EdgeConnect+: Adversarial Inpainting with Edge and Color Guidance

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

Abstract

We present EdgeConnect+, an enhanced deep learning-based image inpainting model that integrates both structural and chromatic guidance to improve the realism of reconstructed images. Our method builds upon the EdgeConnect framework by incorporating a low-frequency blurred color map in addition to the edge map to enhance contextual and chromatic consistency in missing regions. EdgeConnect+ consists of three stages: (1) an edge generation network (G1) that predicts structural contours, (2) generation of a coarse color map to guide chromatic consistency, and (3) a texture completion network (G2) that performs final image reconstruction using predicted edges and color guidance. We outline the methodology and detail the evaluation metrics that will be used to assess reconstruction performance.

1. Introduction

Image inpainting is a fundamental task in computer vision aimed at reconstructing missing or corrupted regions in images, ensuring visual realism and structural coherence. Its practical applications include photo restoration, object removal, and creative editing. Traditional techniques, such as diffusion-based and patch-based methods, have limited effectiveness in reconstructing complex structures and large missing areas, often resulting in blurry or incoherent textures.

Recently, deep learning methods, particularly Generative Adversarial Networks (GANs) [3], have significantly advanced inpainting performance by generating plausible content in missing image regions through learned representations from large datasets. Among these, EdgeConnect [1] explicitly predicts missing edges to ensure structural coherence before image completion. While effective structurally, EdgeConnect lacks explicit color guidance, frequently leading to noticeable artifacts and unnatural color transitions at the boundaries of reconstructed regions.

To address these limitations, we propose **EdgeConnect+**, an enhanced inpainting framework that explicitly integrates both edge structure and color information. Our proposed pipeline involves a three-stage process. First, an edge generator predicts missing structural information conditioned on masked image edges, grayscale input, and the binary mask of missing regions. Second, we introduce color guidance via a Gaussian-blurred version of the original image, providing coarse color priors. Recognizing limitations in Gaussian blur's semantic guidance, we are also exploring advanced color propagation methods such as Partial Convolutions [6] and Contextual Attention (CA) modules [8] to further enhance visual consistency. Finally, a second generative network leverages both the predicted edges and the provided color hints to reconstruct the final realistic inpainted image.

We validate our approach on the CelebA dataset [2], containing diverse facial images with various attributes such as pose, expression, and illumination. Preliminary results indicate improvements in visual realism and coherence compared to existing edge-based inpainting methods.

2. Related Work

Early deep learning-based inpainting methods, such as Context Encoders [4], relied on encoder-decoder architectures with reconstruction and adversarial losses. While effective for coarse structure, they often failed to preserve fine details, especially in large or complex missing regions.

Partial Convolutions [6] and Gated Convolutions [7] improved robustness to free-form masks by adapting convolution operations based on mask validity. However, these methods lacked explicit structure guidance, limiting their effectiveness in preserving geometry.

Attention-based approaches, such as DeepFill [8], introduced contextual attention to propagate textures from known regions, improving visual coherence. Co-ModGAN [9] advanced this further using feature-wise modulation, enabling stronger conditioning on input context and better global consistency.

To incorporate structural priors, two-stage methods have been proposed, such as EdgeConnect [1], which first predicts edge structure and then performs image completion. However, it does not explicitly model color consistency, leading to desaturated or artifact-prone outputs.

Our method builds upon this structural guidance paradigm by introducing a color guidance stream that enhances chromatic continuity, resulting in more realistic and perceptually rich reconstructions.

3. Methodology

Our approach follows a three-stage pipeline designed to integrate edge and color information effectively for realistic image inpainting. The process is divided into the following stages:

3.1. Edge Generation (G1)

In the first stage, we employ an edge generator (G1) that predicts the missing edges in the occluded regions. The generator is conditioned on the masked image, grayscale image, and binary mask. The predicted edges are essential for ensuring that the structure of the image is consistent with the surrounding context. The generator is trained using adversarial loss, L1 loss, and feature-matching loss to achieve stable convergence and high-quality edge predictions.

