EdgeConnect+: Adversarial Inpainting with Edge and Color Guidance

Abhinay Kotla, Sanjana Ravi Prakash
University of Texas at Arlington
axk5827@mavs.uta.edu, sxr8375@mavs.uta.edu

Abstract

We present EdgeConnect+, an enhanced deep learning framework for image inpainting that leverages both structural and chromatic guidance to produce perceptually realistic reconstructions. Building upon the original EdgeConnect model, our approach introduces a three-stage pipeline: (1) an edge generation network (G_1) that predicts structural contours from masked grayscale inputs, (2) a color guidance stage that constructs a low-frequency appearance prior from unmasked regions, and (3) a guided image completion network (G_2) that synthesizes the final inpainted output conditioned on both structural and chromatic cues. By combining precise edge maps with smooth color information, EdgeConnect+ is designed to improve contextual coherence and visual fidelity in complex inpainting scenarios. Preliminary experiments and qualitative results suggest that our method has the potential to outperform the original EdgeConnect and other structure-only baselines, although full training and benchmarking remain ongoing.

1. Introduction

Image inpainting is a fundamental task in computer vision that aims to reconstruct missing or corrupted regions in images while preserving visual realism and structural coherence. It has numerous practical applications, including photo restoration, object removal, and creative image editing. Traditional approaches, such as diffusion-based and patch-based methods, often struggle with large missing regions or complex structures, typically producing blurry or semantically inconsistent outputs.

Recent advances in deep learning, particularly the use of Generative Adversarial Networks (GANs) [3], have significantly improved inpainting performance by enabling models to synthesize plausible content through data-driven feature learning. Among these, EdgeConnect [1] introduced a two-stage approach that first predicts edge maps to guide structural reconstruction before performing texture synthesis. While this method improves geometric coherence, it lacks explicit chromatic modeling, often resulting in color inconsistencies and unnatural transitions near the boundaries of inpainted regions.

To address this limitation, we propose **EdgeConnect+**, a three-stage inpainting framework that explicitly integrates both structural and chromatic guidance. The first stage employs an edge generation network (G_1) to predict missing contours using masked edge maps extracted via Canny edge detection, along with grayscale version of the input image and binary masks extracted from the original input image. In the second stage, a low-frequency color prior is introduced by applying the TELEA inpainting algorithm to the masked input. Finally, a guided image completion network (G_2) synthesizes the inpainted output, conditioned on both the predicted structural edges and the chromatic prior.

Although full training of the model remains ongoing, our framework is designed to yield sharper, more perceptually coherent completions, particularly in facial inpainting tasks where both geometry and color fidelity are critical.

The proposed architecture is generalizable and can be trained on other datasets such as Places2 [26], enabling its application across a broader range of inpainting scenarios beyond facial images.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 4.1 presents the proposed method; Sections 4.3 and 4.5 describe the experimental setup and evaluation metrics; Section 5 discusses the results; and Section 6 concludes the paper with future research directions. The GitHub repository containing our full code and models can be found at: [1].

2. Related Work

Early image inpainting methods relied on hand-crafted priors to propagate information from surrounding regions. Diffusion-based approaches, such as PDE-based diffusion (partial differential equation–based inpainting) [16], and exemplar-based techniques, like Criminisi et al's approach [17], achieved reasonable results for small or texture-homogeneous holes, but struggled with semantically meaningful content and large missing regions.

[1] https://github.com/Abhinaykotla/EdgeConnect_Plus_Inpainting_with_Edge_and_Color_Guidance

The advent of deep learning enabled significant progress. Context Encoders [4] were among the first to leverage encoder-decoder architectures with adversarial losses for inpainting, producing semantically plausible content but often yielding blurry results. Subsequent advances, such as Partial Convolutions [6] and Gated Convolutions [7], improved mask handling by modifying convolutional operations to respect missing regions. Attention-based mechanisms like DeepFill [8] enabled long-range feature borrowing to improve visual continuity, while HiFill [18] offered efficient high-resolution inpainting. CoModGAN [9] further improved global consistency through feature-wise modulation of the generator network.

A parallel line of work explored the use of structural priors. PEN-Net [20] predicted edge maps to guide completion, and RFR-Inpainting [21] employed recursive reasoning to refine structural features. EdgeConnect [1] formalized a two-stage approach in which an edge generation network guides the final image synthesis. However, these methods often ignore chromatic coherence, which can lead to unnatural color transitions or desaturation in completed regions.

