Image Enhancement

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF TECHNOLOGY

IN

Computer Science Engineering

Submitted by:

Sahil Patil (21011102074)

Maran V (21011102059)

Under the supervision of

Prof.(Dr.) Vegesna SM Srinivasavarma

SHIV NADAR UNIVERSITY CHENNAI

DEPT. OF Computer Science Engineering

SHIV NADAR UNIVERSITY CHENNAI

Rajiv Gandhi Salai (OMR), Kalavakkam, Chennai-603110

November 2024

CANDIDATE'S DECLARATION

We, Sahil Patil (21011102074) & Maran V (21011102059) students of B.Tech (IOT), hereby declare that the Project Dissertation titled — "Image Enhancement" which is submitted by us to the Department of Computer Science Engineering, Shiv Nadar University Chennai in fulfillment of the requirement for awarding of the Bachelor of Technology degree, is not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma, Fellowship or other similar title or recognition.

Place: Chennai
Date: 08/11/2024

Sahil Patil
(21011102074)

Maran V
(21011102059)

CERTIFICATE

I hereby certify that the Project titled ”Image Enhancement” which is submitted by Sahil Patil (21011102074) & Maran V (21011102059) for fulfillment of the requirements for awarding of the degree of Bachelor of Technology (B.tech) is a record of the project work carried out by the students under my guidance & supervision. To the best of my knowledge, this work has not been submitted in any part or fulfillment for any Degree or Diploma to this University or elsewhere.

Place : Chennai

Date : 08/11/2024

Prof.(Dr.) Vegesna SM Srinivasavarma

(SUPERVISOR)

Professor

Department of CSE

Shiv Nadar University Chennai

ABSTRACT

Keywords - Vision Transformers , DIBCO

Document images often suffer from degradation, impacting recognition and processing accuracy. This project addresses these challenges through a novel encoder-decoder architecture based on vision transformers to enhance both machine-printed and handwritten document images in a fully end-to-end manner. Unlike traditional methods, the encoder operates on pixel patches and their positional information without using convolutional layers, while the decoder reconstructs a clean image from these encoded patches. Experimental results on multiple DIBCO benchmarks demonstrate that this model outperforms state-of-the-art methods, highlighting its potential for robust document image enhancement in the digital era.

ACKNOWLEDGEMENT

The successful completion of any task is incomplete and meaningless without giving any due credit to the people who made it possible without which the project would not have been successful and would have existed in theory.

First and foremost, we are grateful to **Dr. T Nangarajan**, HOD, Department of Computer Science Engineering,Shiv Nadar University Chennai, and all other faculty members of our department for their constant guidance and support, constant motivation and sincere support and gratitude for this project work. We owe a lot of thanks to our supervisor, **Dr. Vegesna SM Srinivasavarma**, Professor, Department of Computer Science,Shiv Nadar University Chennai for igniting and constantly motivating us and guiding us in the idea of a creatively and amazingly performed Major Project in undertaking this endeavor and challenge and also for being there whenever we needed his guidance or assistance.

We would also like to take this moment to show our thanks and gratitude to one and all, who indirectly or directly have given us their hand in this challenging task. We feel happy and joyful and content in expressing our vote of thanks to all those who have helped us and guided us in presenting this project work for our Major project. Last, but never least, we thank our well-wishers and parents for always being with us, in every sense and constantly supporting us in every possible sense whenever possible.

Sahil Patil (21011102074) **Maran. V (21011102059)**

Contents

List of Tables

List of Figures

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

- **Symbols:**
  - I: Input image
  - Ienhanced: Enhanced output image
  - T: Threshold value for binarization
  - L: Loss function for training
  - p: Patch size in transformers
- **Abbreviations:**
  - **DIBCO**: Document Image Binarization Contest
  - **OCR**: Optical Character Recognition
  - **CNN**: Convolutional Neural Network
  - **ViT**: Vision Transformer
  - **MAE**: Mean Absolute Error
  - **MSE**: Mean Squared Error
- **Nomenclature:**
  - **Binarization**: Process of converting images into black and white format.
  - **Transformer**: A deep learning model architecture based on attention mechanisms.
  - **Degradation**: Loss of image quality due to noise, fading, or artifacts.

Chapter 1

INTRODUCTION

1.1 Overview

In the digital age, the quality of document images plays a crucial role in various applications, from archival digitization and text recognition to digital communication. However, document images, especially those scanned from physical copies or captured under suboptimal conditions, are prone to degradation. Common issues include noise, blurriness, and low resolution, all of which hinder accurate processing and recognition by both human users and automated systems. These degradation factors present a significant challenge in tasks like optical character recognition (OCR) and content retrieval, where clarity and detail are essential for reliable results.

The demand for high-quality document images has driven research into image enhancement techniques that can effectively restore degraded content. Traditional approaches, often reliant on convolutional neural networks (CNNs), have shown considerable success but face limitations in generalizing across diverse document types and degradation patterns. As a result, researchers are now exploring alternative architectures that better capture and reconstruct fine-grained details.

