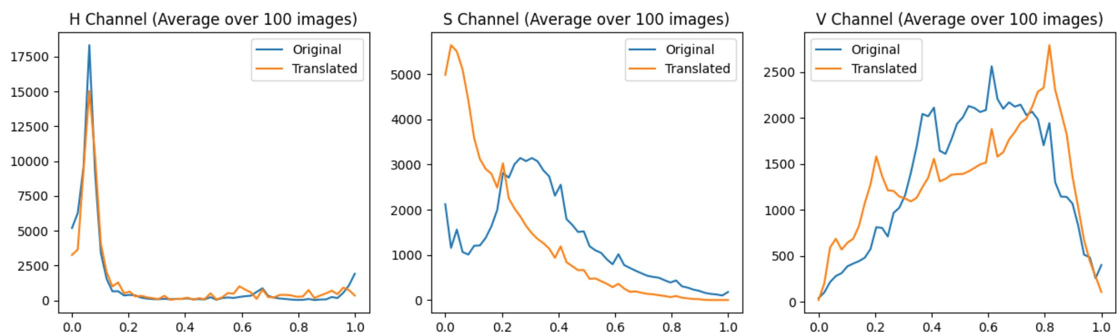

Fig. 2: HSV histograms for Indian→Western translations averaged over 100 images. The similar curve structure for Hue indicates that the model effectively preserved overall color tones across domains. Saturation curve gradually declined in the translated set, suggesting that the generator enhanced vibrancy early on before stabilizing it. Meanwhile, the multi-peaked Value distribution reflects that the model has adapted to varied lighting and fabric brightness, demonstrating its ability to capture stylistic luminance contrasts instead of just replicating texture.

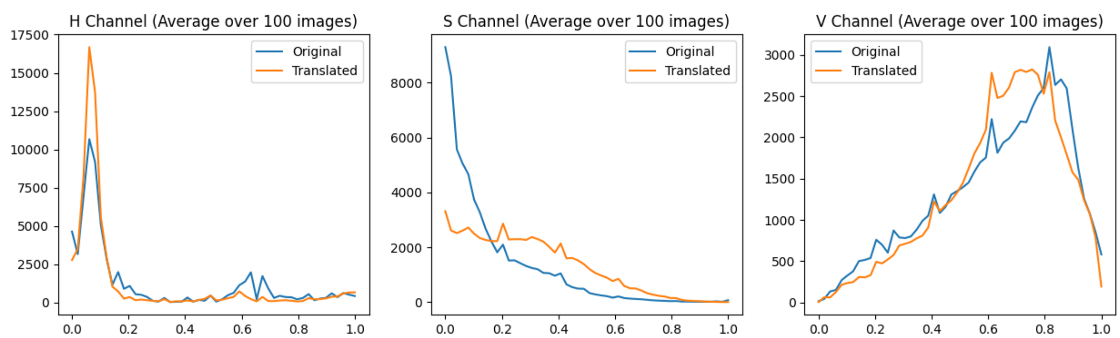

Fig. 3: HSV histograms for Western→Indian translations averaged over 100 images. The similar Hue profiles indicate consistent tone preservation. In contrast, the Saturation distribution of the translated domain is subdued, suggesting that the model underrepresented the richness of color intensity compared the originals. The Value curves maintain comparable structure, reflecting stable brightness mapping despite chromatic disparities.

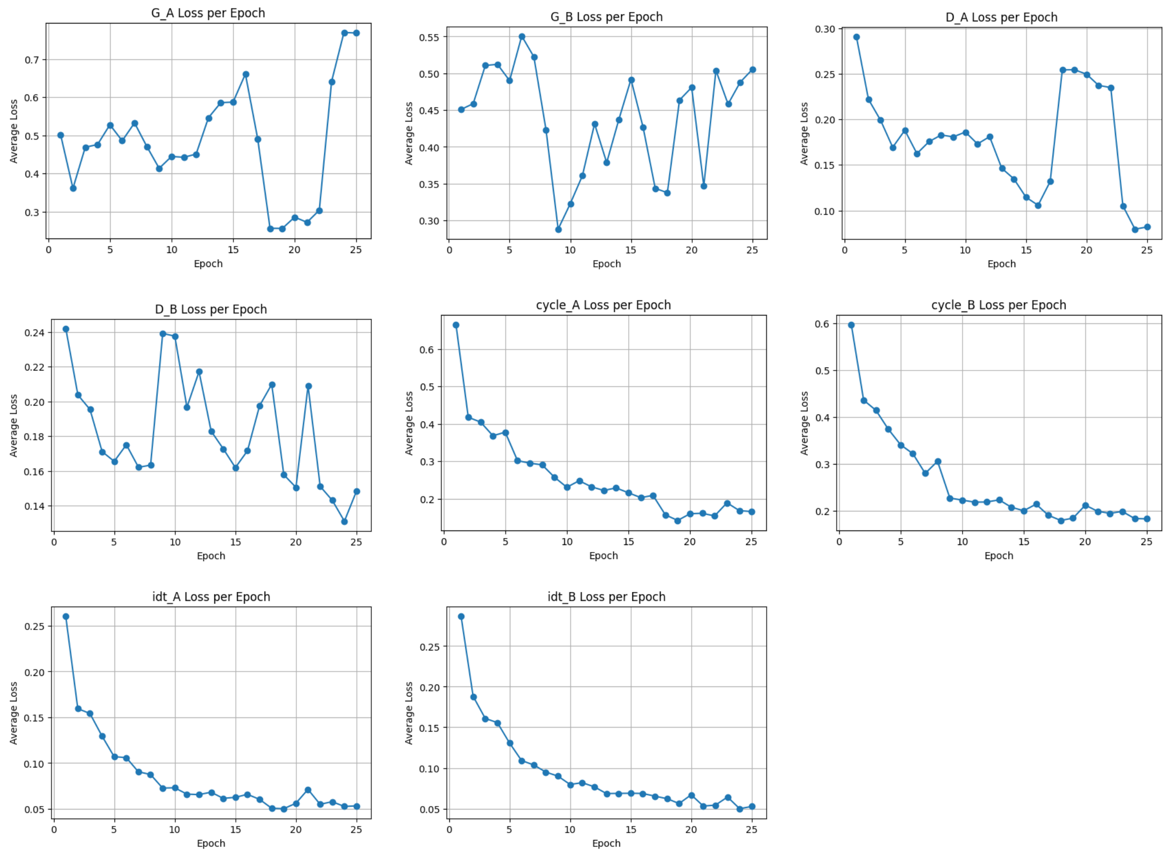

Fig. 4: Loss evolution for the Fashion-MNIST domain translation. For the sneakers domain (B), the generator and discriminator exhibit inverse behavior from early in the training. For the sandals domain (A), both losses remain moderate during the first half of training, after that, the adversarial dynamics intensify. The cycle-consistency and identity losses for both domains have a steady, monotonic decrease, corresponding smooth structural alignment between generated and real images.

