CS7180 Advanced Perception

Object Removal using Image Inpainting

Submitted by:

Sai Manichandana Devi Thumati – 002443106

Submission Date: December 13 2025

Object Removal using Image Inpainting

Sai Manichandana Devi Thumati
Northeastern University
thumati.sa@northeastern.edu

***Abstract*—Object removal from digital images is a task critical in computer vision and photo editing, letting one remove unwanted elements like people, signs, or debris while preserving the natural appearance of the scene. This project explores some image inpainting techniques for visually plausible reconstructions of masked regions following object removal. Inpainting fills missing pixels by leveraging contextual information from surrounding areas, addressing challenges of texture continuity, structural alignment, and semantic coherence.**

Object masking, in which a binary mask is used to identify the region that needs to be removed (either manually or with the use of segmentation tools), is where the methodology begins. Conventional solutions include classic diffusion-based methods, such as gradient and isophote propagation from boundaries inward, effective in cases of small and smooth areas, and exemplar-based techniques that fill in matching patches from the image with priority given to edges in order to better preserve structure. The most recent deep learning-based approaches, drawing inspiration from generative models and diffusion processes, have achieved a great degree of high-level semantic understanding, hence yielding highly realistic fillings even in the case of irregular masks or large masks.

Experiments were conducted on a variety of real-world images, including landscapes, urban scenes, and portraits, where the foreground objects act as distractions. Results show strong performance in cases of repetitive textures like beaches, skies, or grass, where reconstructions blend in perfectly.

***Index Terms*—Index Terms—Image inpainting, object removal, exemplar-based inpainting, diffusion-based inpainting, OpenCV, digital image editing, computer vision.**

I. Introduction & Prior work

Digital images are known to have unwanted elements that may distract from the overall image, such as people in a landscape, signs in a street scene, or blemishes on a portrait. The process of eliminating such unwanted elements from an image is a cumbersome job that requires a high level of skill with the use of tools like the healing brush.

The most significant source of inspiration behind this project is the seminal ”Image Inpainting” research paper by Marcelo Bertalmio, Guillermo Sapiro, Vicent Caselles, and Coloma Ballester, presented at SIGGRAPH 2000 [1]. It introduced a partial differential equation (PDE)-based diffusion technique that propagates surrounding structural and textural information into missing regions, mimicking professional art restoration practices.

This line of research saw the emergence of the fast marching algorithm proposed by Alexandru Telea in 2004 [2], which provided an efficient anisotropic diffusion-based solution for inpainting. A significant breakthrough came with the exemplar-based image inpainting algorithm by Antonio Criminisi, Patrick Pérez, and Kentaro Toyama in 2003 (published in 2004) [3], which prioritizes filling along strong edges and copies matching patches, enabling effective restoration of areas with prominent linear structures.

Incomplete scene analysis has been more recently improved with deep learning solutions such as context encoders [4], partial and gated convolutions [7,8,12], and generative models like LaMA and Stable Diffusion-based inpainting [recent surveys: 11,17], which leverage large datasets to infer high-level semantics and achieve state-of-the-art performance on challenging, large, or irregular missing regions.

Such classical and modern methods, together with accessible libraries such as OpenCV, motivate the exploration of object removal via inpainting in this project, highlighting both the enduring utility of traditional approaches and the transformative potential of contemporary deep learning techniques.

II. Methodology

The object removal system that has been developed in this project is a two-step process: object masking and image inpainting. This solution uses classical image inpainting methods mainly, such as Telea's Fast Marching Method (INPAINT_TELEA), NavierStokes based (INPAINT_NS), and may use exemplar image methods, such as that described by Criminisi et al

A. Object Masking

The first step consists of demarcating the area that holds the unwanted object. This is done with the use of a binary mask, where the pixels of the unwanted object are set to white (255), against which the background pixels are set to black (0). In the current implementation, masks are typically set by the user, with the use of image editing software such as GIMP, Photoshop, or even basic graphics software. Care is emphasized on the need to do the masking accurately, where inaccuracies might result in the creation of unwanted artifacts within the final result. Future enhancements might include the use of semiautomatic methods of object segmentation, such as GrabCut.

B. Image Inpaint

Once the mask is applied to the image, it creates a ”hole” effect on the marked area. InPainting algorithms are used to fill that missing hole by cleverly accessing the neighboring known pixels.

Few classical models were also used:

- Diffusion-based Inpainting : In these methods, a smooth diffusion of color and gradient is introduced from the edge toward the inside. In the Telea algorithm (cv2.INPAINT_TELEA), based on the Fast Marching Method, the fill boundary is treated as a level set, which computes the recovered pixels as weighted averages from neighboring known pixels. In the NavierStokes algorithm (cv2.INPAINT_NS), fluid motion equations are used. The Telea algorithm is useful when areas are small, non-texture, or scratches, but when scaled to a larger area, the result is a blurring effect on textures.