More recently, transformer-based and diffusion-based models have evolved. ICT [22] introduced external memory mechanisms for capturing long-range dependencies, and DFI [23] applied denoising diffusion probabilistic models for iterative image refinement. While powerful, such models tend to be computationally intensive and less interpretable, with limited control over fine-grained guidance.

Our work builds on EdgeConnect's structure-first approach by introducing explicit chromatic guidance alongside structural priors. We incorporate a low-frequency blurred color map to provide coarse color consistency and fuse it with edge predictions in a unified framework. Unlike prior works that treat structure and texture separately or rely solely on implicit learning, EdgeConnect+ jointly models structural and chromatic cues to enhance perceptual realism in complex inpainting scenarios.

3. Problem Statement

Image inpainting is a longstanding challenge in computer vision, aiming to reconstruct missing or corrupted regions of an image such that the completed output appears both natural and seamless. Traditional methods, including diffusion-based and exemplar-based techniques, are often inadequate for large or irregularly shaped holes, as they fail to preserve high-level semantics and complex structures.

The advent of deep generative models, such as Context Encoders and DeepFill, has significantly advanced the ability to learn semantic priors from large-scale datasets. However, these models frequently exhibit limitations when handling intricate textures, structural alignment, or chromatic consistency. In challenging scenarios such as facial features, man-made structures, or textual regions, artifacts like over-smoothed textures and unnatural color transitions remain common.

EdgeConnect introduced a two-stage pipeline that first predicts structural edges and then performs image completion guided by those edges. Although this framework improves geometric fidelity, it lacks an explicit mechanism for modeling chromatic information. As a result, inpainted regions may suffer from noticeable seams, color bleeding, or inconsistent tones, particularly near the boundaries of missing areas.

In this work, our aim is to address the limitations of structure-only inpainting methods by incorporating both structural and chromatic guidance into the reconstruction process. Specifically, we propose an extension of the EdgeConnect framework with a parallel color guidance stream, represented by a low-frequency blurred color prior. This integrated approach is designed to enforce consistency in both spatial layout and appearance, resulting in inpainted outputs that are perceptually more coherent and visually realistic.

4. Problem Solution

EdgeConnect+ introduces a novel three-stage image inpainting framework designed to address limitations in structure-only generative approaches by jointly modeling both edge and color guidance. While the original EdgeConnect framework emphasizes structural priors through a two-stage pipeline, it lacks a dedicated mechanism for ensuring chromatic consistency, often leading to desaturated or visually discordant results. To overcome this, EdgeConnect+ augments the structural pipeline with an explicit color guidance stage, creating a more semantically coherent and visually realistic output.

Our full pipeline comprises: (1) an edge generation network (G_1) that reconstructs structural contours from incomplete images, (2) a color guidance module that provides low-frequency chromatic context by blending TELEA-inpainted color maps with predicted edge maps, and (3) a guided image completion network (G_2) that synthesizes the final output conditioned on both structure and color priors.

Each component is carefully designed and integrated to improve both pixel-level accuracy and perceptual realism while remaining computationally tractable. The architecture is modular and extensible, making it adaptable to additional cues such as depth or semantic segmentation in future work.

4.1. Methodology

4.1.1. Edge Guidance (G_1)

The first stage of the EdgeConnect+ pipeline focuses on predicting structural contours in masked regions using a dedicated edge generation network, G_1. This network