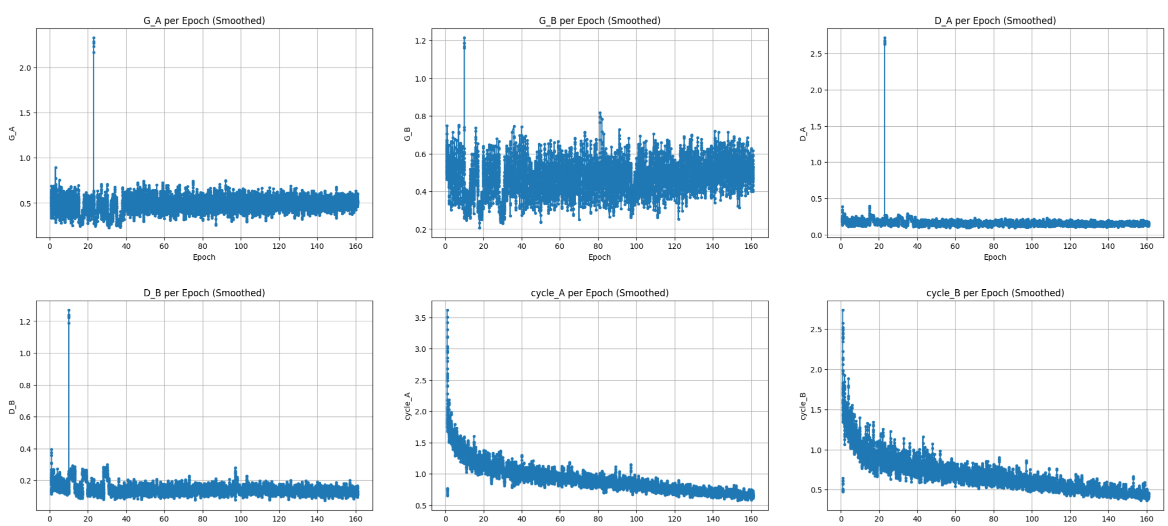

Fig. 5: Loss evolution for the Bridal Couture domain translation. The discriminator losses for both domains (A and B) have a minimal oscillation, which indicates a early stabilization despite the dataset's visual complexity. In contrast, the generator losses oscillate more noticeably, indicating that its continuously adapting to the intricate detailing. The cycle-consistency loss decreased steadily throughout training, indicating improving structural preservation.

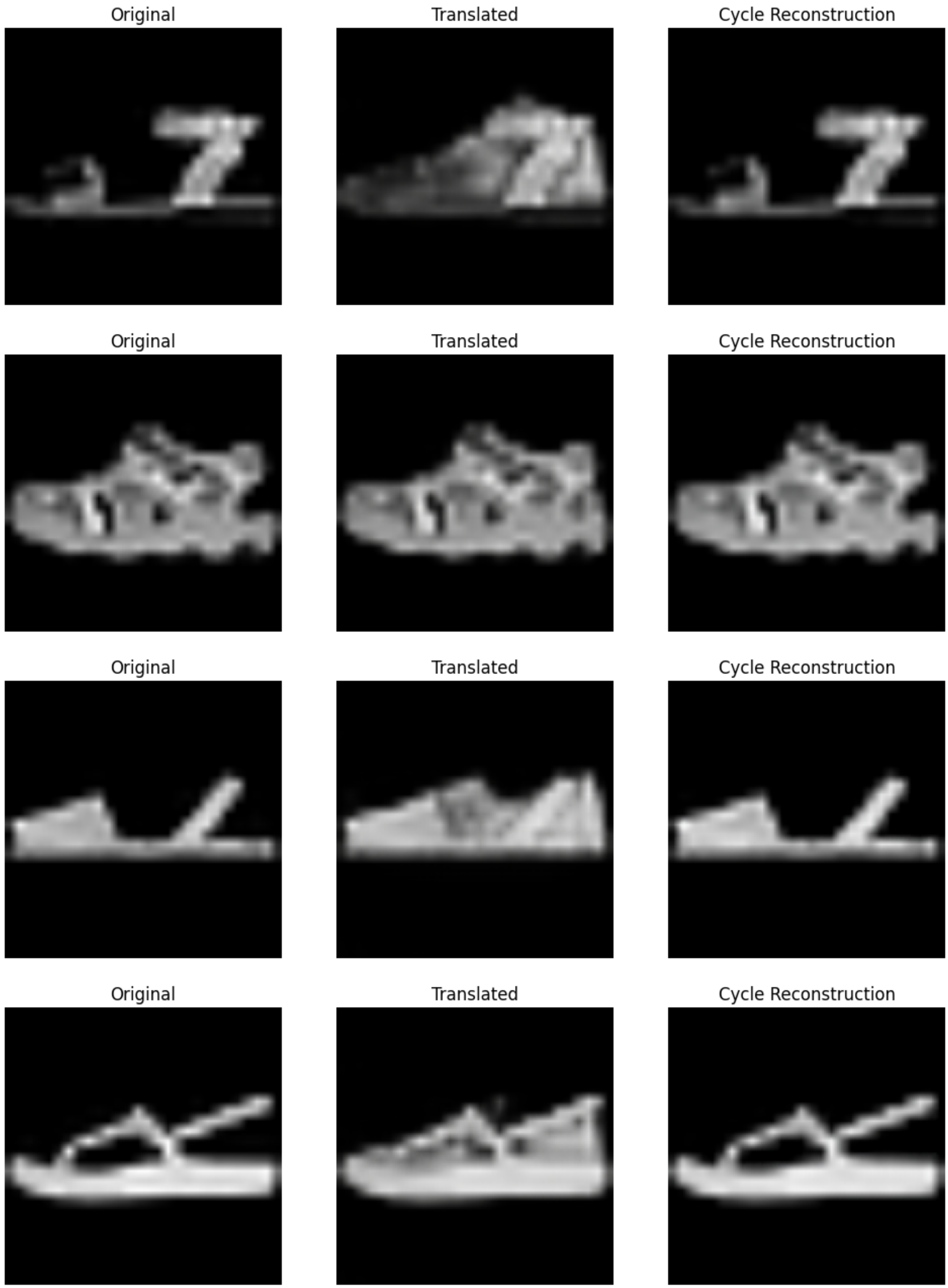

Fig. 6: Representative examples of successful translations for the Sandals $\rightarrow$ Sneakers direction. Each triplet shows the original sandal, the translated sneaker, and the reconstructed sandal.