3.2. Color Map Generation

Once the edges are predicted, we move to the second stage, where color guidance is provided to fill in the missing regions. Initially, a Gaussian blur is applied to the unmasked portions of the image to generate a low-frequency color map. This map provides coarse color information, which helps maintain color consistency across the inpainted regions. Additionally, to improve the color propagation, we are exploring advanced techniques like Partial Convolutions [6] and Contextual Attention (CA) modules [8], which allow more semantic and context-aware color filling, further improving the realism of the inpainted regions.

3.3. Final Inpainting (G2 - Planned)

In the third and final stage, a second generative network (G2) completes the image reconstruction. The G2 model takes as input a composite RGB image where the unmasked regions retain the original content, while the masked regions are filled with the predicted edges and the generated color map. Along with the composite image, the binary mask is fed into G2 to indicate the missing regions. The output of this stage is a fully inpainted image that combines both structural and color information to ensure a realistic and visually coherent result.

4. 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.

4.1. Data Preparation

To simulate realistic inpainting scenarios, we apply irregular binary masks to the images, ensuring each mask covers at least 20% of the image area. Masked regions are filled with white pixels to create input images for training.

The preprocessing steps are as follows:

- Binary masks indicating missing areas are derived from white regions.
- 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.

4.2. Dataset Splits

We split the CelebA dataset as below:

- **Training:** 162,079 images
- **Validation:** 10,129 images
- **Testing:** 30,391 images

Figures 1 and 2 show samples from the prepared dataset.

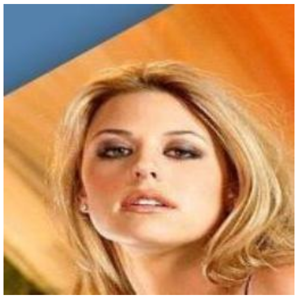

Figure 1. Ground Truth Image

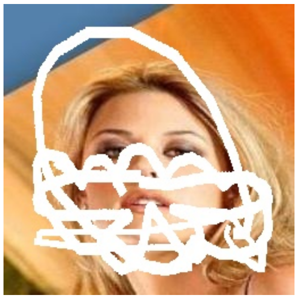

Figure 2. Input Image

5. Loss Functions

EdgeConnect+ employs a combination of loss functions during training to ensure structural accuracy, perceptual realism, and texture consistency. These losses are applied across both stages of the pipeline: edge generation (G1) and image completion (G2).

5.1. L1 Loss (Pixel-wise Reconstruction)

L1 loss computes the mean absolute difference between the predicted and ground truth images (or edges). It encourages pixel-wise accuracy and helps maintain structural alignment.

5.2. Adversarial Loss

We use a non-saturating GAN (NS-GAN) objective to train both G1 and G2. This loss encourages the generators to produce outputs indistinguishable from real data, promoting naturalness in edges and textures.

5.3. Feature Matching Loss

Applied in G1, this loss minimizes the difference between discriminator feature activations for real and generated edge maps, promoting training stability and structural realism.

5.4. Perceptual Loss

Perceptual loss is computed using feature activations from a pretrained VGG16 network [10]. It helps G2 preserve high-level content semantics and overall scene consistency.

5.5. Style Loss

Style loss [11] ensures texture coherence by matching the Gram matrices of feature maps between predicted and ground truth images, helping to preserve texture and fine patterns.

5.6. Gradient Penalty

To enforce Lipschitz continuity and improve training stability, we apply a gradient penalty [12] on the discriminator in the edge generation stage.

6. Evaluation Metrics

To quantitatively evaluate the performance of EdgeConnect+, we plan to report the following widely accepted metrics:

6.1. PSNR (Peak Signal-to-Noise Ratio)

PSNR measures the ratio between the maximum possible pixel intensity and the mean squared error between the inpainted and ground truth images. It serves as a basic indicator of pixel-level reconstruction fidelity. Higher PSNR values correspond to lower distortion.

6.2. SSIM (Structural Similarity Index Measure)

SSIM [14] evaluates structural and perceptual similarity by comparing luminance, contrast, and structure between the predicted and ground truth images. It ranges from -1 to 1, where values closer to 1 signify higher perceptual similarity.

6.3. Mean Absolute Error (L1 Loss)

This metric computes the average of the absolute pixel-wise differences. Since it is one of the training objectives, evaluating it at test time provides consistency with training. Lower values indicate better reconstruction quality.