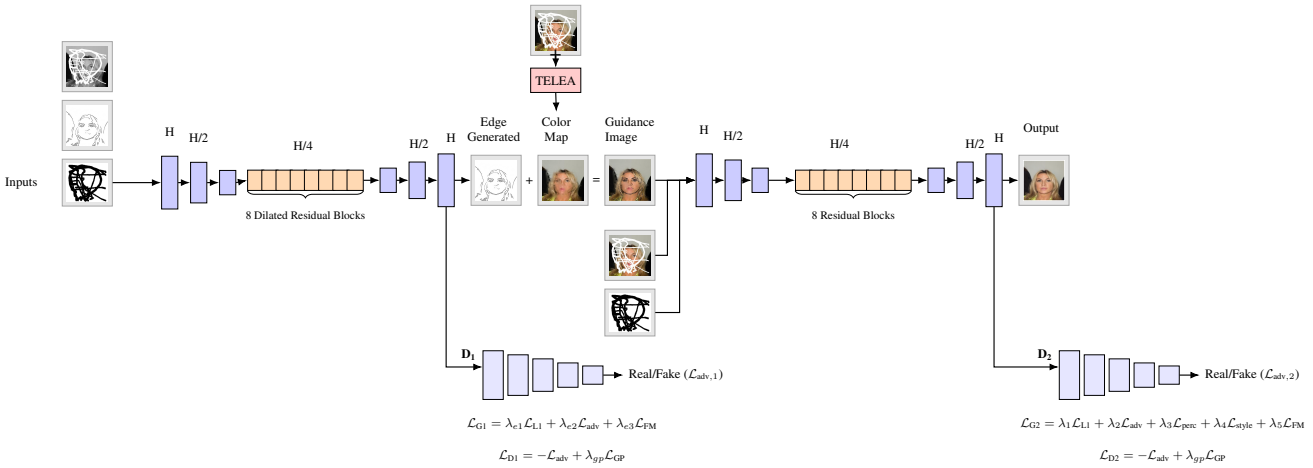

Figure 1. EdgeConnect+ Inpainting Architecture.

builds upon the encoder-decoder architecture introduced in EdgeConnect, incorporating dilated convolutions and residual connections to effectively capture both fine-grained details and global context. Unlike EdgeConnect, which requires a manually provided binary mask alongside a grayscale image as input, EdgeConnect+ extracts the binary mask directly from the masked RGB image by identifying uniformly white pixels. This automated approach eliminates the need for separate mask annotations, thereby simplifying preprocessing and improving the modularity of the pipeline.

The input to G_1 comprises three channels: the grayscale version of the masked image, an edge map computed using the Canny edge detector, and the automatically extracted binary mask. These components are concatenated and passed through a series of downsampling layers, multiple dilated residual blocks, and upsampling layers to generate the predicted edge map. The objective is to ensure that the predicted edges are semantically aligned with the visible structures in the unmasked regions.

The network is trained using a combination of three losses. An L_1 loss encourages pixel-wise accuracy, an adversarial loss facilitated by a PatchGAN-based discriminator (D_1) promotes the generation of realistic edge structures, and a feature matching loss ensures perceptual stability by minimizing discrepancies in internal feature representations between real and generated outputs. While these losses are inspired by the original EdgeConnect framework, they are tailored in EdgeConnect+ to accommodate the automatically derived input masks.

4.1.2. Color Guidance

Following edge prediction, the second stage of the pipeline focuses on constructing a color guidance map to support the subsequent inpainting process. While EdgeConnect relies exclusively on structural priors, our method introduces explicit chromatic context to guide G_2 more effectively. The primary objective of this stage is to provide low-frequency color cues that promote smooth transitions and maintain chromatic coherence across masked and unmasked regions.

To construct the color guidance, we initially explored the use of Gaussian blur applied to the unmasked regions as a way to approximate the global color distribution. While Gaussian blurring effectively removes high-frequency noise, it lacks spatial awareness and often results in unnatural color transitions near mask boundaries. As a more semantically meaningful alternative, we adopt the TELEA inpainting algorithm, an efficient, non-learning-based method that fills missing regions by propagating nearby pixel values based on geometric and photometric continuity. Compared to Gaussian blur, TELEA produces smoother and more spatially coherent color priors that better preserve local structures. To further reduce edge artifacts, we apply a slight dilation to the binary mask before inpainting, softening the boundaries and minimizing residual white spaces.

After TELEA-based inpainting is applied to generate a low-frequency color prior, we refine the guidance image by overlaying the predicted edge map from G_1 onto the color map using a thin black stroke. This fusion process yields a composite guidance map that simultaneously encodes structural details and chromatic information. The resulting image serves as a critical input to the final inpainting generator, enhancing its ability to synthesize perceptually realistic and contextually consistent outputs.

4.1.3. Final Inpainting (G_2)

The final stage in our pipeline is handled by G2, which generates the completed image by taking in three inputs: the masked RGB image, the fused guidance map, and the binary mask. These are combined into a single 7-channel input, allowing the network to process structural and color

cues together in a more unified and efficient way.

In contrast to the original EdgeConnect architecture, which employs a U-Net-based generator with skip connections and handles edge and mask inputs separately, our design adopts a more streamlined approach. G_2 eliminates skip connections between the encoder and the decoder, thereby reducing memory usage and architectural complexity. The network consists of initial convolutional layers for downsampling, followed by a series of residual blocks for semantic feature learning, and transposed convolutions for upsampling and image reconstruction. This simplification facilitates faster convergence and improved training stability without heavily compromising output quality.