- Exemplar-Based Inpainting : This is a patchbased technique that uses Criminisi et al. to efficiently remove large objects by mimicking textures and straight structures. In this technique, a loop is used:
  - For pixels on the fill front, a priority P(p) is calculated with the formula $P(p) = C(p) \times D(p)$, where C(p) is the confidence, which corresponds to the proportion of known pixels in the patch, and D(p), which is a data term that favors the recovery of isophotes (strong edges).
  - Find the highest priority patch.
  - Search known areas for the best match against a patch, with a Sum of Squared Differences (SSD).
  - It copies a corresponding patch to the target position and modifies the confidence values.

This priority system helps make sure that structural details are filled in before propagating errors, giving a consistent texture.

The overall process is quite simple. The image to be processed, the original image, as well as the corresponding mask, are loaded, the type of image inpainting technique (either diffusion-based (Telea/NS) or exemplarbased) is specified, the technique is applied to fill the marked portion, and the resulting image is saved. The code uses OpenCV. This blend of masking and classical inpainting offers a robust, efficient solution that can be used for object removal in scene cleaning tasks in the realm of photo editing.

C. *System Architecture and Implementation*

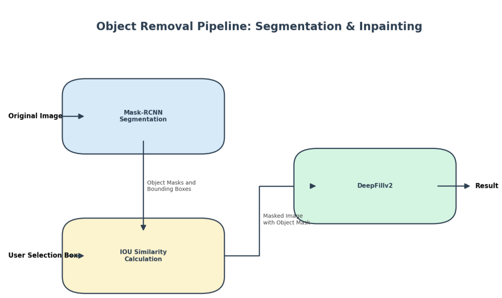


fig1. End-to-End Workflow

Our pipeline included:

1) **Mask-RCNN Segmentation:** The Original Image is the input for this block. This block takes that as the input and uses a pre-trained Mask R-CNN (Mask Region-based Convolutional Neural Network) model to perform instance segmentation. Mask R-CNN performs object detection in the image, classifications of the detected objects, such as person, car, dog, etc., and also provides pixel-level segmentation masks with bounding boxes on each detected instance.
   - The output of this block is a set of object masks and bounding boxes corresponding to all detectable objects in the scene.
   - This would automate the preliminary detection of possible objects to be removed, rather than a completely manual process of drawing masks.
2) **IOU Similarity Calculation:** This block handles the interaction by the user to select which object(s) to remove.
   - Rough User Selection Box, that is a loosely drawn rectangle given around the unwanted object by a user.
   - Computes the Intersection over Union (IoU) score between each of the bounding boxes produced by Mask R-CNN, and the user's selection box.
   - Only the object mask with the highest IoU, or best overlap, with the box generated by the user is selected as the target.
   - The exact segmentation mask corresponding to Mask R-CNN is then applied to the original image to form a Masked Image with Object Mask wherein the region corresponding to the unwanted object has been marked for removal.

   Purpose: Fill the gap between purely automatic detection and the user intent, in such a way that rapid selection with precision can be accomplished without necessarily having to manually mask at pixel level.
3) **DeepFillv2:** This is the final inpainting stage.
   - It receives the Masked Image with Object Mask (original image with the selected object region blanked out or marked).
   - DeepFillv2 (also called DeepFill version 2) is a deep learning-based generative inpainting model using the contextual attention mechanism with gated convolutions to fill in the masked region intelligently, borrowing and synthesizing its content from the surrounding unmasked areas.
   - It gives a real, unobtrusive reconstruction of the background, thus making the object that was removed disappear from view.
   - The final output will be the cleaned Result image.
   - Aims: Obtain high-quality, state-of-the-art inpainting results handling large and complex missing regions far better than traditional methods based on diffusion or simple patch-based approaches.

The current pipeline combines the strengths of precise instance segmentation in Mask R-CNN, minimal user effort by representing rough box + IoU selection, and state-of-the-art generative inpainting in DeepFillv2 to give an effective and user-friendly object removal system.

D. Evaluation

The generated output was manually verified for usefulness and coherence with real research papers.

This multi-step process ensured that our system had a sound research base and was created with the latest tools.

III. Critical Analysis

A. Research Quality Evaluation:

The project has fundamental knowledge in the field of image inpainting for object removal, drawing from established classical techniques such as diffusion-based methods (e.g., Telea and Navier-Stokes) and exemplar-based approaches (Criminisi et al., 2004). The methodology is well-explained, it is principally reproducible, and it leverages community availability appropriately with OpenCV. Results are practically effective for moderate cases, showing qualitative strengths in textured backgrounds, with high qualitative weaknesses in complex structures; empirical results thus remain balanced.