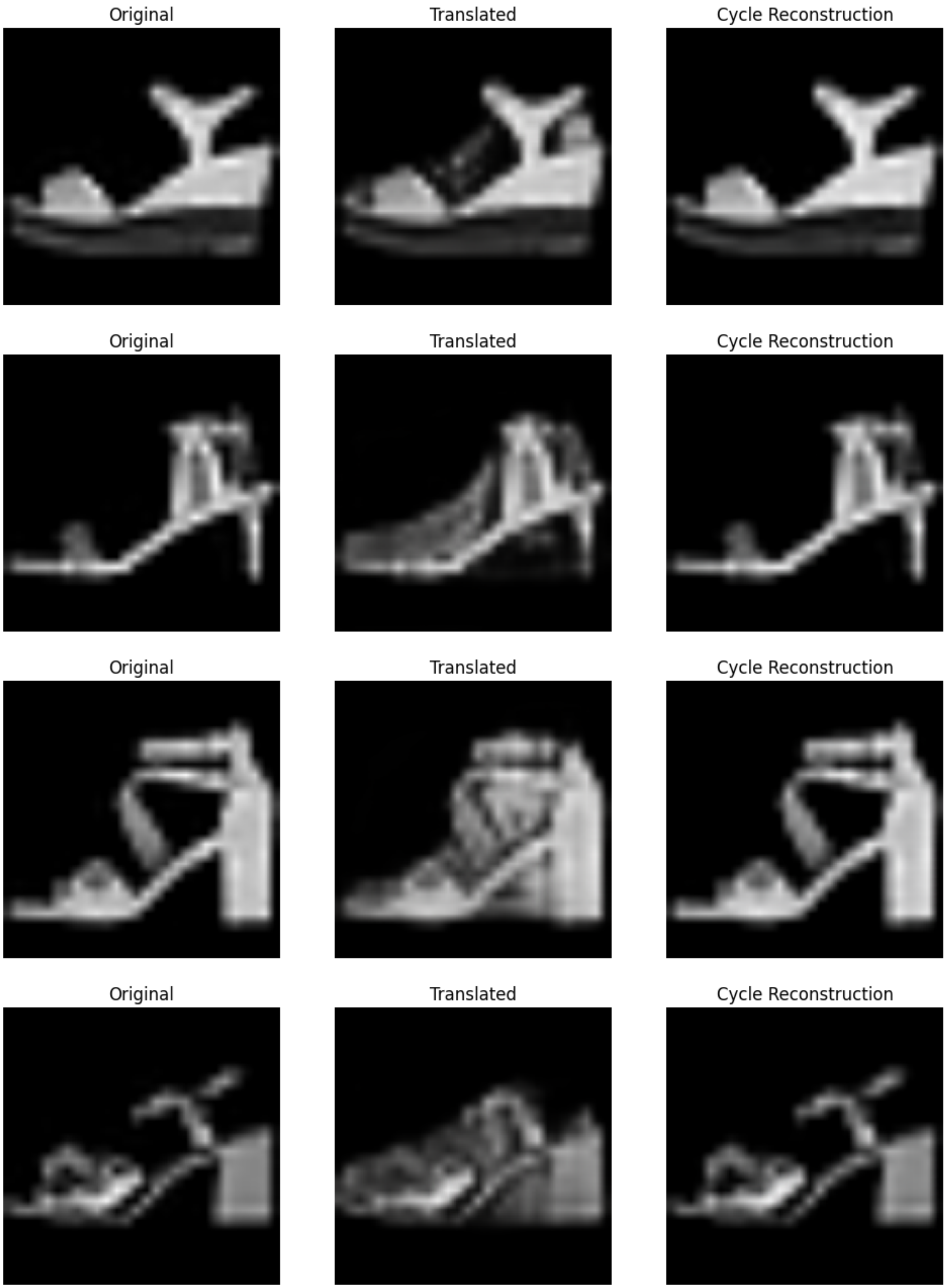

Fig. 7: Representative mid-level translations for the Sandals $\rightarrow$ Sneakers direction where the generator partially adapted the source sandals into more enclosed footwear, resembling an intermediate form between sandals and sneakers.Each triplet shows the original sandal, the translated sneaker, and the reconstructed sandal.