During training, G_2 is supervised using a combination of loss functions. An L1 loss is applied within the masked regions to prioritize learning in areas of missing content. Perceptual and style losses, computed from feature activations of a pretrained VGG network, help preserve high-level semantics and textural consistency. In addition, an adversarial loss, facilitated by a patch based discriminator, D_2, encourages the generation of visually sharp and realistic results. To maintain consistency and reproducibility during training, all input data is normalized and resized, and any missing guidance images are automatically regenerated as needed.

Figure 1 illustrates the overall EdgeConnect+ pipeline, highlighting the interaction between the edge generation, color guidance, and final reconstruction stages.

4.2. Dataset

We evaluate our method on the CelebA dataset [2], a large-scale face dataset containing over 200,000 celebrity images with diverse facial attributes such as pose, expression, age, and lighting conditions. All images are center-cropped and resized to 256×256 pixels. We partition the CelebA dataset into a training set of 162,079 images, validation set of 10,129 images, and testing set of 30,391 images. Figures 2 and 3 show samples from CelebA dataset.

Figure 2

Figure 3

To simulate realistic inpainting scenarios, we apply irregular binary masks [27] to the images, ensuring each mask covers at least 20% of the image area. These wide masks are designed to challenge the model with substantial missing regions, including facial features and background structures. Masked regions are filled with white pixels to create input images for training.

The preprocessing involves deriving binary masks to indicate missing regions from white areas in the images. Input edges are generated by applying the Canny edge detector to the masked image, subsequently removing edges corresponding to mask regions, thus retaining only edges from visible image areas. Grayscale versions of masked images are created as additional input for the edge generation network (G1). Ground truth edges are generated by applying the Canny edge detector to the original (unmasked) images, supervising G1 during training.

Figures 4 and 5 show samples from the prepared dataset.

Figure 4. Ground Truth Image

Figure 5. Input Image

4.3. Training Setup

EdgeConnect+ is trained on the CelebA dataset, with each training batch comprising 192 samples striking a balance between memory efficiency and convergence stability. The edge generation network (G_1) is trained for 25 epochs, followed by 5 epochs of training for the image completion network (G_2). Both networks are optimized using the Adam optimizer, configured with a learning rate of 1×10^{-4} and a weight decay of 5×10^{-5} to encourage stable convergence and reduce overfitting.

To accommodate fused edge and color guidance while maintaining computational efficiency, EdgeConnect+ is designed with a leaner architecture compared to its predecessor. Whereas the original EdgeConnect framework utilizes approximately 22 million parameters [15] across its generators, EdgeConnect+ operates with a reduced footprint of roughly 21.5 million parameters, distributed across G_1 and G_2. Despite the lighter architecture, the model is capable of processing richer input representations by incorporating structural and chromatic cues, leading to perceptually and structurally coherent inpainting results.

Training is performed using mixed-precision arithmetic with gradient scaling, which accelerates computation while preserving numerical stability. An Exponential Moving Average (EMA) is maintained over the generator weights to

smooth updates and improve generalization. To mitigate overfitting, an early stopping mechanism halts training if validation loss fails to improve for five consecutive epochs.

The training loops are modular, fault-tolerant, and fully instrumented. All training progress is logged, including loss trends and key metrics, with periodic checkpointing that enables seamless interruption and resumption. Scripts are integrated to save generated sample outputs every 200 batches (configurable), allowing for real-time qualitative monitoring. Additionally, the training setup supports on-the-fly modification of loss weights based on observed outputs, enabling dynamic tuning of hyperparameters mid-training. This design facilitates iterative experimentation, making it possible to resume from a previous checkpoint while adjusting model behavior to improve convergence or visual fidelity.

All experiments are conducted on CUDA-enabled NVIDIA A100 GPUs, leveraging GPU parallelism for efficient training of large-scale generative models.

4.4. Loss Functions

The EdgeConnect+ framework consists of two generator–discriminator pairs: the edge generation network G_1 and its discriminator D_1, followed by the inpainting generator G_2 and its corresponding discriminator D_2. Each component is trained using a composite of loss functions designed to encourage structural accuracy, perceptual fidelity, and stylistic realism.

