# Unpaired Day-to-Night Image Translation using CycleGAN

Tarun Gangadhar Vadaparthi

## 1 Overview of the Method

Our objective is straightforward but difficult to achieve: transforming street scenes from the day into realistic nighttime scenes (and vice versa) *without* having matched day/night pairs of the same location. We want the model to alter the color temperature and lighting while maintaining the traffic lights, buildings, lanes, and cars in the layout.

We employ a cycle consistent GAN that has two discriminators and two generators. One generator learns from day to night, while the other learns from night to day. The discriminators assess each domain's realism. The discriminators are PatchGANs, which look at small patches and are concerned with local texture like headlight glow, asphalt grain, and sky tone; the generators are small ResNet-9 models, which offer a good trade-off between capacity and speed at 256×256. Training depends on three losses, each of which serves a distinct purpose:

- **Adversarial (LSGAN).** uses least-squares targets for smoother, more stable gradients than BCE, it pushes translated images to resemble real samples in the target domain.

  $$\mathcal{L}_{\text{GAN}}(G_{AB}, D_B) = \mathbb{E}_{b\sim p_B}\big[(D_B(b) - 1)^2\big] + \mathbb{E}_{a\sim p_A}\big[D_B(G_{AB}(a))^2\big],$$

  and analogously for $(G_{BA}, D_A)$.

- **Cycle consistency (L1, $\lambda_{\textbf{cyc}}$=10).** The original should be recovered if we go from point A to point B and back again. This prevents object identity and scene geometry from slipping.

  $$\mathcal{L}_{\text{cyc}} = \mathbb{E}_{a\sim p_A}\big[\|G_{BA}(G_{AB}(a)) - a\|_1\big] + \mathbb{E}_{b\sim p_B}\big[\|G_{AB}(G_{BA}(b)) - b\|_1\big].$$

- **Identity (L1, $\lambda_{\textbf{id}}$=5).** The generator should largely ignore images that are already from the target domain. This lessens needless color changes, particularly in Night-Day.

  $$\mathcal{L}_{\text{id}} = \mathbb{E}_{b\sim p_B}\big[\|G_{AB}(b) - b\|_1\big] + \mathbb{E}_{a\sim p_A}\big[\|G_{BA}(a) - a\|_1\big].$$

**Total loss.**

$$\mathcal{L} = \mathcal{L}_{\text{GAN}}(G_{AB}, D_B) + \mathcal{L}_{\text{GAN}}(G_{BA}, D_A) + \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}} + \lambda_{\text{id}}\mathcal{L}_{\text{id}},$$

with $\lambda_{\text{cyc}}$=10 and $\lambda_{\text{id}}$=5.

**What we really constructed.** Using a luminance heuristic, we created a balanced, unpaired split from the BDD100K (10k subset): TrainA (Day) = 3500, TrainB (Night) = 3500, ValA = 500, ValB = 500. The resized images are 256×256. Using Adam and a linear decay schedule, we trained for five epochs on a Colab Pro T4 (learning rate $2\times10^{-4}$, $\beta = (0.5, 0.999)$). To make runs easily reproducible, the code exports qualitative grids and metrics and saves checkpoints for every epoch.

## 2 Training Details

**Data and split.** In order to iterate rapidly while still covering a variety of scenes, we worked with the BDD100K *images/10k* subset. Since the raw folders lack a time of day label, we used a straightforward luminance heuristic to create Day/Night domains: convert to grayscale, take the global median, and assign darker images to Night ($B$) and brighter images to Day ($A$). The sets were then balanced:

$$\text{TrainA} = 3500,\ \text{TrainB} = 3500, \qquad \text{ValA} = 500,\ \text{ValB} = 500.$$

**Preprocessing and augmentation.**

- **Train:** resize to 1.12× (286), normalize to $[-1, 1]$ (mean=0.5, std=0.5 per channel), random crop to 256×256, and random horizontal flip.
- **Validation:** fixed resize with the same normalization to 256×256.
- To sanity check the heuristic (to catch dusk/dawn leakage), we validated a small number of samples per split.

**Model configuration.**

- **Generators ($G_{AB}$, $G_{BA}$):** ResNet 9 with instance normalization and ReLU/Tanh activations. Downsample to 256 channels, 9 residual blocks, then upsample back to 3×256×256.
- **Discriminators ($D_A$, $D_B$):** PatchGAN critics that output a grid of real/fake decisions, encouraging realistic local texture (lights, asphalt, sky).
- **Image replay buffer:** size 50, so $D$ sees a mix of fresh and slightly older fakes (reduces oscillation).