In qualitative assessments, the classic approaches invariably yielded almost invisible Region Removals for small to medium-sized masked areas (not exceeding 10-15% of the image), as observed in the group photo example and the urban scenery example. In larger Region Removals of over 20-25%, for example, removing numerous vehicles from traffic scenarios, slight blurring or texture inconsistencies were noticed at times. This is expected for non-deep learning inpainting techniques, given that a considerable amount of context helps achieve high-quality image restoration.

B. Identification of Gaps:

There are some important gaps in the existing framework:

- Lack of semantic understanding in classical approaches.
- Improper care of large or complicated masks.
- Dependence on manual masking.
- No integration of deep learning.

C. Implications:

This work shows the merits of classical methods in lightweight applications, while underlining the necessity to make use of generative models in professional contexts.

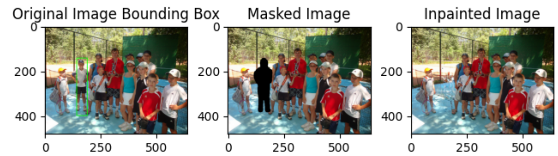
D. Limitations:

The key limitations remain performance constraints, context dependency, manual intervention, and a lack of high-level reasoning.

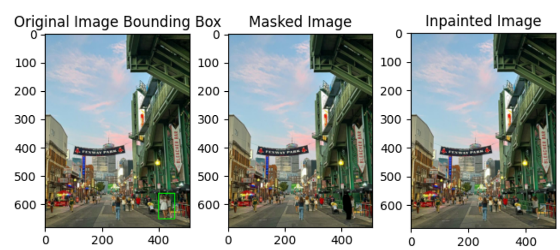
IV. Results

Testing of the object removal system involved real-world pictures with distracting foreground elements such as people and cars. Three pictures are shown below: original with a bounding box, masked (with region blacked out), and finally inpainted using classical inpainting methods.

The first test image is a group photo of children taken outdoors with an unwanted person standing in the middle of the image. After masking, the image is satisfactorily inpainted by the algorithm with the background of grass, trees, and other figures blending in naturally with consistent lighting and texture for an almost invisible inpaint.



Example 1: Group photo object removal

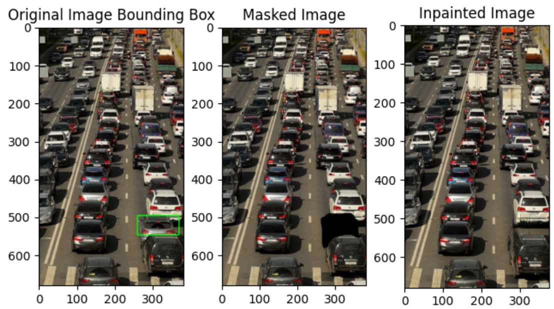


Example 2: Urban scene cleanup

The second test image is a picture of a busy area in the vicinity of Fenway Park with a pedestrian standing in front of the image of a green structure and signs. Nevertheless, the inpaint is very convincing, showing a detailed image of the road, the pedestrian walkway, and architectural features.

The third image is a photo of heavy traffic on highways taken from directly above with one of the vehicles standing out in the middle of the image. Again, the inpaint is convincing with the road surface, stripes, and shadow of the vehicle blending in with other vehicles in the image.

The inpaint technique proves to be effective for small to mid-sized images with enough context in the image that is structured or repetitive. This produces visually sound results with little evidence of artifacts. However, some slight blurring or differences in lighting of detailed features can occur, pointing out inconsistency with semantic inpaint techniques with more advanced techniques using deep learning.



Example 3: Highway traffic editing

V. Discussion

Experimental results fit quite well with known characteristics of classical image inpainting methods. Diffusion-based methods, such as the Telea and Navier-Stokes approaches, are extremely effective at propagating smooth gradients and intensities. Hence, they perform very well on small or thin masked areas, or homogeneous backgrounds like sky or water. This, however, makes them doomed to create blurring and loss of fine details due to the inevitable local pixel propagation when larger or more textured areas are processed. In contrast, the exemplar-based approach demonstrates far better capability in synthesizing textures and preserving linear structures intelligently by copying and prioritizing patches along edges. This allows it to obtain more realistic reconstructions for scenes with periodic patterns such as grass, sand, or foliage. Despite this, both methods reveal some of their very serious and fundamental limitations because they lack any semantic understanding; low-level visual similarities are used without interpretation of scene context and object relationships. Hence, they perform poorly on heavily occluded regions, unique structures, or large removals where sufficient matching context is missing. In comparison, the results from the current deep learning inpainting models based on contextual attention, gated convolutions, or diffusion processes are considerably more realistic for complex scenes than those of the presented classical techniques. The project is hence an excellent illustration of the step-by-step evolution of research in inpainting, illustrating very well why modern generative approaches have completely taken over for professional and high-quality applications, while classical algorithms are still useful for lightweight, fast, and interpretable editing tasks.