Pixel Reconstruction Loss (L1): Both G_1 and G_2 are trained with an L1 pixel reconstruction loss to ensure that the outputs, edge maps in the case of G_1, and completed images for G_2 remain close to the ground truth at a pixel level:

$$\mathcal{L}_{\text{L1}} = \|y - \hat{y}\|_1$$

where y and $\hat{y}$ represent the ground truth and the predicted outputs, respectively.

Adversarial Loss: To improve realism, both generators are trained adversarially using PatchGAN based discriminators (D_1, D_2), which evaluate the authenticity of local image patches:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,\hat{y}}[\log(1 - D(x,\hat{y}))]$$

Gradient Penalty: To stabilize discriminator training and enforce a soft Lipschitz constraint, we incorporate a gradient penalty term:

$$\mathcal{L}_{\text{GP}} = \mathbb{E}_{\hat{x}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right]$$

where $\hat{x}$ is an interpolated sample between real and generated data.

Feature Matching Loss: To reduce mode collapse and stabilize adversarial training, a feature matching loss is used:

$$\mathcal{L}_{\text{FM}} = \sum_{i=1}^{L} \|D_i(x,y) - D_i(x,\hat{y})\|_1$$

which compares internal feature activations of the discriminator between real and generated outputs.

Perceptual and Style Losses (G_2 only): For the image synthesis network G_2, we add perceptual and style-based losses using a pre-trained VGG16 network. The perceptual loss is defined as:

$$\mathcal{L}_{\text{perc}} = \sum_{l} \|\phi_l(I) - \phi_l(\hat{I})\|_1$$

and the style loss, derived from Gram matrices of VGG features, as:

$$\mathcal{L}_{\text{style}} = \sum_{l} \|G_l(I) - G_l(\hat{I})\|_1$$

Total Loss Formulations: The full objective functions for each generator and discriminator are:

$$\mathcal{L}_{G_1} = \lambda_{e1}\mathcal{L}_{\text{L1}} + \lambda_{e2}\mathcal{L}_{\text{adv}} + \lambda_{e3}\mathcal{L}_{\text{FM}}$$

$$\mathcal{L}_{D_1} = -\mathcal{L}_{\text{adv}} + \lambda_{gp}\mathcal{L}_{\text{GP}}$$

$$\mathcal{L}_{G_2} = \lambda_1\mathcal{L}_{\text{L1}} + \lambda_2\mathcal{L}_{\text{adv}} + \lambda_3\mathcal{L}_{\text{perc}} + \lambda_4\mathcal{L}_{\text{style}} + \lambda_5\mathcal{L}_{\text{FM}}$$

$$\mathcal{L}_{D_2} = -\mathcal{L}_{\text{adv}} + \lambda_{gp}\mathcal{L}_{\text{GP}}$$

Here, λ_{e1} through λ_{e3}, and λ_1 through λ_5, are empirically chosen weights for balancing the individual loss components. The gradient penalty weight λ_{gp} plays a critical role in ensuring discriminator stability. In our experiments, these losses collectively contribute to producing inpainted outputs that balance pixel-level accuracy with high perceptual and structural quality.

Figures 8 and 13 visualize the training progression of each loss term.

4.5. Evaluation Metrics

We evaluate EdgeConnect+ using standard metrics that assess both pixel-level fidelity and perceptual quality. Although the model is not fully trained, preliminary results show promising improvements across several dimensions, suggesting the potential of combining structural and chromatic guidance.

PSNR: Peak Signal-to-Noise Ratio evaluates reconstruction fidelity by comparing pixel-level differences. EdgeConnect+ achieves a PSNR of 25.23, slightly lower than EdgeConnect's 25.28. This marginal difference may be attributed to the model's emphasis on perceptual realism rather than strict pixel-level matching.

SSIM: The Structural Similarity Index (SSIM) measures perceptual quality in terms of luminance, contrast, and structure. EdgeConnect+ achieves 0.864 compared to 0.846

for EdgeConnect, suggesting improved semantic coherence due to integrated edge and color guidance.

ℓ_1 **Loss:** EdgeConnect+ reports a slightly higher ℓ_1 error (4.83%) versus EdgeConnect (3.03%), which aligns with the design goal of prioritizing perceptual alignment over exact pixel recovery.

FID: Fréchet Inception Distance (FID) evaluates image realism and diversity. EdgeConnect+ achieves a FID score of 2.94, indicating reasonably good alignment with natural image distributions, though slightly less effective than the original EdgeConnect (FID 2.82), potentially due to limited training or added model complexity.

