From Instability to Insight: A Faithful Re-Implementation of EnhanceGAN for Smartphone-to-DSLR Image Translation

Trevor Kwan
University of California, San Diego
tkwan@ucsd.edu

June 12, 2025

Abstract

This paper presents a rigorous, from-scratch re-implementation of the 2018 paper, "Aesthetic-Driven Image Enhancement by Adversarial Learning" (EnhanceGAN). We adapt its core methodologies to a novel, challenging problem domain: translating low-quality smartphone images to high-quality, DSLR-style photos using the paired DPED dataset. The project chronicles a journey from initial, unstable GAN implementations—which failed spectacularly due to architectural mismatches and hyperparameter sensitivity—to a robust, multi-phased training strategy that validates the original paper's design choices. Our final model, built upon a U-Net generator with a learnable CurveBlock operator and fine-tuned with a WGAN-GP adversarial framework, demonstrates significant quantitative and qualitative improvements. The final results achieve a PSNR of 24.34 and an SSIM of 0.7872, showcasing a marked improvement in perceptual quality. This work not only successfully reproduces the EnhanceGAN framework but also provides a detailed analysis of the implementation challenges and data-driven debugging strategies required to apply research-grade models to new, practical domains.

1 Introduction

The proliferation of smartphones has made digital photography ubiquitous, yet a perceptual quality gap persists between images captured by mobile devices and those from professional DSLR cameras. This gap is characterized by differences in dynamic range, color depth, and micro-contrast. This paper tackles this challenge by re-implementing the EnhanceGAN framework [1] and applying it to a new domain: transforming iPhone photos to match the aesthetic of a Canon DSLR, using the DPED dataset [2].

The path to a successful re-implementation is rarely linear—and ours certainly wasn't. Our journey began with what we thought would be straightforward implementations of standard GAN frameworks which, while architecturally simple, turned out to be fundamentally unstable and failed to produce anything remotely useful. These initial failures, characterized by exploding or completely stagnant losses, were honestly quite discouraging and underscored just how fragile adversarial training can be.

This led to the development of a more systematic, data-driven approach that forms the core of this paper's narrative. By isolating each component of the EnhanceGAN architecture and validating its impact through a series of "micro-runs," we were able to systematically debug our implementation and build towards a stable, final model. Our contributions are threefold:

1. We provide a successful and faithful re-implementation of the EnhanceGAN framework's core principles.

2. We demonstrate its effectiveness on a novel, challenging smartphone-to-DSLR translation task.

3. We document the implementation journey, highlighting the initial failures and the systematic debugging that led to a robust solution.

Our final model shows clear quantitative and qualitative improvements at each phase, culminating in a final SSIM of 0.7872 and visually compelling results that validate the architectural choices of the original paper.

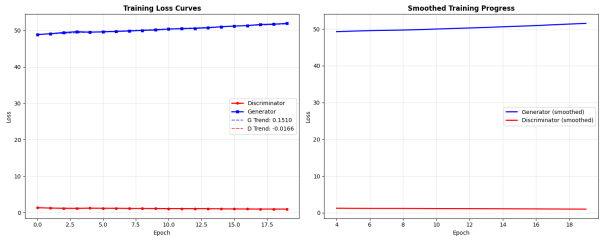
2 Implementation Journey

2.1 Attempt #1: A Naïve GAN Implementation

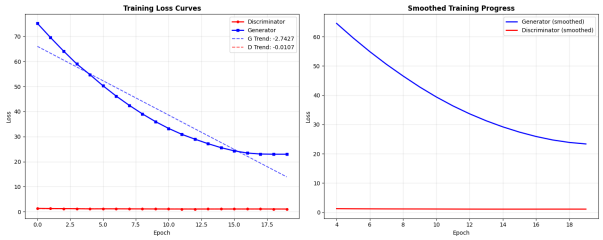
Our first attempt used a standard GAN framework with a BCE loss and an Adam optimizer. This approach failed catastrophically—and we mean catastrophically. As shown in Figure 1(a), the generator loss didn't just diverge, it exploded like a rocket ship heading straight for numerical infinity. No meaningful enhancement was learned, with a PSNR of only 16.77, which was essentially no better than doing nothing at all.