**Optimization.**

- **Objective weights:** cycle L1 $\lambda_{\text{cyc}} = 10$, identity L1 $\lambda_{\text{id}} = 5$; adversarial uses least squares targets (LSGAN).
- **Optimizer:** Adam with learning rate $2\times10^{-4}$, $\beta = (0.5,\ 0.999)$ for all nets.
- **Schedule:** linear decay over epochs (keeps late training stable).
- **Batch size / epochs:** batch = 2, epochs = 5.

**Implementation notes.**

- Hardware: Colab Pro, Tesla T4 GPU. PyTorch + torchvision.
- DataLoader: shuffle every epoch, `drop_last=True`. On Colab we fall back to single worker loading to avoid the known shutdown warning.
- Checkpoints: generators and discriminators saved at the end of each epoch; periodic sample grids written to disk.
- Reproducibility: fixed random seeds for Python/NumPy/PyTorch; code exports both `state_dict` and TorchScript versions of the generators.

Table 1: Key hyperparameters used in all reported runs.

| Setting | Value |
|---|---|
| Image size | 256×256 |
| Batch size | 2 |
| Epochs | 5 |
| Optimizer | Adam (lr = $2\times10^{-4}$, $\beta = (0.5, 0.999)$) |
| Adversarial loss | LSGAN (MSE targets) |
| Cycle loss | L1, $\lambda_{\text{cyc}} = 10$ |
| Identity loss | L1, $\lambda_{\text{id}} = 5$ |
| Discriminator type | PatchGAN |
| Replay buffer | 50 images |
| Augmentation (train) | resize → random crop → flip |
| Normalization | to $[-1, 1]$ (per channel mean=0.5, std=0.5) |

## 3 Quantitative Evaluation

Both directions improve steadily over epochs: SSIM and PSNR climb, while LPIPS drops. By the final checkpoint we see consistent structure preservation (lanes, vehicles, façades) alongside the intended change in illumination. The two directions are close; Night→Day edges up slightly on SSIM, Day→Night is a touch better on LPIPS.

**Trends over epochs.** On a fixed validation subset, SSIM rises from early epoch values into the low–mid 0.8 range by the last epoch, PSNR moves into the mid-20 dB range, and LPIPS falls into the $\approx 0.18$–$0.19$ band. This joint movement (SSIM/PSNR up, LPIPS down) is a good sign that the model is both *preserving structure* and *producing perceptually closer reconstructions* after a full cycle.

**Best checkpoint selection.** We select the final model using a simple two step rule:

$$e^{\star} = \arg\min_{e} \frac{\mathrm{LPIPS}_{A\to B\to A}(e) + \mathrm{LPIPS}_{B\to A\to B}(e)}{2},$$

and if two epochs tie on mean LPIPS, we pick the one with higher mean SSIM across both directions. This favors perceptual realism while still rewarding structural fidelity.

**Distribution view (not just means).** The per-image histograms (A→B and B→A) show tight clusters with short tails. Most samples land around SSIM 0.80–0.86 and LPIPS 0.18–0.22. The few low-SSIM/high-LPIPS outliers typically correspond to difficult cases such as strong glare, very dark frames, or scenes dominated by uniform sky/walls where deconvolution artifacts are most visible.

**Protocol details (for reproducibility).** We evaluate up to 500 images per domain at a fixed 256×256 resolution, using the same normalization as training. Cycle metrics are computed on the round-trip reconstructions with a consistent color channel order. We keep the validation subset fixed across epochs so that the curves reflect *training progress*, not sampling noise.

**What these metrics do and don't capture.** Cycle SSIM/PSNR/LPIPS are useful stability probes in the *unpaired* setting: they check whether the model can change style and recover the original content. They are not a substitute for paired ground truth evaluation. To complement them, one can report distributional scores (e.g., FID/KID) between translated images and the target domain; those capture how well the translations match the *target style* even without pairs.