LPIPS: The LPIPS metric measures perceptual similarity using deep feature comparisons. With a score of 0.193, EdgeConnect+ shows encouraging perceptual closeness to the ground truth. Since LPIPS was not reported for EdgeConnect, this serves as a supplemental indication of perceptual improvements.

These initial results suggest that integrating edge and color priors can positively influence inpainting performance. We anticipate that further training and tuning will enhance these metrics further and strengthen the model's performance relative to established baselines.

Metric	Fusion Label [25]	EdgeConnect	Ours
PSNR	29.16	25.28	25.23
SSIM	0.9235	0.846	0.864
ℓ_1 Loss (%)	Not reported	3.03	4.83
FID	Not reported	2.82	2.94
LPIPS	Not reported	Not reported	0.193

Table 1. Quantitative comparison of inpainting performance on the CelebA dataset of different models: Fusion label, EdgeConnect and EdgeConnect+

Table 1 presents quantitative results comparing the performance of EdgeConnect, our proposed EdgeConnect+, and the fusion-based method introduced by the paper - *Generative image inpainting via edge structure and color aware fusion* [25]. We include the Fusion Label model in this comparison as it also combines edge and color information using a dual-encoder architecture with gated feature fusion and spatial-channel attention to guide the final inpainting. Including this baseline offers a meaningful point of comparison, helping contextualize the effectiveness of our own modular guidance strategy in balancing structural alignment and perceptual quality.

5. Results

This section presents qualitative and quantitative outcomes from the EdgeConnect+ pipeline, covering edge generation, intermediate guidance fusion, full inpainting results, and comparative evaluations.

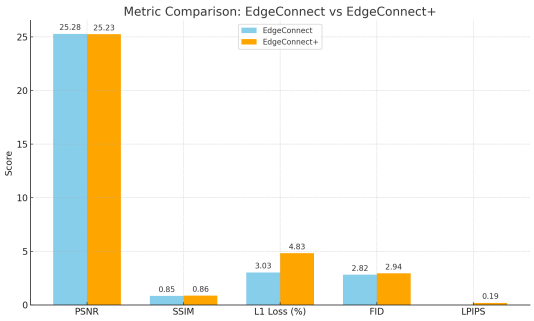

Figure 6. Visual comparison of evaluation metrics between EdgeConnect and EdgeConnect+ on CelebA.

Figure 7 illustrates the output of the edge generation network G_1. The top row displays: (1) input edge map extracted from the masked image, (2) the corresponding binary mask, and (3) the predicted edge map. The bottom row shows the ground truth edge map derived from the unmasked image and its grayscale counterpart. The results highlight G_1's ability to reconstruct plausible edge structures despite missing regions.

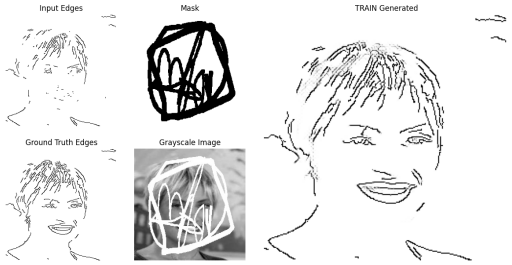

Figure 7. Edge prediction outputs from G_1. Top: input edges, binary mask, and predicted edges. Bottom: ground truth edge map and grayscale input.

Training dynamics for G_1 and its discriminator D_1 are shown in Figure 8. The top subplot visualizes batch-level losses, including L1, adversarial, and feature matching terms, as well as discriminator performance. The bottom subplot presents epoch-wise averages, revealing consistent loss reduction and stable adversarial training.

Figure 9 presents the intermediate guidance representations passed to the inpainting network G_2. Each triplet shows: (1) the predicted edge map, (2) the low-frequency color map generated using TELEA inpainting, and (3) a fused overlay combining edges and color. These multi-modal cues jointly guide G_2 to synthesize perceptually realistic and structurally coherent completions.

Figures 10, 11, and 12 display complete inpainting outputs from EdgeConnect+. Each result consists of six com-

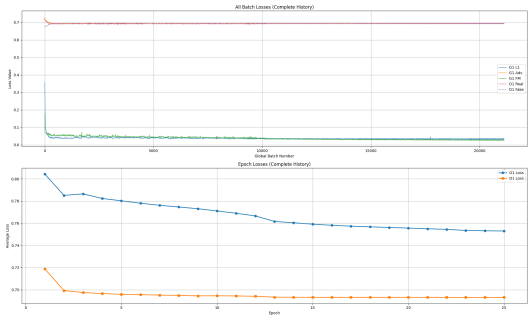

Figure 8. Training loss trends for G_1 and D_1.