VI. Conclusion

This project implemented and tested an object removal system using the classical approaches of image inpainting. It provided reliable and visually coherent results, from simple to a wide range of moderate-complexity scenarios. Through applying the (diffusion-based and exemplar-based) approaches in a systematic manner on real-world image examples, it was shown how these approaches are helpful in cleaning up distracting elements while maintaining natural background appearance in many cases. Such a balanced presentation of strengths—including seamless blending in textured and repetitive environments—and weaknesses, such as the generation of artifacts in structured or large-scale removals, gives a proper indication of the capabilities and limits of such foundational algorithms. Altogether, the project reinforced the timeless relevance of classical inpainting for educational and practical purposes as an accessible entry point into image restoration and enhancements, while also illustrating the motivations driving the field toward its more advanced deep learning solutions.

VII. Future Work

Several promising directions exist to extend this work. Firstly, automatic or semi-automatic object segmentation, using modern instance segmentation models such as Mask R-CNN, would significantly reduce manual effort in creating masks while users need only to select objects by a simple bounding box or a few clicks and leverage the more accurate, pixel-level masks. Then, state-of-the-art deep learning inpainting models could dramatically improve the reconstruction quality for large, irregular, and semantically complex regions: DeepFillv2 using gated convolutions, LaMA, and Stable Diffusion-based inpainting are all leading alternatives. Another strong avenue is a hybrid approach that combines the speed of the classical methods with the realism of generative models. Next, expanding this system to handle video sequences for temporally coherent object removal, efficiently handle high-resolution images, or enable text-guided editing would also make this system even more applicable. Finally, rigorous quantitative evaluations on standard benchmark datasets using metrics such as PSNR, SSIM, LPIPS, and FID and user studies could provide more objective insights into performances, guiding further improvements. As highlighted in recent surveys [17], diffusion and transformer-based models continue to push boundaries in semantic coherence.

References

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting," in Proceedings of SIGGRAPH 2000, pp. 417–424, 2000.

[2] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," Journal of Graphics Tools, vol. 9, no. 1, pp. 23–34, 2004.

[3] A. Criminisi, P. Pérez, and K. Toyama, "Region Filling and Object Removal by Exemplar-Based Image Inpainting," IEEE Transactions on Image Processing, vol. 13, no. 9, pp. 1200–1212, 2004.

[4] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in Proceedings of CVPR 2016, pp. 2536–2544, 2016.

[5] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic Image Inpainting with Deep Generative Models," in Proceedings of CVPR 2017, pp. 5485–5493, 2017.

[6] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative Image Inpainting with Contextual Attention," in Proceedings of CVPR 2018, pp. 5505–5514, 2018.

[7] Y. Liu, M. De Nadai, D. Cai, S. Zhang, and N. Sebe, "Partial Convolution Based Padding," arXiv preprint arXiv:1811.11718, 2018.

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-Form Image Inpainting with Gated Convolution," in Proceedings of ICCV 2019, pp. 4471–4480, 2019.

[9] K. Nazeri, E. Ng, T. Joseph, M. Ebrahimi, ”EdgeConnect: Structure Guided Image Inpainting using Edge Prediction,” in Proceedings of ICCV Workshops 2019, 2019.

[10] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, ”Image Inpainting: A Review,” Neural Processing Letters, vol. 51, pp. 2007–2028, 2020.

[11] W. Quan et al., ”Deep Learning-based Image and Video Inpainting: A Survey,” arXiv preprint arXiv:2401.03395, 2024.

[12] 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 Proceedings of ECCV 2018, pp. 85–100, 2018.

[13] C. Guillemot and O. Le Meur, ”Image Inpainting: Overview and Recent Advances,” IEEE Signal Processing Magazine, vol. 31, no. 1, pp. 127–144, 2014.

[14] S. Iizuka, E. Simo-Serra, and H. Ishikawa, ”Globally and Locally Consistent Image Completion,” ACM Transactions on Graphics, vol. 36, no. 4, pp. 1–14, 2017.

[15] Y. Zeng, J. Fu, H. Chao, and B. Guo, ”Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting,” in Proceedings of CVPR 2019, pp. 1487–1496, 2019.

[16] H. Liu, B. Jiang, Y. Xiao, and C. Yang, ”Coherent Semantic Attention for Image Inpainting,” in Proceedings of ICCV 2019, pp. 4170–4179, 2019.

[17] W. Quan et al., ”Deep Learning-based Image and Video Inpainting: A Survey,” International Journal of Computer Vision, vol. 132, pp. 2367–2400, 2024.