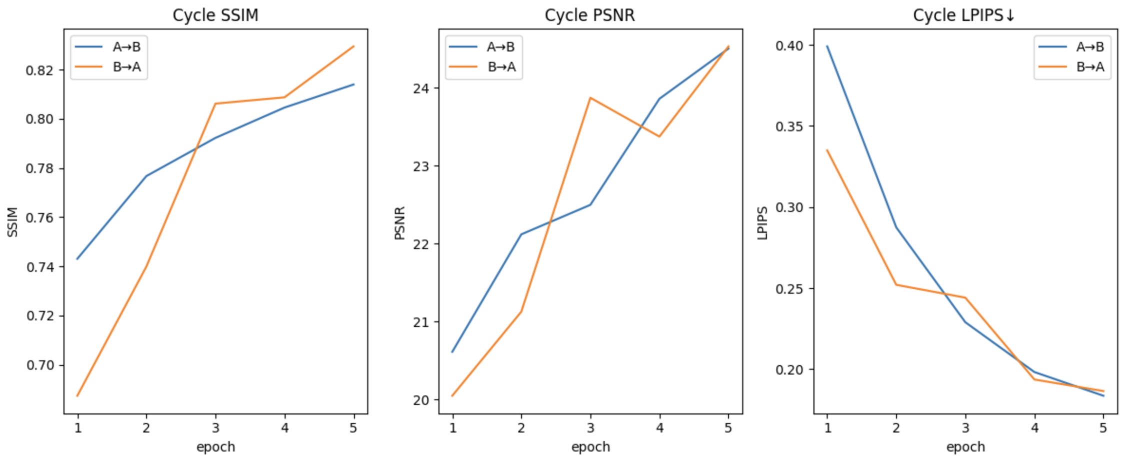


Figure 1: Metric trends across epochs for the fixed validation subset (SSIM/PSNR up, LPIPS down).

# 4 Qualitative Evaluation

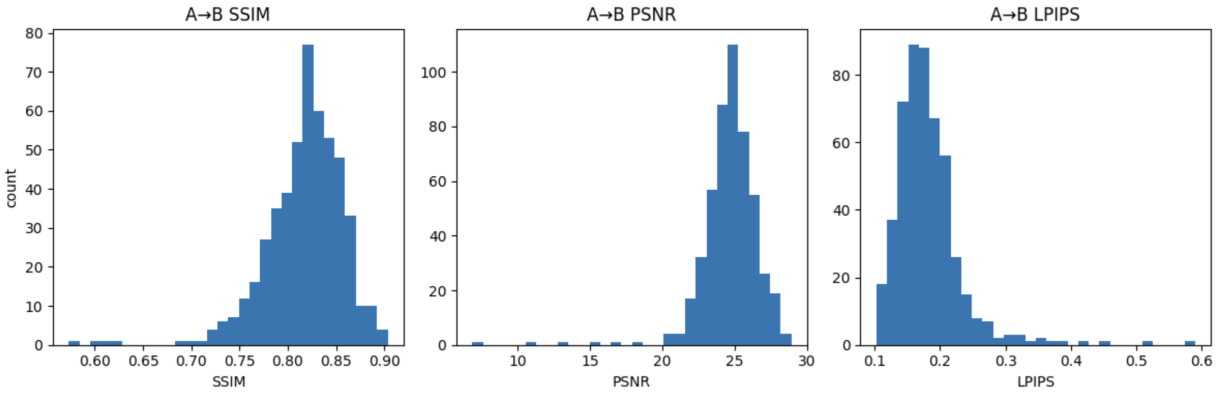


Figure 2: A→B (Day→Night): validation distributions for SSIM, PSNR, and LPIPS.