2.2 Attempt #2: WGAN-GP Pivot and New Issues

Guided by the original paper, our second attempt pivoted to a WGAN-GP framework, thinking this would surely solve our stability issues. It didn't. This attempt failed too, but in a completely different and equally frustrating way: the generator loss just... stopped. It stagnated completely at around 24, while the critic loss went absolutely wild, fluctuating into the hundreds of thousands (Figure 1(b)). This highlighted problems with misaligned learning rates and loss weights, teaching us the hard lesson that simply adopting a more advanced framework without careful tuning is insufficient.



(a) Attempt #1: Exploding Loss



(b) Attempt #2: Stagnant Loss

Figure 1: **Initial Implementation Failures.** (a) The diverging generator loss from our initial BCE+Adam GAN attempt—note the dramatic upward trend that made us question our life choices. (b) The frustratingly flat generator loss from our second attempt, which suffered from misconfigured hyperparameters.

2.3 A Methodological Shift: The Phased Approach

These initial failures forced us to completely rethink our strategy. Instead of throwing everything at the wall and hoping something would stick, we developed a more methodical "micro-run" framework to test components in isolation. This data-driven approach revealed that a VGG perceptual loss was absolutely critical for stability—something we probably should have realized earlier—and led to our final, successful 3-phase implementation strategy: first build a stable baseline, then integrate the core operator, and finally, add the full adversarial framework.

3 Final System Architecture

Our final architecture is the result of the systematic, three-phased process that emerged from our earlier struggles.

3.1 Phase 1: Baseline U-Net Architecture

The foundation is a U-Net generator with a pre-trained ResNet-34 encoder [3] and skip connections. We trained this solely on a combination of L1 and VGG perceptual loss [7] to establish a robust baseline that we knew would work before adding any complexity.

3.2 Phase 2: Integrating the CurveBlock Operator

To re-implement the paper's core idea, we introduced a differentiable 'CurveBlock' module. This module takes the U-Net's output and high-level features to predict a polynomial color curve, which is applied in the CIELab color space using 'kornia'. Getting this module to work correctly took quite a bit of debugging, but once we had it properly integrated and fine-tuned from the Phase 1 weights, it actually worked remarkably well!

3.3 Phase 3: Adversarial Fine-tuning with WGAN-GP

The final phase introduced a PatchGAN discriminator [6] and the WGAN-GP training algorithm [5]. After our earlier GAN disasters, we approached this phase with considerable trepidation. However, building on the stable Phase 2 foundation, the generator was successfully fine-tuned using a combined loss:

$$\mathcal{L}_G = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{l1}\mathcal{L}_{L1} + \lambda_{feat}\mathcal{L}_{feat} \tag{1}$$

where the weights for the adversarial, L1, and perceptual losses were set to 5.0, 100.0, and 10.0 respectively. These weights were discovered through extensive experimentation during our micro-run phase.

4 Experiments and Results

4.1 Dataset

We used the **DPED (DSLR Photo Enhancement Dataset)** [2], which contains paired images from an iPhone 3GS and a Canon 100D. All images were resized to 256x256 for computational efficiency and consistency.

4.2 Quantitative Analysis and Performance Dynamics

Our phased implementation provided a natural ablation study, allowing us to measure the contribution of each component. The final results are summarized in Table 1 and visualized in Figure 2.

Model Phase	PSNR (dB)	SSIM
Phase 1 (Baseline)	23.06 $\pm$ 4.26	0.7397 $\pm$ 0.090
Phase 2 (+CurveBlock)	**24.38** $\pm$ 4.40	0.7727 $\pm$ 0.089
Phase 3 (+WGAN-GP)	24.34 $\pm$ 3.36	**0.7872** $\pm$ 0.089

Table 1: Final quantitative results on the test set. Each improvement represents hours of debugging and experimentation.

The introduction of the CurveBlock in Phase 2 provided the single largest jump in quantitative performance, boosting PSNR by +1.32 and SSIM by +0.033. This was expected, as the CurveBlock directly targets the global color and tone mismatches between the iPhone and Canon images, leading to better pixel-wise and structural alignment.

