

# **Data-Driven Image Inpainting Approaches for Removing Handwritten and Scene Text from Natural Images.**

**Somanka Maiti (P18MA002)**

**Supervisor: Dr. Gaurav Bhatnagar**

**Department of Mathematics, IIT Jodhpur**

## **Abstract**

Image inpainting is a key problem in computer vision that aims to fill in missing or damaged parts of an image so that the result looks plausible. Over time, techniques for inpainting have changed from classical mathematical models to data-driven deep learning methods. Diffusion-based and exemplar-based methods and other classical methods work well for small and organized gaps but causes trouble when it comes to handle complex, and irregular occlusion, and cannot restore complicated textures and semantics. These shortcomings are more visible in actual situations where images are usually damaged by handwritten or printed texts, as this adds irregular patterns, sharp edges, and strong contrast changes. While CNNs and GANs have significantly improved inpainting quality, they often produce oversmoothed results, require extensive paired training data, and struggle with global structural consistency for complex or irregular occlusions. The primary objective of this thesis is to develop effective image inpainting frameworks for removing handwritten and scene text from natural images.

The work systematically reviews the evolution of inpainting techniques, from regression-based classical approaches to generative and transformer-based modern frameworks. Initially, a regression-based inpainting model is proposed in the thesis to predict missing pixels values based on K-NN Regression and Support Vector Regression between the neighbourhood pixels. This model serves as an interpretable baseline and highlights the limitations of purely classical methods

To address the lack of realistic training data, a handwritten-aware dataset is introduced by overlaying handwritten masks on natural images. This dataset includes variations in stroke thickness, shape, and coverage, which helps to develop inpainting models that are more representative of real-world situations involving the removal of handwritten text. Building on this dataset, a CycleGAN-based framework is proposed to perform handwritten text removal using unpaired image translation, avoiding the need for paired ground truth data.

Further, the thesis proposes a residual Swin Transformer based architecture for handwritten text removal by capitalizing shifted window self attention to capture long range dependencies and contextual relationships. To make this model applicable for scene text removal, an architectural enhancement is made where the last inpainting block of the previous architecture is replaced with pixel shuffle up-sampling modules. The reconstruction reorganize feature representations while preserving the spatial resolution of the input image, resulting in improved texture consistency and cleaner background restoration. The extended models are tested on standard scene text datasets like SCUT-EnsText and SCUT-Syn, and compared with state of the art text removal methods.

The experimental results prove that transformer based inpainting, when combined with suitable reconstruction strategies and realistic datasets, outperforms traditional CNN and GAN based approaches for both handwritten as well as scene text removal. The results of this thesis work will help in creating strong and widely applicable image inpainting frameworks for real-world applications involving text removal from images.

### **List of Publications:**

1. S. Maiti, A. Kumar, S. Jain, and G. Bhatnagar, "A Novel Image Inpainting Framework Using Regression," ACM Transactions on Internet Technology (TOIT), vol. 21, no. 3, pp. 1–16, 2021.
2. S. Maiti, S. Nath Panuganti, G. Bhatnagar, and J. Wu, "Efficient Image Inpainting For Handwritten Text Removal using CycleGan framework," Mathematics, vol. 13, no. 1, p. 176, 2025.
3. S. Maiti, S. Nath Panuganti, G. Bhatnagar, "Bridging the Gap: A Novel Dataset For Handwriting Aware Image Inpainting," International Journal Image, Graphics and Signal Processing, (under review).
4. S. Maiti, S. Nath Panuganti, G. Bhatnagar, "RSwinHTR: Residual Swin Transformer based HandWritten Text Removal," IEEE Transactions on Emerging Topics in Computational Intelligence, (under review).
5. S. Maiti, S. Nath Panuganti, G. Bhatnagar, "SwinErase-R: A Robust Framework for Scene Text Removal Using Residual Swin Transformer," (manuscript to be submitted).