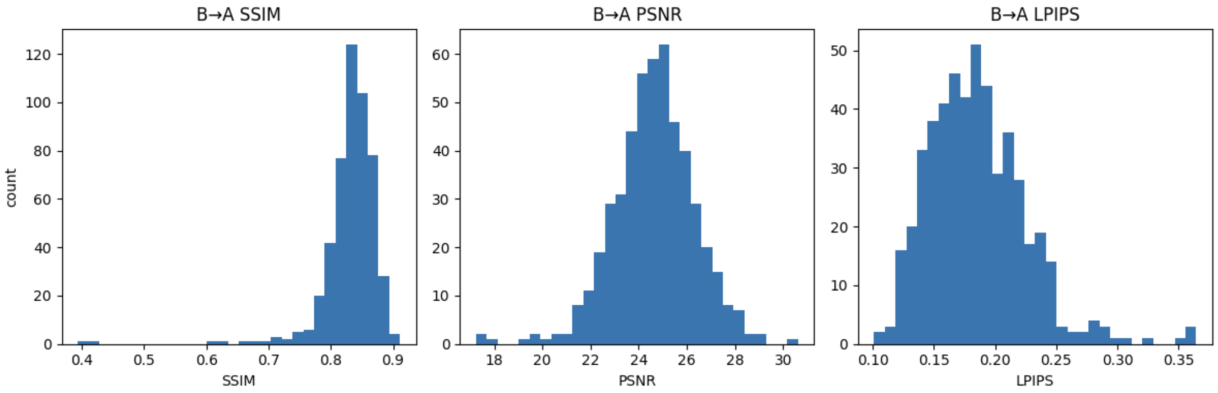


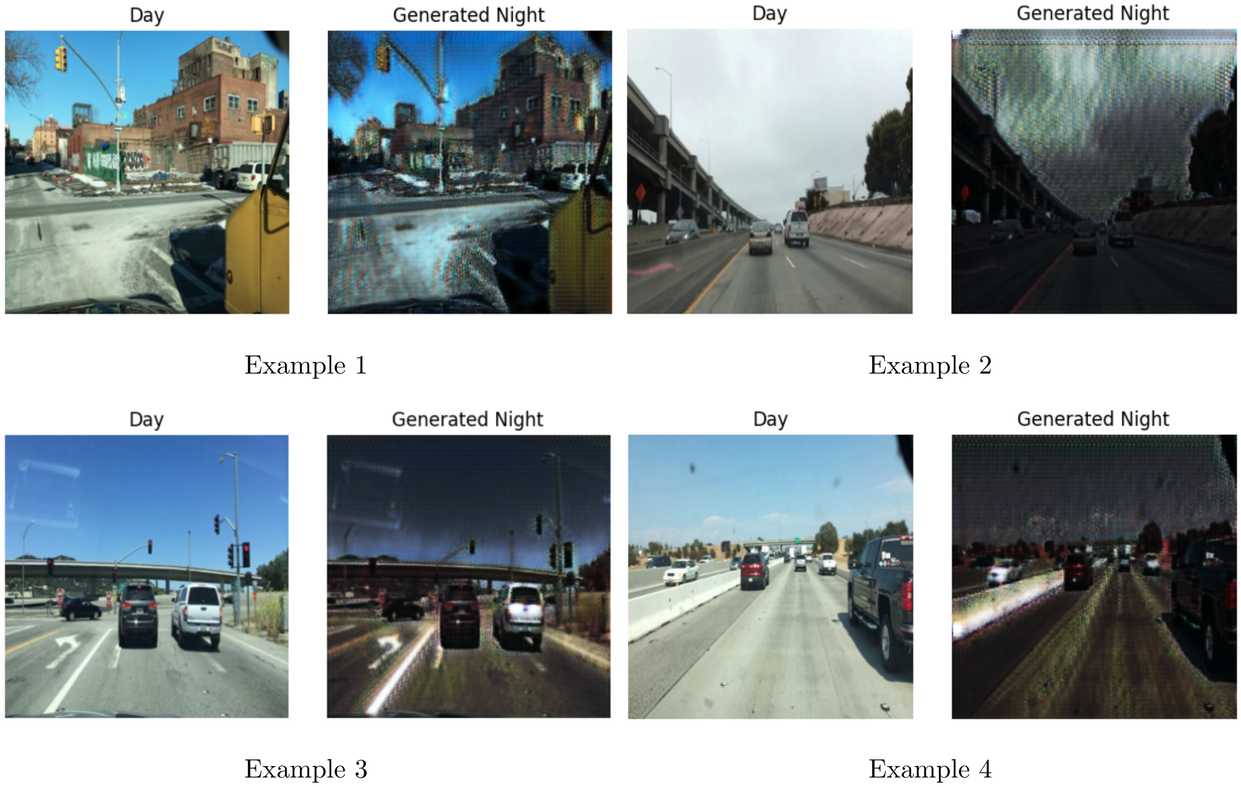
Figure 3: B→A (Night→Day): validation distributions for SSIM, PSNR, and LPIPS.

**What to look for (detailed). Day→Night.**

- *Global tone & exposure:* overall brightness should drop; mid tones compress while blacks deepen, but lane markings and road edges remain legible.
- *Color temperature:* sky and pavement shift cooler (blue/cyan); sodium/LED lights, brake lights, and signs keep their hue (no gray wash).
- *Highlights and reflections:* small specular spots (headlights, windows) become brighter and slightly bloom, without bleeding across edges.
- *Geometry consistency:* poles, façades, traffic lights, and car silhouettes stay aligned with the day image—no bending or stretching.
- *Shadows:* darker regions form in plausible areas (under vehicles, curb lines) while shadow *directions* remain consistent with the scene.