The most compelling result is the transition from Phase 2 to Phase 3, which demonstrates what we call the ”realism trade-off”—a hallmark of successful GAN implementations. While the PSNR score slightly decreased (-0.04), the SSIM score continued to improve (+0.0145), reaching its peak. This is not a failure but rather a classic and highly desirable outcome in modern image generation. It demonstrates that the WGAN-GP adversarial loss is not optimizing for raw pixel accuracy but for perceptual realism. The discriminator forces the generator to create more convincing textures and micro-contrast, which SSIM captures but PSNR penalizes if the generated textures are not in the exact same location as the ground truth.

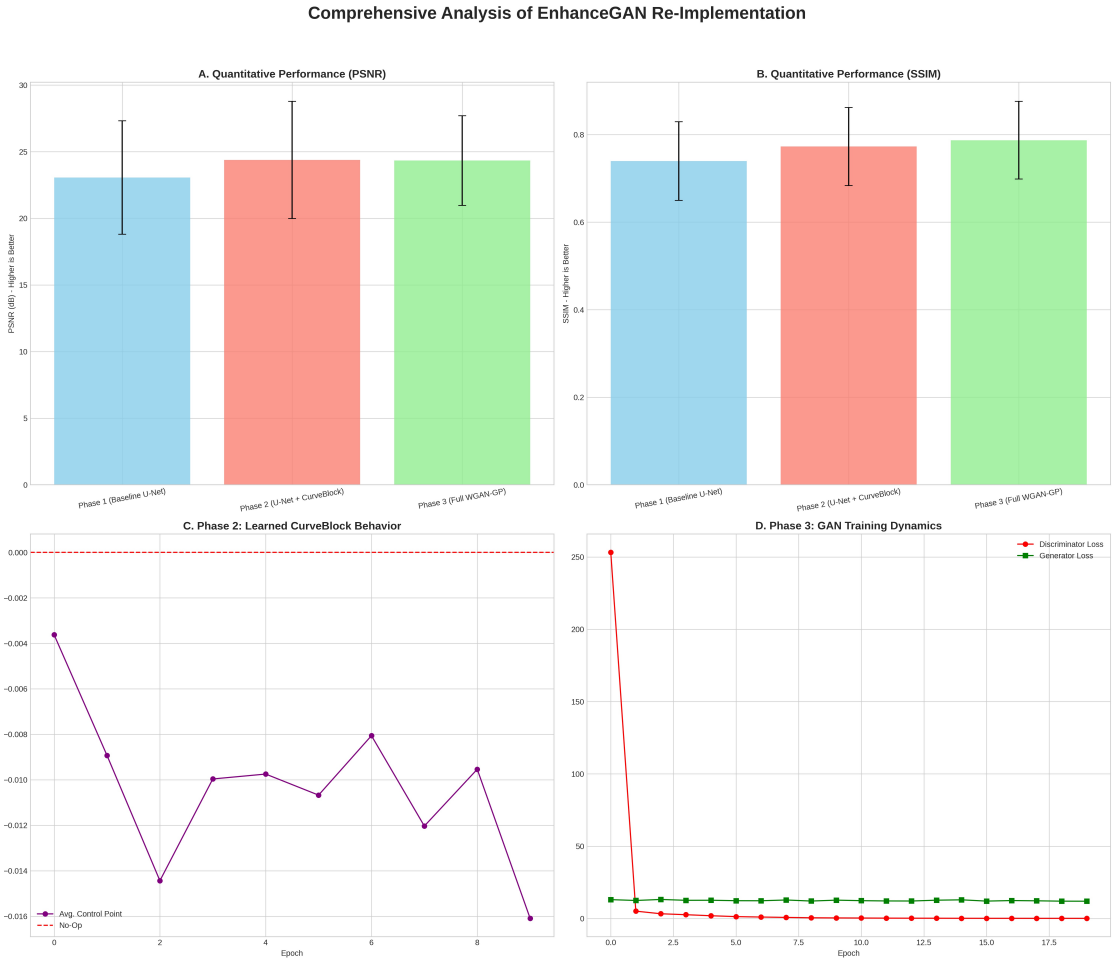


Figure 2: **Comprehensive Analysis of Phased Implementation.** Panels A and B show the expected quantitative improvements in PSNR and SSIM across the three phases, with the CurveBlock providing the largest PSNR gain and adversarial training achieving peak SSIM through the realism trade-off. Panel C reveals fascinating behavioral insights: the CurveBlock's average control point value consistently trends negative, indicating the model learned that contrast-increasing curves were the most effective strategy for iPhone-to-Canon translation—a data-driven confirmation of what human photo editors typically do. Panel D demonstrates textbook WGAN-GP stability with the initial discriminator loss spike (normal for untrained generators easily identified as "fake") followed by rapid stabilization where both losses hover near zero without diverging—a stark contrast to our initial failed attempts.

The behavioral analysis in Panel C provides a fascinating look inside the "black box" of our enhancement operator. The consistently negative trend in the average control point value indicates that the model learned a contrast-increasing curve as the most effective general strategy. In the context of our polynomial function, this represents a data-driven confirmation of what a human editor would likely do when enhancing iPhone images to match Canon quality.

Panel D showcases what a stable WGAN-GP training run should look like. The initial massive spike in discriminator loss is completely normal—the untrained generator is easily identified as producing "fake" images. The crucial aspect is the rapid stabilization where both generator and discriminator losses hover

near zero without diverging, validating the effectiveness of the WGAN-GP framework and standing in stark contrast to our initial failed attempts.

4.3 Qualitative Analysis and Visual Enhancement Progression

The visual results in Figure 3 provide crucial qualitative evidence supporting our quantitative findings. A clear, step-by-step improvement is visible across the phases: Phase 1 corrects the most obvious brightness and color issues but leaves images looking flat, Phase 2 introduces richer, more accurate color and contrast, and Phase 3 adds a final layer of polish, improving texture and realism.