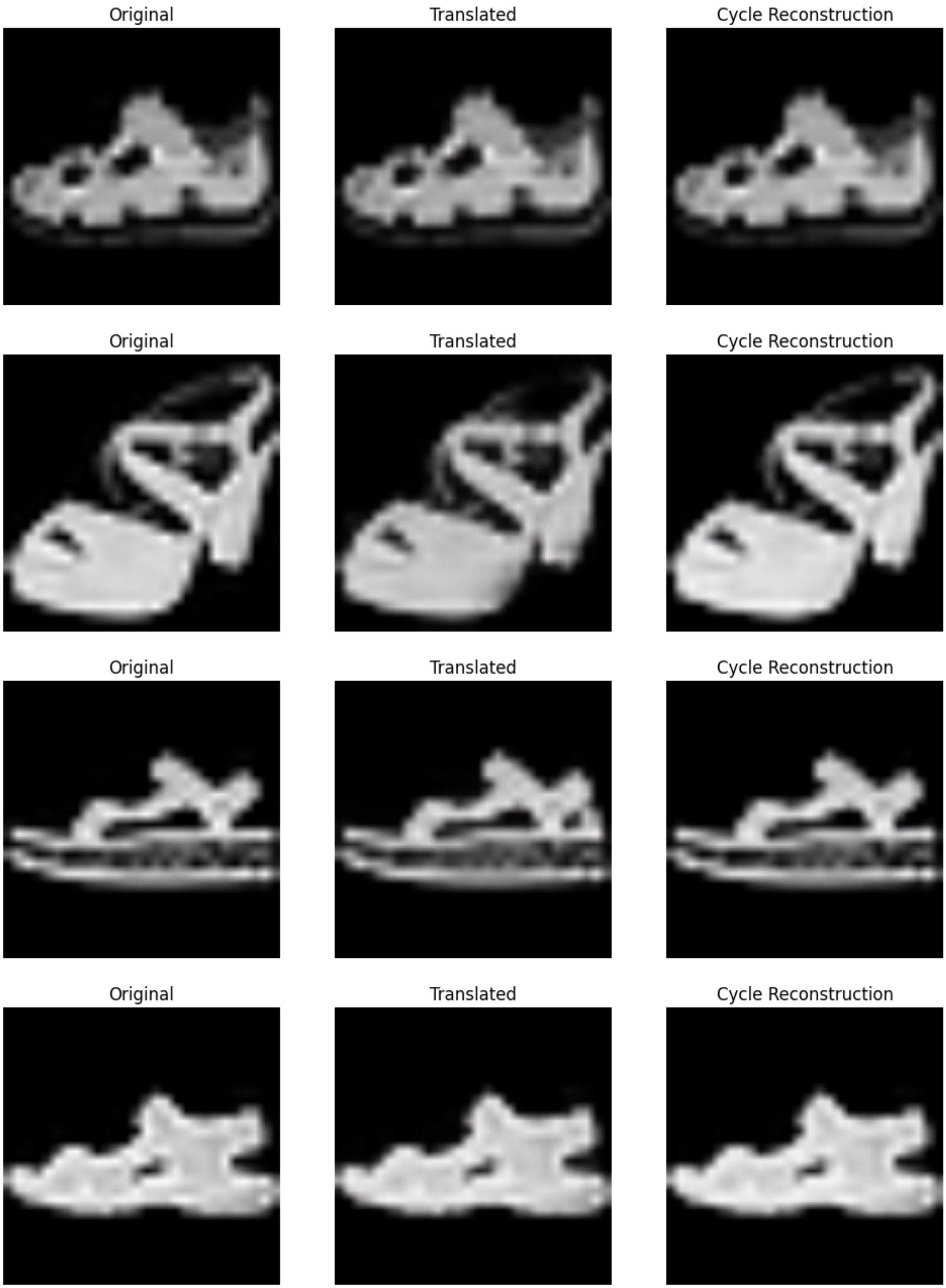

Fig. 8: Representative examples for the Sandals $\rightarrow$ Sneakers direction where the generator failed to perform domain translation, leaving the input sandals nearly unaltered.Each triplet shows the original sandal, the translated sneaker, and the reconstructed sandal.

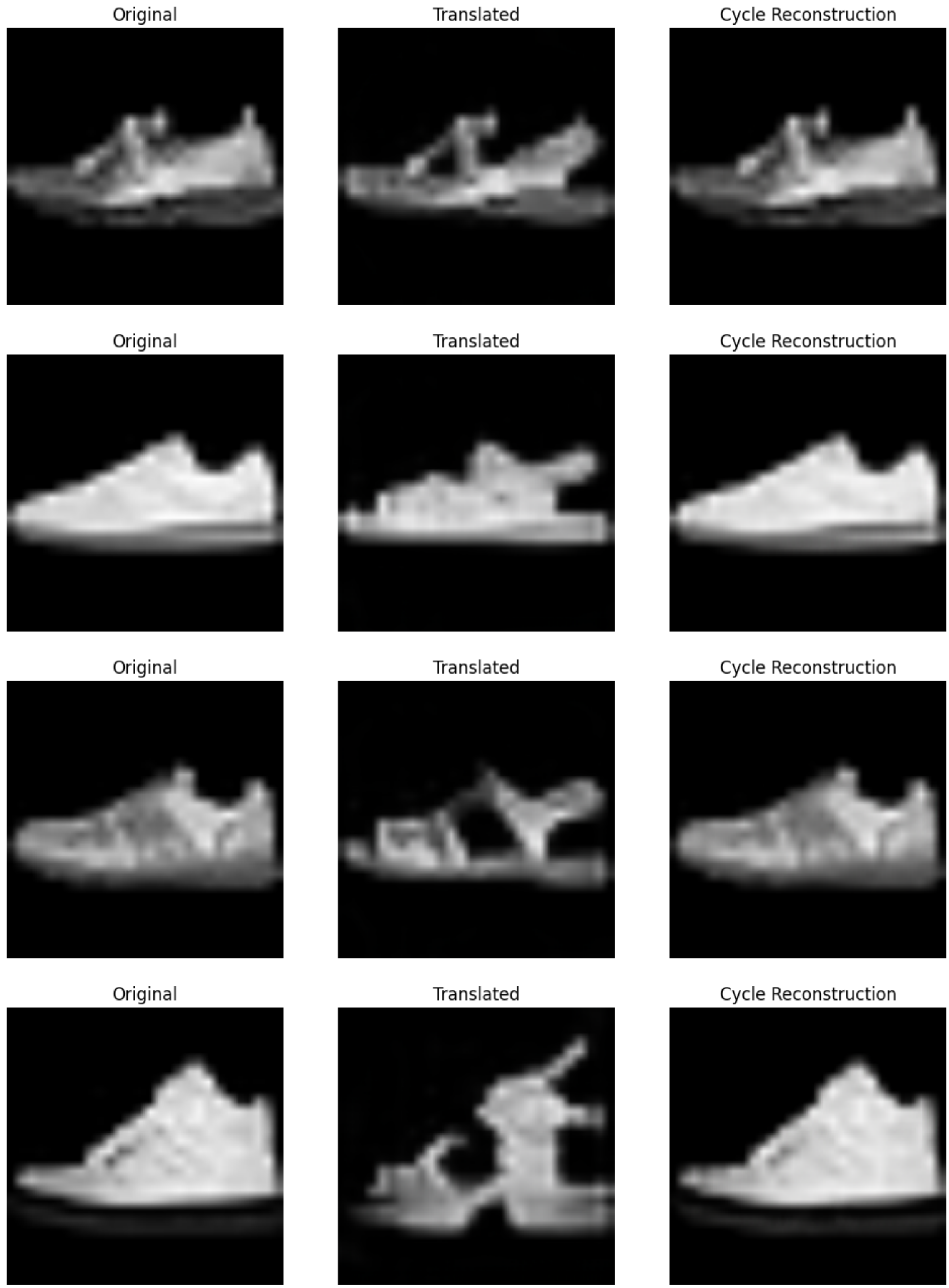

Fig. 9: Representative examples of successful translations for the Sneakers $\rightarrow$ Sandals direction. Each triplet shows the original sneaker, the translated sandal, and the reconstructed sneaker.