6.4. LPIPS (Learned Perceptual Image Patch Similarity)

LPIPS [13] compares deep features extracted from pretrained networks to assess perceptual similarity. Unlike PSNR and SSIM, LPIPS aligns more closely with human perception of image quality. Lower values indicate stronger perceptual resemblance to the ground truth.

Metric	Description	Preferred Direction
PSNR	Pixel-wise fidelity via signal-to-noise ratio	Higher
SSIM	Structural and perceptual similarity	Closer to 1
L1 Loss	Mean absolute pixel error	Lower
LPIPS	Perceptual similarity via deep features	Lower

Table 1. Quantitative metrics for evaluating inpainting quality.

7. Training Setup

We train EdgeConnect+ using the CelebA dataset with the following configuration:

- **Batch Size:** 192 datapoints per iteration.
- **Epochs:** 250 full passes over the training set.
- **Optimizer:** Adam optimizer with a learning rate of 1×10^{-4} and weight decay of 5×10^{-5}.
- **Precision:** Mixed precision training is employed to reduce memory footprint and improve computational efficiency.
- **Stabilization:** Exponential Moving Average (EMA) is applied to smooth generator weights during training.
- **Early Stopping:** Training halts if validation loss does not improve for 10 consecutive epochs.
- **Environment:** All experiments are conducted on CUDA-enabled GPUs.

8. Preliminary Results

Figures 3 and 4 show samples of current G1 results from training, and Figure 5 shows training loss trends up to epoch 25. The model demonstrates good convergence behavior, producing progressively sharper edge predictions. Training is ongoing, with additional epochs planned. Future stages will evaluate full G2 integration.

9. Conclusion

This work presents EdgeConnect+, a structure- and color-aware inpainting framework that enhances visual realism

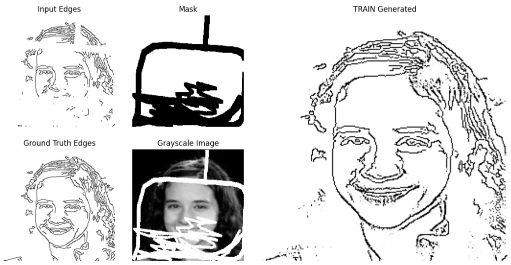

Figure 3. Generated Sample 1

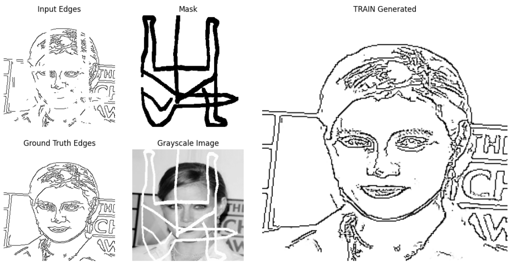

Figure 4. Generated Sample 2

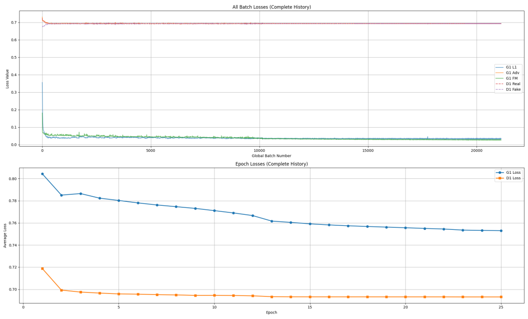

Figure 5. Loss Trends

by jointly leveraging edge predictions and chromatic guidance. By integrating enhanced low-frequency color priors with structural contours, the model produces inpainted results with improved fidelity, continuity, and perceptual quality.

While promising, training the full pipeline is computationally intensive, which imposes constraints on batch size and training efficiency. Moving forward, we aim to complete G2 training, explore transformer-based attention mechanisms for better global context modeling, and evaluate the model on larger and more diverse datasets such as Places2 to assess generalizability.

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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (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 Transactions on Graphics (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. *Proceedings of the European Conference on Computer Vision (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. *Proceedings of the IEEE/CVF International Conference on Computer Vision (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. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (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. *Advances in Neural Information Processing Systems (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 *European Conference on Computer Vision (ECCV)*, pages 694–711. Springer, 2016.

[11] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems (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 *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition (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 Transactions on Image Processing*, 13(4):600–612, 2004.