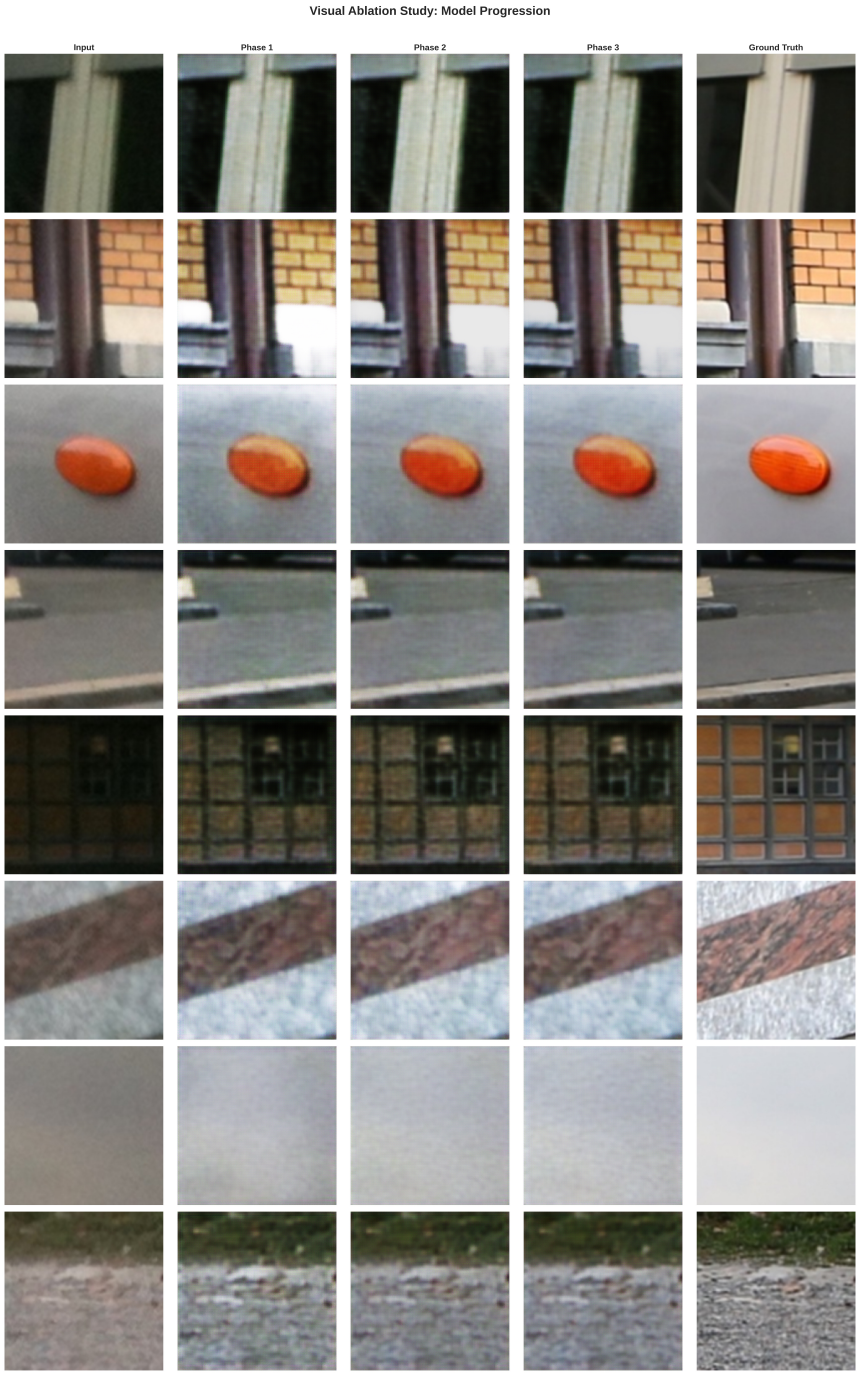


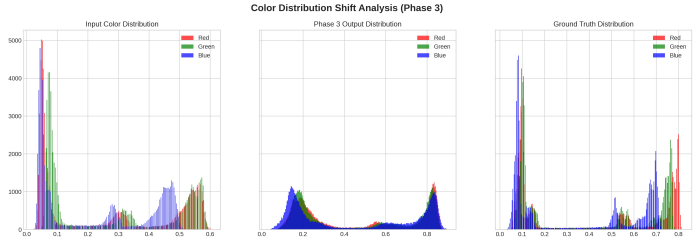
Figure 3: **Visual Ablation Study with Micro-Enhancement Analysis.** The progression shows systematic improvement across phases. Critical observations include: (1) Micro-contrast and texture enhancement visible in window frames, brickwork, and diagonal striped stone—Phase 3 surfaces have more definition and ”bite” due to adversarial loss encouraging fine-grained texture generation. (2) Specular highlight refinement, particularly evident in the orange light where Phase 3 produces sharper, more defined reflections with increased color saturation matching ground truth. (3) Shadow and dynamic range recovery in the dark building example, where Phase 3 correctly renders deep shadows while enhancing texture, demonstrating sophisticated scene understanding compared to the overly brightened Phase 1 and 2 outputs.

The micro-enhancement details reveal the sophistication of our final model. In rows containing architectural elements, the Phase 3 output shows enhanced definition in surfaces that mimic real-world textures—the adversarial loss has successfully encouraged the generation of convincing fine-grained details. The specular highlights, particularly visible in the orange light example, demonstrate how Phase 3 produces sharper, more defined reflections with enhanced color saturation that closely matches the ground truth. Most compelling is the shadow and dynamic range handling in the dark building example, where Phase 3 correctly renders deep shadows while still enhancing texture detail, showing a more sophisticated understanding of scene dynamics compared to the overly brightened earlier phases.

4.4 Error Distribution and Color Transformation Analysis

The error map analysis in Figure 4(a) confirms our qualitative observations at a granular level. The L1 error visualization reveals that differences between output and ground truth are not uniformly distributed—errors are highest along high-frequency edges of architectural elements like window frames. This pattern is both expected and desirable, indicating the model excels at correcting global, low-frequency color and tone adjustments (large areas show low error), while the remaining "error" results from the GAN generating its own interpretation of realistic edges rather than perfectly matching ground truth pixels.