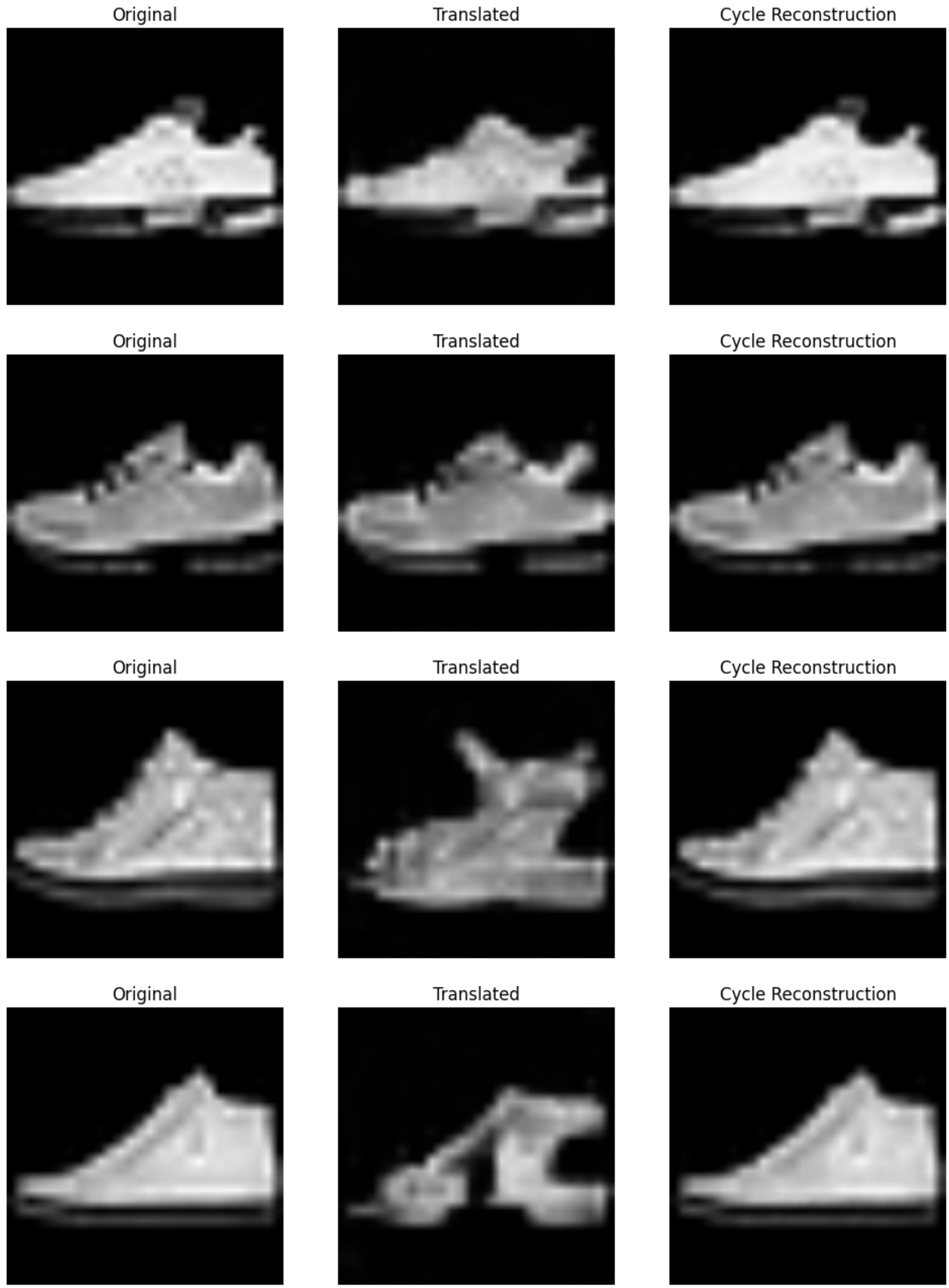

Fig. 10: Representative examples where the generator produced unconventional or ambiguous sandal-like outputs from sneaker inputs. Each triplet shows the original sneaker, the translated sandal, and the reconstructed sneaker.

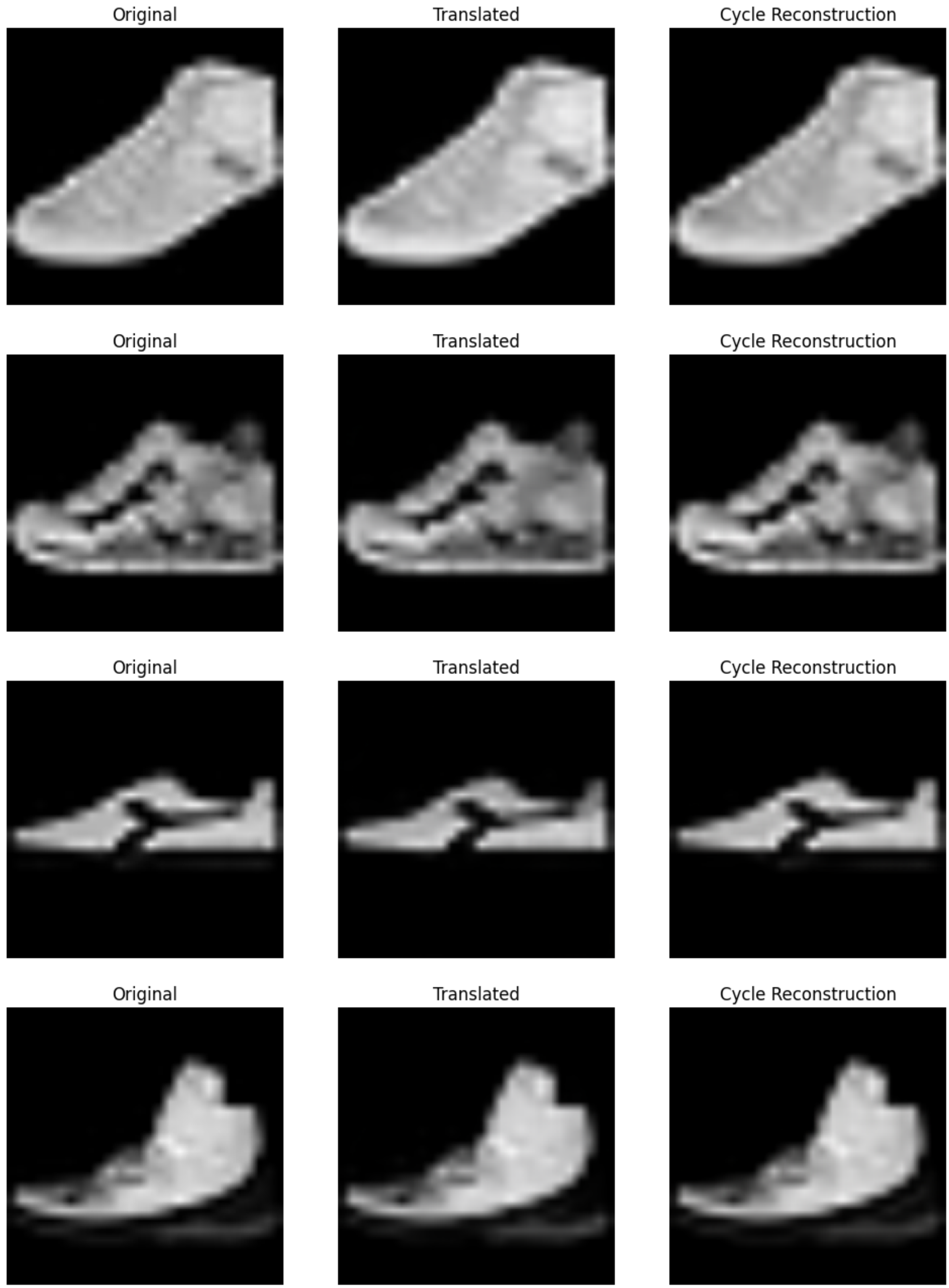

Fig. 11: Representative examples for the Sneakers $\rightarrow$ Sandals direction where the generator failed to perform domain translation, leaving the input sandals nearly unaltered. Each triplet shows the original sneaker, the translated sandal, and the reconstructed sneaker.