**Night→Day.**

- *Exposure lift:* shadows open up, texture appears on façades and pavement; whites do not clip (no blown sky/road).
- *Color temperature:* scene warms slightly; vegetation regains green, façades regain natural hue, sky moves toward pale blue/gray.
- *Edge fidelity:* high contrast edges (wires, poles, car edges) stay crisp—no halos or ringing around them.
- *Noise behavior:* sensor like night grain should *decrease* after translation.



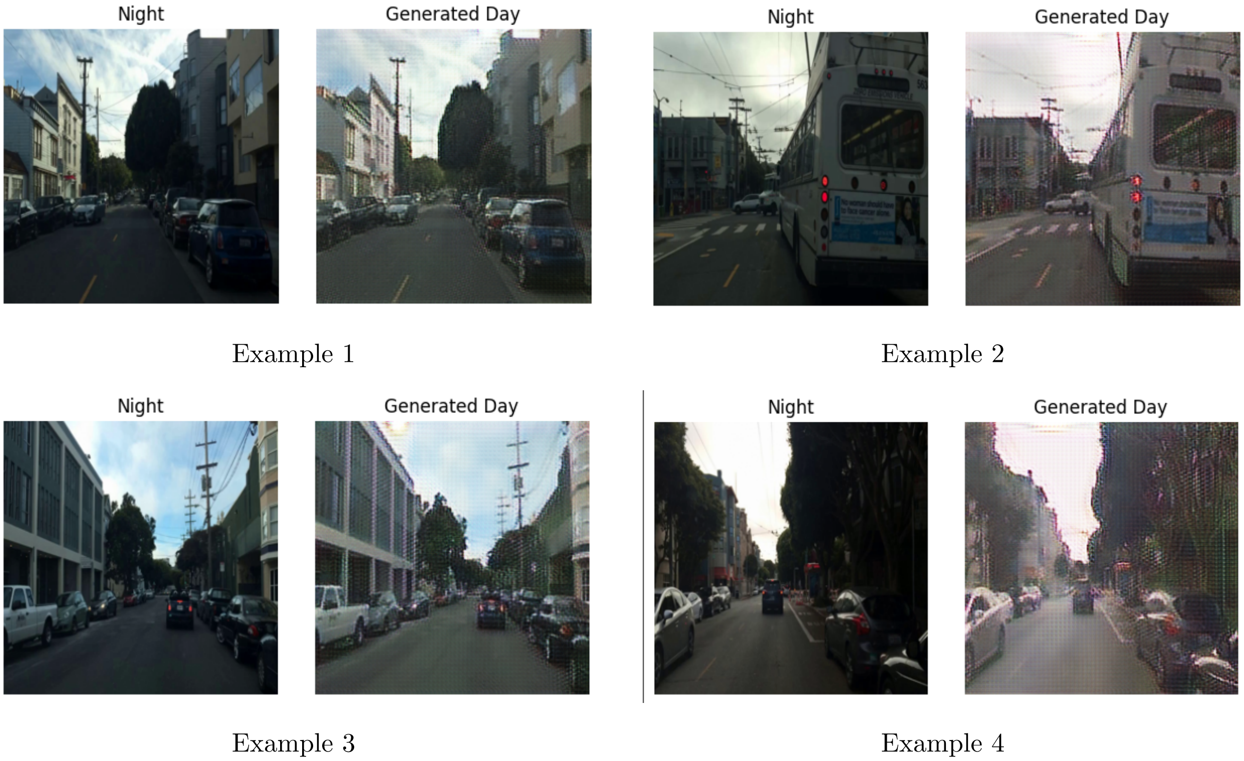
Example 1

Example 2

Example 3

Example 4

Figure 4: Four Day→Night comparisons (left: input Day, right: generated Night).



Example 1

Example 2

Example 3

Example 4

Figure 5: Four Night→Day comparisons (left: input Night, right: generated Day).

# 5 Analysis and Observations

**Learning behaviour.** Across 5 epochs the model improves steadily: SSIM and PSNR go up while LPIPS goes down (see epoch curves). There is a small wobble on Night→Day PSNR around epoch 4, which is typical for adversarial training, but the final checkpoint is consistently the best across all three metrics.

**Distributions, not just means.** The histograms give a fuller picture: A→B (Day→Night) has a tight SSIM cluster around 0.80–0.86 and LPIPS concentrated near 0.18–0.22 with a light right tail. B→A (Night→Day) is similarly compact but a touch broader on PSNR, suggesting a few harder cases. Outliers in both directions correspond to images with strong glare, heavy occlusions, or extreme sky/road dominance (very large uniform regions).