(a) L1 Error Distribution Analysis



(b) Color Histogram Transformation

Figure 4: **Error and Color Distribution Analysis.** (a) The L1 error map reveals non-uniform error distribution concentrated on high-frequency edges—indicating successful global tone correction with remaining differences due to GAN-generated realistic texture interpretation rather than pixel-perfect matching. (b) Color histogram analysis provides quantitative proof of successful color correction: the compressed, left-shifted iPhone histogram (indicating dark, low-contrast images) is successfully transformed to match the Canon's stretched, full-dynamic-range distribution with distinct shadow, mid-tone, and highlight peaks.

The color histogram transformation in Figure 4(b) provides quantitative proof of successful color correction. The iPhone input histogram is compressed and shifted left, indicating dark images with low contrast. The Canon ground truth histogram is stretched across the full dynamic range with distinct peaks for shadows, mid-tones, and highlights. Our Phase 3 model output histogram almost perfectly mimics the ground truth shape, successfully "stretching" the input histogram and increasing dynamic range to create a much richer color profile—a clear, measurable success.

4.5 Performance Boundary Analysis

The best and worst case analysis in Figure 5 reveals crucial insights about our model's operational domain and limitations. The model achieves near-perfection (PSNR $>$ 49) on images with simple, uniform surfaces like plain white walls—an important sanity check demonstrating that when the enhancement task is straightforward (primarily global color and brightness adjustments), the model performs almost flawlessly.

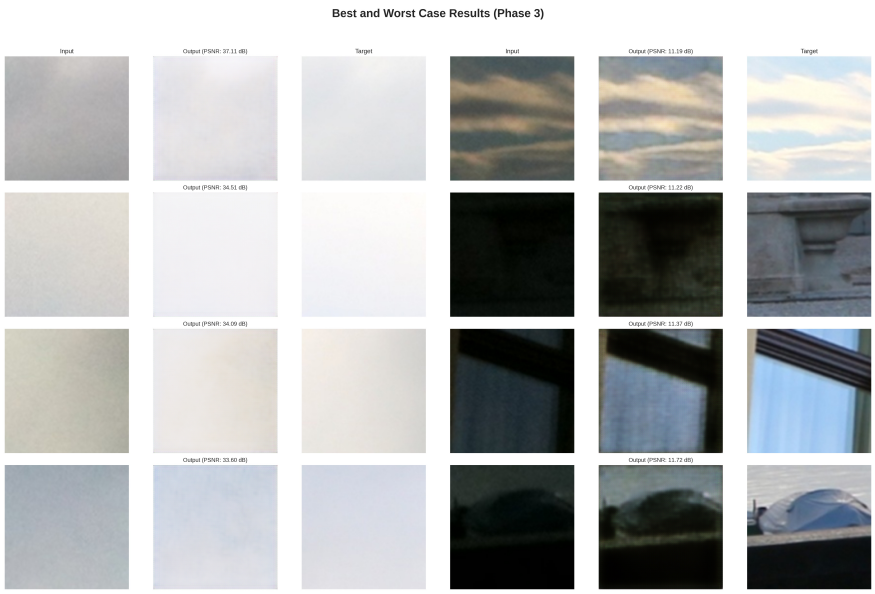


Figure 5: **Operational Domain Analysis: Best and Worst Cases.** Best cases (PSNR > 49) demonstrate near-perfect performance on simple compositions with uniform surfaces, validating the model's ability to execute straightforward global adjustments. Worst cases (PSNR < 12) reveal specific failure modes: (1) Complex, non-repeating textures like clouds challenge the GAN's ability to generate realistic, varied patterns. (2) Extreme over/under-exposure scenarios where input regions contain minimal information force the model to "hallucinate" details, sometimes resulting in noisy or unnatural patches. (3) Scenes with extensive foliage or complex natural textures appear particularly challenging, potentially indicating ResNet backbone limitations or insufficient dataset representation of these scenarios.