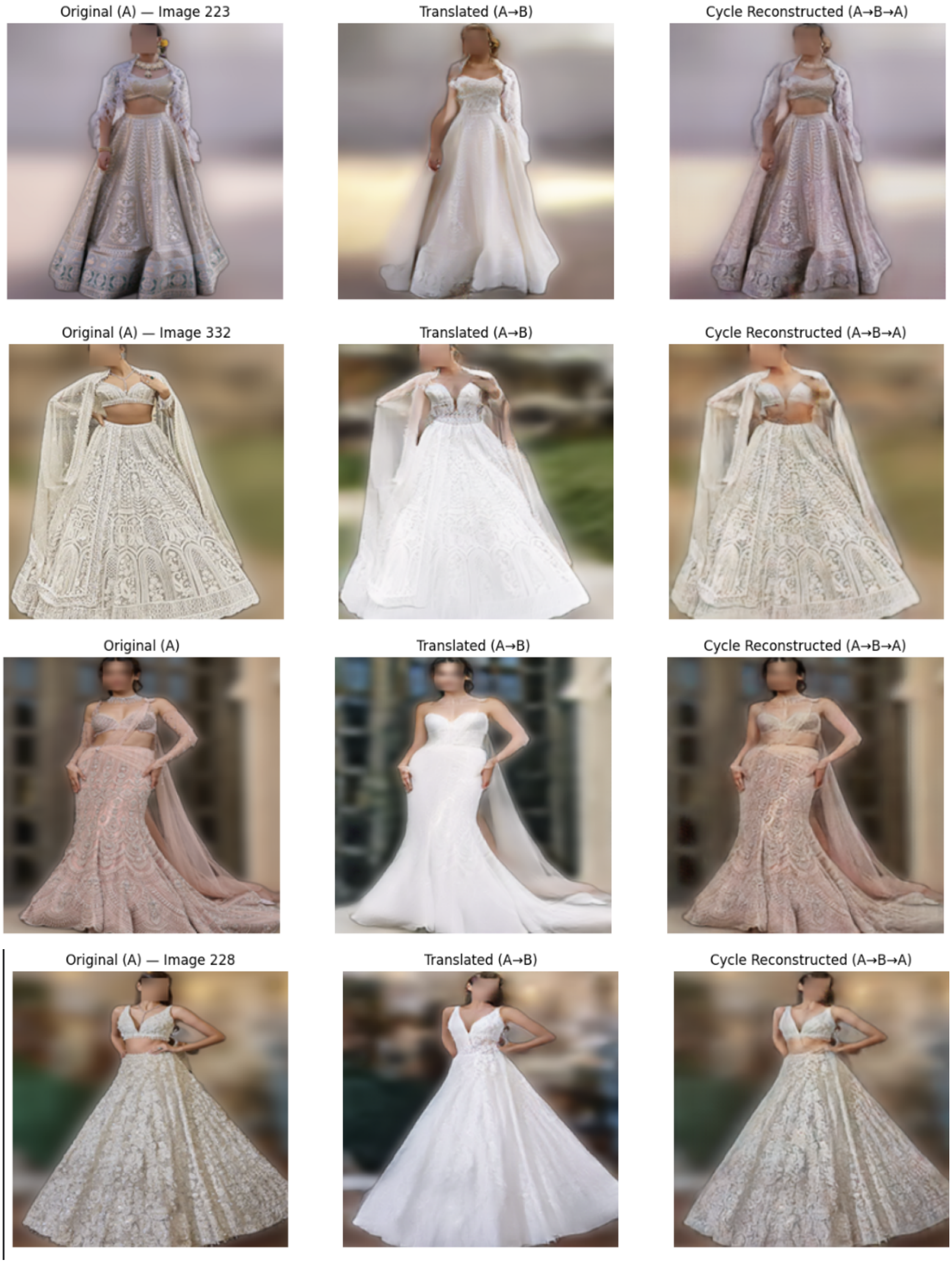

Fig. 12: Representative examples of successful translations from Indian bridal attire to Western couture. The generator effectively captured the main stylistic traits of the Western domain like lighter fabric tones, textural patterns and reduced ornamentation, while preserving the original structure. The reconstructions further confirm strong cycle consistency and preservation of fine details.