**Direction asymmetry.** Night→Day has slightly higher SSIM on average, while Day→Night often scores a bit lower LPIPS. Intuitively, brightening night scenes tends to reveal structure that is already present, whereas synthesizing night textures introduces additional high-frequency detail (headlights, neon, starry/noisy skies) that PatchGAN rewards perceptually but can affect SSIM.

**What the visuals show.** Qualitatively, the model changes illumination and color temperature convincingly while preserving geometry: lane boundaries, building edges, poles, and vehicles stay in place. Failure cases match the tails in the histograms: mild tint drift, moiré/speckle in sky or walls, and occasional blotchy shadows in large flat areas.

**Why the artifacts appear.**

- **Checkerboard/moire in flat regions.** Transposed convolutions at $256^2$ with small batch sizes can leave grid patterns, and PatchGAN pushes high frequency detail.
- **Tint drift.** With a weak identity term, the generator sometimes "over edits" colors, especially when the input already looks close to the target domain (dusk or dawn).
- **Domain noise.** The luminance based split mixes twilight images into both sets, which blurs the target style and makes training targets less consistent.

**Data and compute constraints.** We intentionally used the BDD100K *10k* subset and trained for 5 epochs to stay within Colab Pro limits (batch 2 on a T4). This budget is sufficient to establish a solid baseline and reproducible trends, but the artifacts indicate the model has not yet fully converged.

**Takeaways.**

- The cycle objective effectively preserves structure; geometry stays stable across both directions.
- Discriminator feedback captures believable night/day textures, but flat regions expose deconvolution artifacts.
- A cleaner Day/Night split and modest training extensions are likely to improve both the histogram tails and the visual crispness.

# 6 Recommendations for Future Work

Below are concrete, low effort upgrades first, followed by medium and stretch goals. They target the artifacts we observed (tint drift, moiré in flat regions, occasional blotchy shadows).

- **Cleaner domains.** Replace the luminance split with official `timeofday` labels. Keep the sets balanced; remove obvious dusk/dawn from both sides.
- **Identity tuning.** Raise $\lambda_{\text{id}}$ from 5 to 7–10 to curb over-editing in Night→Day (less washed sky/signs).
- **TTUR.** Use different learning rates: $lr_G = 2{\times}10^{-4}$, $lr_D = 1{\times}10^{-4}$ to stop $D$ from dominating and reduce speckle.

- **Upsampling fix.** Replace transposed convolution with *nearest neighbor upsample* + $3\times3$ *conv* to reduce checkerboard/moire.
- **Light color jitter.** Small brightness/contrast/saturation jitter during training to broaden style coverage without breaking geometry.
- **Longer training.** Extend to $20-50$ epochs with the same linear decay; evaluate every 5 epochs and keep the best LPIPS/SSIM.

**Expected impact.** Cleaner splits and identity tuning address color drift; TTUR and upsample+conv reduce moiré; more epochs and higher resolution sharpen textures. Together these should tighten the metric histograms (fewer tail outliers) and visibly improve sky/wall regions without harming structure.

## 7 Conclusion

We set up and trained a compact CycleGAN to translate between Day and Night street scenes without paired examples. The model changes illumination and color temperature while keeping road geometry and object layout intact. On a balanced BDD100K 10k split at 256×256, cycle SSIM and PSNR improve steadily over five epochs and LPIPS decreases, landing around 0.82 SSIM, 6 dB PSNR, and 18–0.19 LPIPS at the final checkpoint. The histogram views show most images clustered with a few understandable tails (glare, flat skies, heavy occlusions).

Visually, Day→Night darkens exposure and cools tone; Night→Day restores brightness and warmth, with structure preserved. Remaining artifacts—mild tint drift and moiré/speckle on large flat regions—are consistent with small batch adversarial training at this resolution.

The path forward is clear and practical: use official time of day labels for cleaner domains, raise the identity weight, adopt TTUR, switch deconvolution to upsample+conv, and train longer (20–50 epochs). These adjustments should tighten the metric distributions and noticeably clean up flat regions without compromising geometry.

## References

[1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proc. ICCV*, 2017.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In *Proc. CVPR*, 2017.

[3] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley. Least Squares Generative Adversarial Networks. In *Proc. ICCV*, 2017.

[4] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proc. CVPR*, 2018.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004.

[6] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell. BDD100K: A Diverse Driving Video Database for Heterogeneous Multitask Learning. In *Proc. CVPR*, 2020.