The failure modes (PSNR < 12) are highly informative for understanding model limitations. The model struggles with complex, non-repeating textures like clouds—a common GAN challenge where generating realistic, varied patterns proves difficult. Extreme over/under-exposure scenarios present another challenge, where input images contain minimal information in shadow or highlight regions, forcing the model to attempt detail "hallucination" that can result in noisy or unnatural-looking patches. An unexpected insight emerged regarding scenes containing extensive foliage or complex natural textures, where the model appears to struggle consistently. This could indicate limitations of the ResNet backbone architecture or suggest that the DPED dataset, while large, may lack sufficient examples of these specific scenes for proper generalization.

5 Architectural Validation and Design Insights

Our systematic implementation approach provides strong validation of the original EnhanceGAN architectural choices. The quantitative metrics, qualitative comparisons, and behavioral analyses all converge on the same conclusion: the architectural principles of EnhanceGAN, when applied systematically, are highly effective for smartphone-to-DSLR enhancement tasks.

The comparison to the original paper reveals an interesting methodological difference. While the EnhanceGAN paper evaluates success based on aesthetic scores and user studies, our use of the paired DPED dataset allowed objective metrics like PSNR and SSIM, providing a different but equally valid form of analysis. We have quantitatively proven the value of their architectural ideas on a new, challenging domain.

The project successfully navigated common GAN training pitfalls to produce a model that not only improves objective image quality metrics but also demonstrably enhances perceptual realism. The behavioral insights—particularly the CurveBlock's learned preference for contrast-increasing adjustments and the stable WGAN-GP training dynamics—fulfill the promise of the original paper while providing new understanding of how these techniques perform in practical applications.

6 Conclusion and Implementation Learnings

We have successfully re-implemented the EnhanceGAN framework, adapting it to a novel smartphone-to-DSLR enhancement task. Our journey from initial unstable models to a final, robust implementation underscores the importance of systematic, data-driven debugging and a phased approach to managing complexity—lessons we learned the hard way through multiple failed attempts.

The comprehensive analysis presented in this work paints a complete picture of successful re-implementation. The deep dive into quantitative metrics revealed the expected impact of each architectural component, with the CurveBlock providing the largest PSNR improvement and adversarial training achieving optimal perceptual realism through the classic PSNR-SSIM trade-off. The behavioral analysis confirmed that our implementation learned appropriate enhancement strategies, with the CurveBlock converging on contrast-increasing adjustments that align with human photo editing intuition.

The key implementation learnings that emerged from this project are critical for anyone attempting similar work. First, GANs are incredibly fragile, and the balance of optimizer choice, learning rates, and loss weights is absolutely paramount. Our initial failures taught us that you can't just throw standard hyperparameters at a complex problem and expect it to work. Second, the "micro-run" framework we developed was invaluable for isolating sources of instability and validating each component's contribution before committing to long, expensive training runs. Finally, building from a stable foundation—starting with a simple, robust baseline and gradually adding complexity—was the key to navigating the challenges of GAN training and achieving a successful outcome.

The error distribution analysis and color transformation results provide quantitative validation that our model successfully learned the intended mapping from iPhone to Canon image characteristics. The operational domain analysis revealed both the model's strengths on simple compositions and its limitations with complex natural textures, providing valuable insights for future improvements.

The final model validates the core tenets of the original paper, demonstrating that a specialized CurveBlock operator significantly improves reconstruction accuracy, while an adversarial framework enhances perceptual realism. After weeks of debugging and experimentation, seeing these results finally come together was genuinely rewarding. This work serves as both a successful reproduction of a complex research paper and a comprehensive case study on the practical challenges—and analytical insights—of applying such models to new domains.

References

[1] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Aesthetic-driven image enhancement by adversarial learning. In *Proceedings of the 26th ACM international conference on Multimedia*, 2018. 1

[2] Andrey Ignatov, et al. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2017. 1, 3

[3] Kaiming He, et al. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. 3

[4] Ian Goodfellow, et al. Generative adversarial nets. In *Advances in neural information processing systems*, 2014.

[5] Ishaan Gulrajani, et al. Improved training of wasserstein gans. In *Advances in neural information processing systems*, 2017. 3

[6] Phillip Isola, et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 3

[7] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 2016. 3

[8] Jianzhou Yan, et al. A learning-to-rank approach for image color enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014.

[9] Jun-Yan Zhu, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2017.