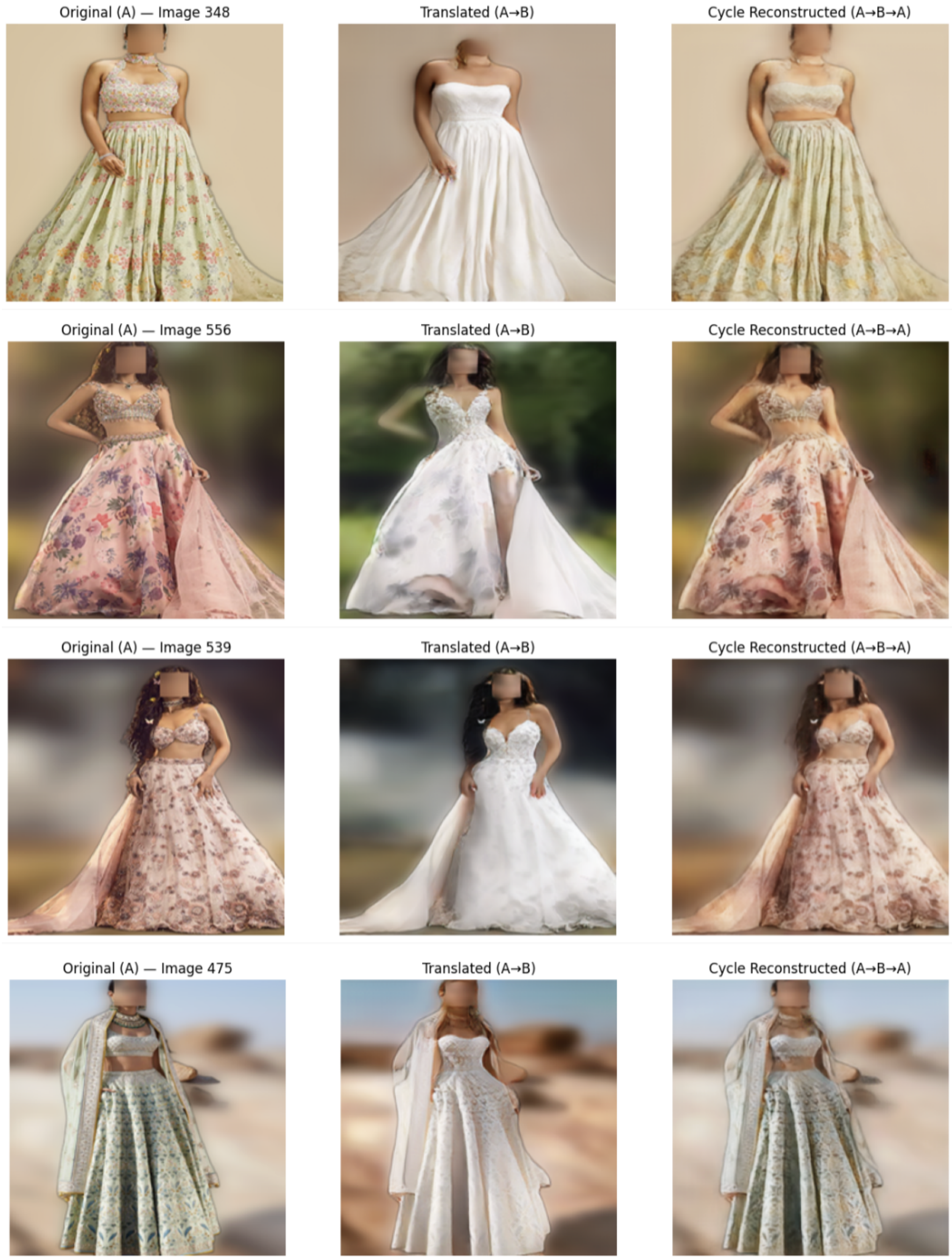

Fig. 13: Representative examples of successful translations from Indian bridal attire to Western couture. The generator effectively captured the main stylistic traits of the Western domain like lighter fabric tones, textural patterns and reduced ornamentation, while preserving the original structure. The reconstructions further confirm strong cycle consistency and preservation of fine details.

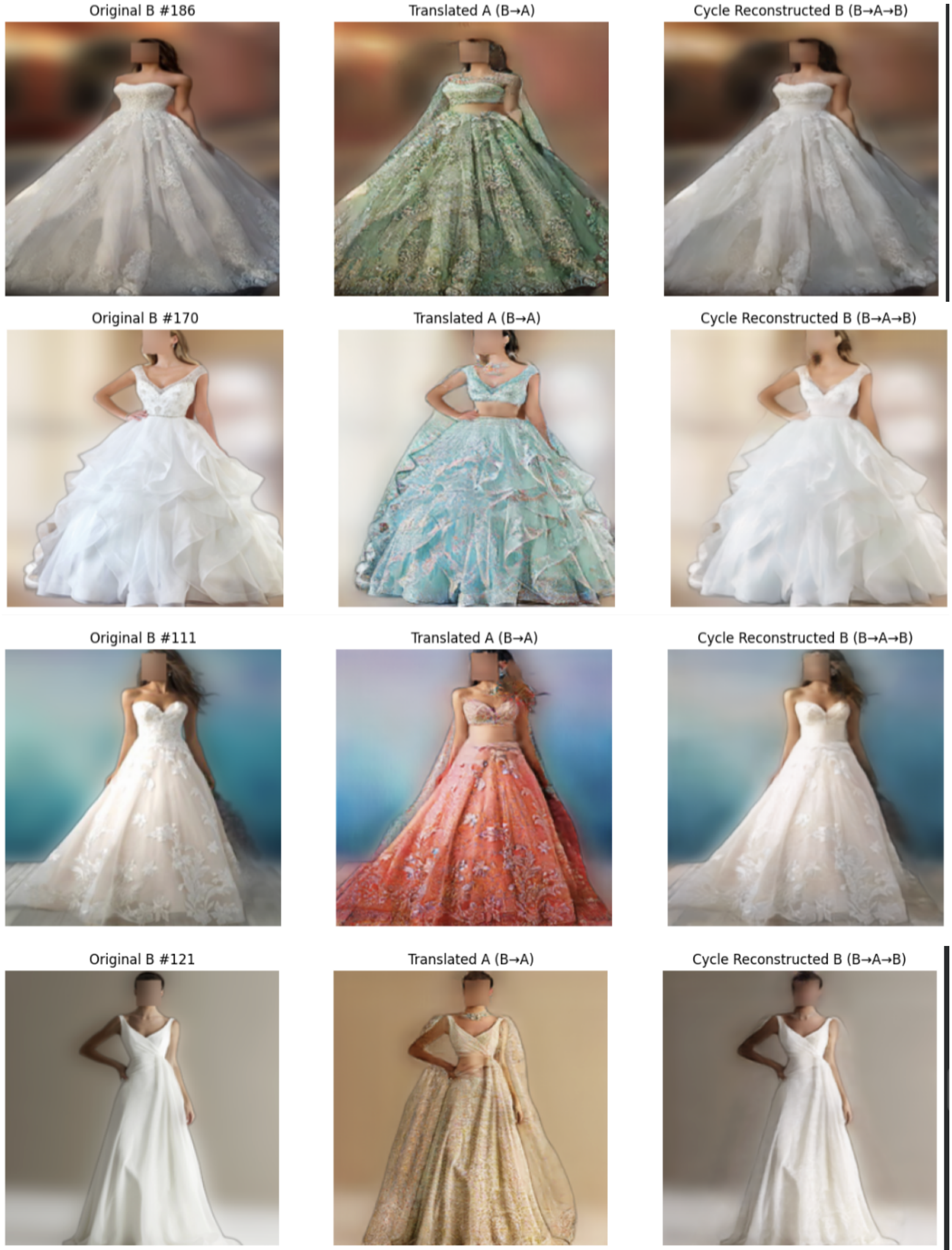

Fig. 14: Representative examples of high-quality translations from Western bridal gowns to Indian couture. The model accurately infused traditional Indian elements such as richer color palettes, intricate embellishment textures, and heavier fabric styles, while preserving the original dress contours. The reconstructions demonstrate strong semantic retention and consistent mapping between domains.