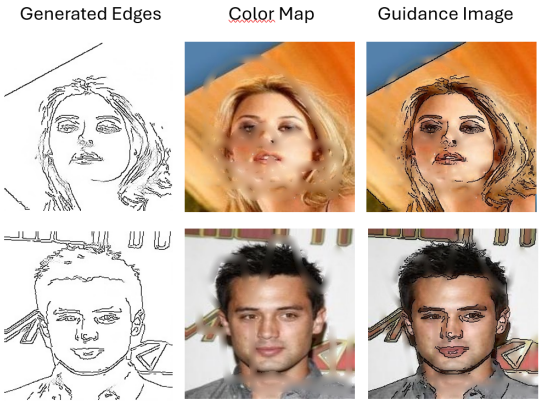

Figure 9. Intermediate representations: edge map, blurred color map, and guidance image.

ponents: the masked input image (top-left), fused guidance (top-center), and final output (top-right), followed by the binary mask (bottom-left), ground truth image (bottom-center), and pixel-wise absolute difference map (bottom-right). The difference maps, which are largely dark, indicate high alignment between prediction and ground truth, demonstrating the potential of the model for high-quality image reconstruction.

Figure 10. Final Generated Output 1

Figure 13 shows loss progression during G_2 training. The top plot tracks batch-level losses: L1, adversarial, per-

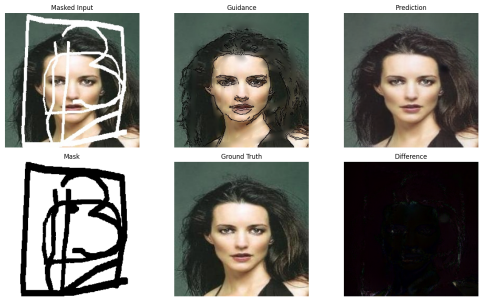

Figure 11. Final Generated Output 2

Figure 12. Final Generated Output 3
Top: masked input, fused guidance, predicted output. Bottom: binary mask, ground truth, and absolute error map.

ceptual (scaled), style, and feature matching, as well as discriminator classification loss. The bottom plot summarizes average epoch losses, indicating early convergence.

Due to computational constraints, G_2 was trained for only 5 epochs. Nonetheless, early-stage outputs are encouraging and highlight the effectiveness of multimodal guidance. With extended training and hyperparameter tuning, we anticipate further improvements in texture quality, semantic accuracy, and visual coherence.

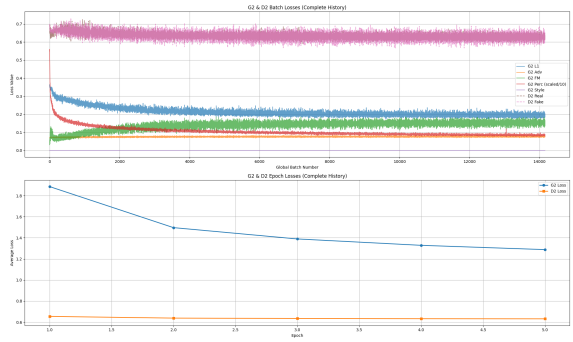

Figure 13. Loss trends for G_2 and D_2 during training.

Figure 14 compares inpainting results from the original EdgeConnect (top) and EdgeConnect+ (bottom) for the same masked input. Although both methods produce plau-

sible completions, EdgeConnect+ shows improved alignment in structure, texture continuity, and color coherence, particularly around fine details such as facial features and background patterns. EdgeConnect sometimes exhibits sharp but semantically inconsistent edges or color mismatches, which EdgeConnect+ mitigates through joint structural and chromatic guidance.

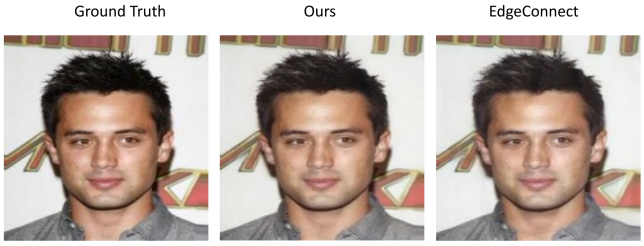

Figure 14. Comparison of inpainting results. Top: EdgeConnect; Bottom: EdgeConnect+.

6. Conclusion

In this paper, we presented **EdgeConnect+**, a modular image inpainting framework that extends the original EdgeConnect architecture by incorporating both structural and chromatic guidance. By fusing edge predictions from a dedicated generator with a low-frequency color prior derived via TELEA inpainting, our method enhances visual coherence and semantic alignment in the reconstructed regions.

Through qualitative examples, intermediate visualizations, and training loss analyses, we demonstrated that EdgeConnect+ effectively addresses limitations of structure only models, such as incomplete contour recovery and color inconsistency [1, 7, 24]. The integration of dual guidance encourages the inpainting generator to synthesize content that aligns with both spatial layout and appearance context.

Although our results were obtained after training the second stage generator for only five epochs, the model exhibited stable convergence and encouraging visual performance. This early success highlights the efficiency and promise of our architecture. With extended training and more computational resources, we expect further improvements in detail preservation, texture fidelity, and robustness across diverse scenes.

Looking ahead, EdgeConnect+ offers a flexible foundation for future work. Potential extensions include learning-based color guidance modules, attention enhanced fusion mechanisms, and semantic conditioning via vision language models. By enabling controllable and context-aware image restoration, our approach contributes to advancing the capabilities of deep generative inpainting systems.

References

[1] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi. EdgeConnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019.

[2] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *International Conference on Computer Vision (ICCV)*, 2015.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[4] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, pages 2536–2544, 2016.

[5] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen. Image melding: Combining inconsistent images using patch-based synthesis. *ACM TOG*, 31(4):82–1, 2012.

[6] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *ECCV*, pages 85–100, 2018.

[7] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Free-form image inpainting with gated convolution. In *ICCV*, pages 4471–4480, 2019.

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In *CVPR*, pages 5505–5514, 2018.

[9] S. Zhao, Z. Liu, Z. Lin, J.-Y. Zhu, and W. Xu. Co-ModGAN: Co-modulated generative adversarial networks. In *NeurIPS*, 34:4439–4452, 2021.

[10] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2016.

[11] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *CVPR*, pages 2414–2423, 2016.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of Wasserstein GANs. In *NeurIPS*, pages 5767–5777, 2017.

[13] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[15] B. Xia, Y. Zhang, S. Wang, Y. Wang, X. Wu, Y. Tian, W. Yang, and L. Van Gool. DiffIR: Efficient diffusion model for image restoration. *arXiv preprint arXiv:2308.07950*, 2023.

[16] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. SIGGRAPH*, pages 417–424, 2000.

[17] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE TIP*, 13(9):1200–1212, 2004.

[18] L. Yi, Y. Liu, Y. Luo, Y. Xu, and J. Tang. Contextual residual aggregation for ultra high-resolution image inpainting. In *CVPR*, pages 7508–7517, 2020.

[19] Y. Liu, S. Wang, Y. Xu, J. Tang, and B. Li. Structure-aware image inpainting with multi-scale gated convolutions. In *CVPR*, pages 2756–2765, 2022.

[20] Y. Zeng, J. Fu, H. Chao, and Y. Zheng. Learning pyramid-structure attention for image inpainting. In *ECCV*, pages 481–497, 2020.

[21] Y. Li, S. Liu, J. Yang, and M.-H. Yang. Recurrent feature reasoning for image inpainting. In *CVPR*, pages 7760–7768, 2020.

[22] R. Wan, Y. Zhang, Z. Li, Y. Wang, and L. Ma. Image inpainting via learning contextual residual aggregation. In *CVPR*, pages 8281–8290, 2021.

[23] A. Lugmayr, M. Danelljan, and R. Timofte. RePaint: Inpainting using denoising diffusion probabilistic models. *arXiv preprint arXiv:2201.09865*, 2022.

[24] H. Liu, X. Zhang, Y. Wan, and D. Lin. PD-GAN: Probabilistic diverse GAN for image inpainting. In *CVPR*, pages 9371–9381, 2021.

[25] H. Shao, Y. Wang, Y. Fu, and Z. Yin. Generative image inpainting via edge structure and color aware fusion. *Journal of Visual Communication and Image Representation*, 2021.

[26] `https://www.kaggle.com/datasets/nickj26/places2-mit-dataset`

[27] `https://github.com/karfly/qd-imd`