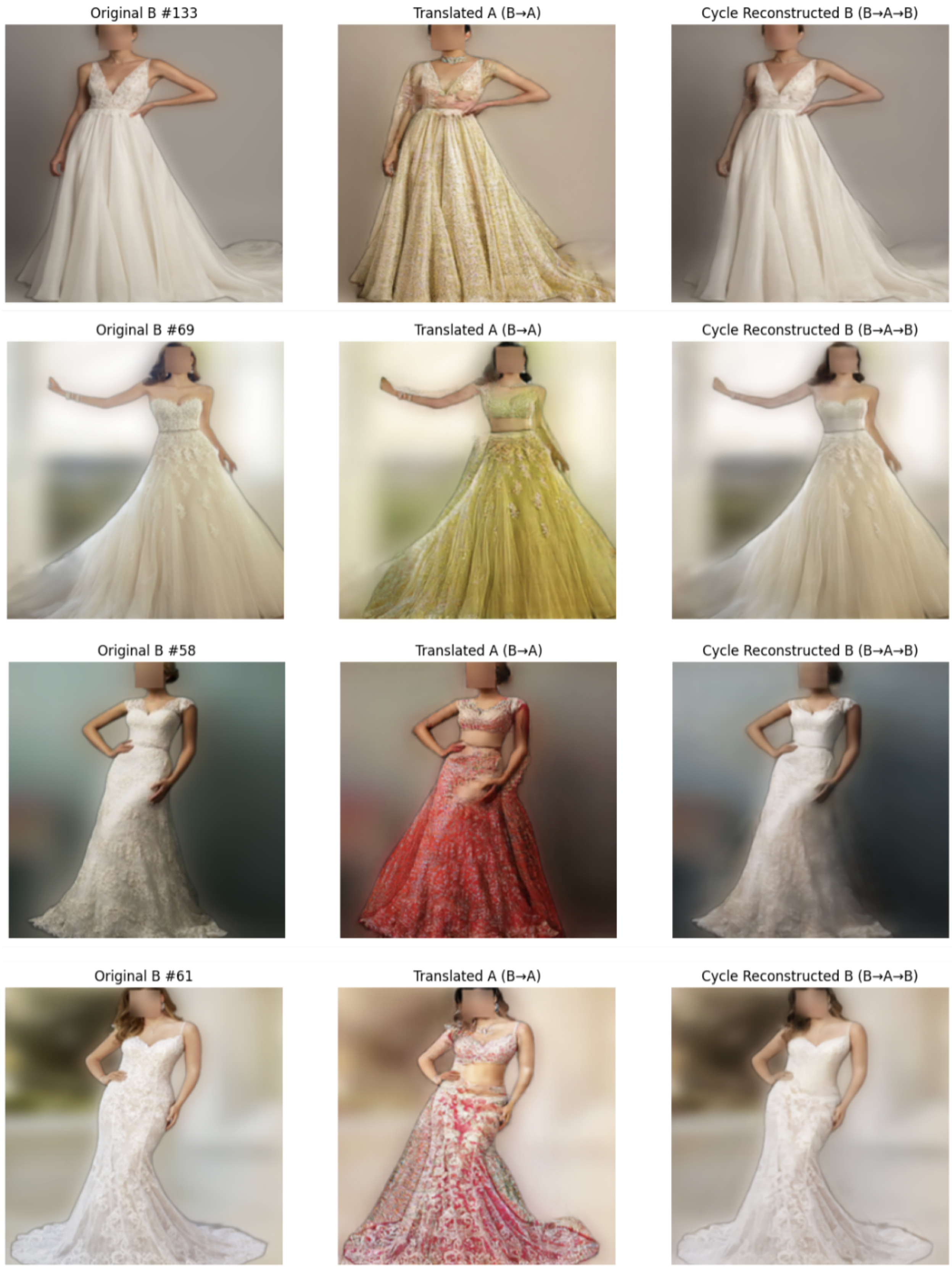

Fig. 15: Representative examples of high-quality translations from Western bridal gowns to Indian couture. The model accurately infused traditional Indian elements such as richer color palettes, intricate embellishment textures, and heavier fabric styles, while preserving the original dress contours. The reconstructions demonstrate strong semantic retention and consistent mapping between domains.

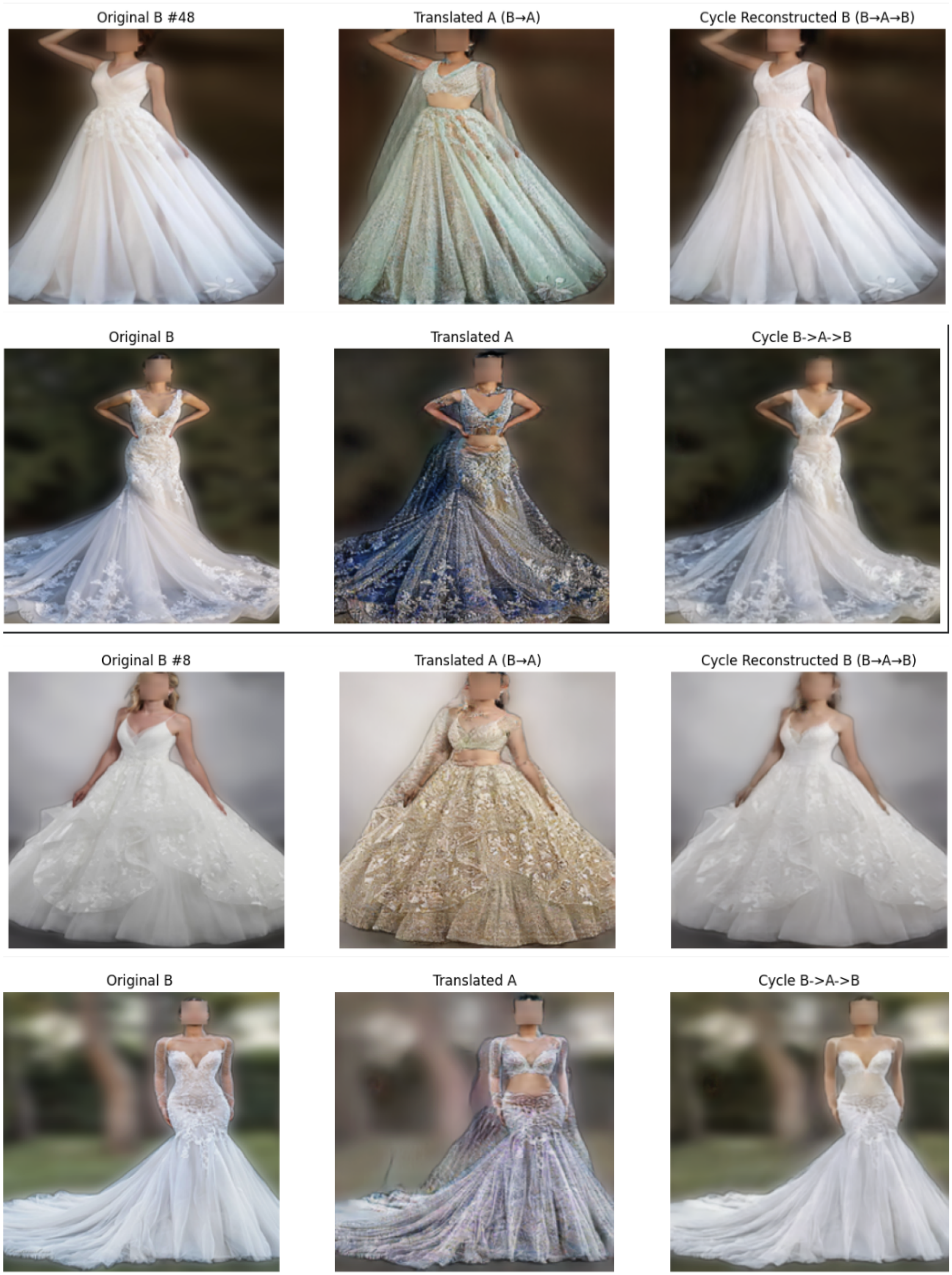

Fig. 16: Representative examples of high-quality translations from Western bridal gowns to Indian couture. The model accurately infused traditional Indian elements such as richer color palettes, intricate embellishment textures, and heavier fabric styles, while preserving the original dress contours. The reconstructions demonstrate strong semantic retention and consistent mapping between domains.