In this project, we introduce a novel encoder-decoder architecture based on vision transformers designed to enhance both machine-printed and handwritten document images. By directly operating on pixel patches and their positional information, this architecture avoids the need for convolutional layers, providing a unique approach to document enhancement. The end-to-end model structure ensures that the encoding captures rich contextual information, while the decoding reconstructs the image in a clean, high-

resolution form.

1.2 Problem Formulation

The degradation of document images, whether through noise, blurriness, or low resolution, poses a major obstacle in the digitization and processing of valuable information. These degradation issues can result from various factors, including low-quality scanning equipment, aging or damaged documents, and inconsistent lighting conditions during capture. As a result, degraded document images lead to substantial difficulties in tasks such as optical character recognition (OCR), where fine details are essential for accurate text extraction.

1.3 Objectives

- Enhance Document Image Quality: Develop an effective model to restore clarity and detail to degraded document images, including both machine-printed and handwritten text.
- Improve Recognition Accuracy: Ensure that enhanced images allow for more accurate optical character recognition (OCR) and other image-processing tasks.
- Utilize Non-Convolutional Architecture: Employ a vision transformer-based architecture that processes pixel patches and positional information without relying on convolutional layers, addressing limitations in traditional CNN-based approaches.
- Outperform Existing Methods: Achieve superior performance on standard benchmarks, particularly the DIBCO dataset, demonstrating the proposed model's effectiveness against state-of-the-art methods.

1.4 Motivation

This project seeks to address these challenges with a novel approach: a vision transformer-based encoder-decoder architecture. Unlike CNNs, vision transformers excel at capturing

global relationships by processing images as sequences of pixel patches. This shift enables the model to grasp complex structures and fine-grained details in document images without convolutional operations, a characteristic that could redefine how degraded document images are enhanced. By incorporating positional information directly within the encoding process, this method aligns more closely with the natural structure of documents, where layout and spatial relationships are crucial.

Our approach is motivated by the vision of providing an end-to-end solution that improves both machine-printed and handwritten document images, making them more accessible and recognizable for automated systems. Experimental results indicate the model's potential to set new standards in document image quality restoration, outstripping traditional methods and pushing the boundaries of current image enhancement techniques.

Chapter 2

BACKGROUND

2.1 How are vision transformers superior to traditional methods?

Document image processing has long been a focal area in computer vision and machine learning, driven by the need to digitize, store, and retrieve information from physical documents. This need spans industries such as healthcare, law, education, and historical preservation, where digitized documents enable faster, more accessible, and scalable data processing. However, real-world document images often suffer from various types of degradation, including noise, low resolution, blurriness, and distortion. Such issues can result from aging, poor scanning quality, or inconsistent lighting during image capture. Degraded document images hinder accurate processing, complicating tasks such as optical character recognition (OCR) and digital archiving.

Over the years, numerous methods have been developed to address document image degradation. Early approaches primarily used basic image processing techniques such as filtering, edge detection, and histogram equalization to reduce noise and enhance contrast. However, these methods were often limited in their ability to handle complex degradations, prompting a shift toward more sophisticated models based on deep learning. Convolutional Neural Networks (CNNs) became widely adopted due to their success in various computer vision tasks, including document enhancement. CNNs leverage convolutional layers to extract features from images, allowing models to reconstruct and enhance degraded images through layered transformations. Despite their success, CNN-

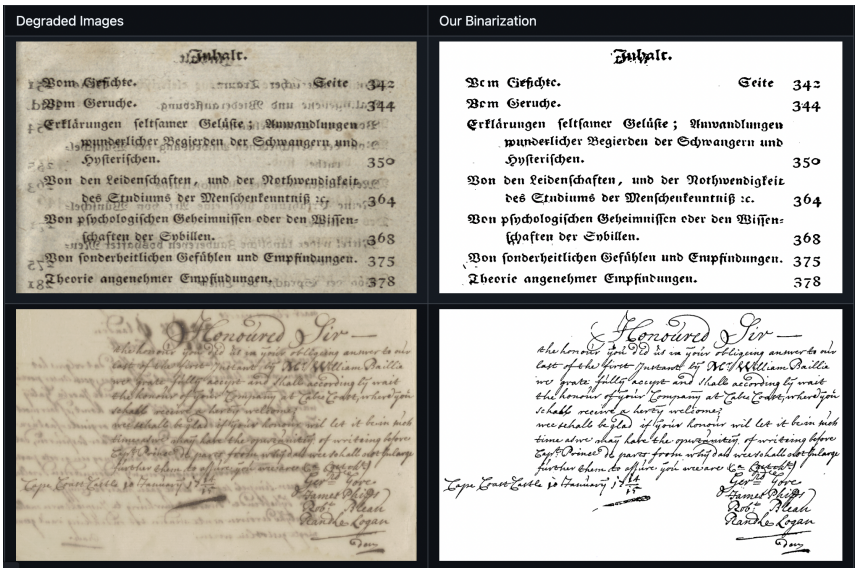

Figure 2.1: Enhancement

based methods struggle with certain limitations. For instance, CNNs may have difficulty capturing long-range dependencies and contextual information across the document, especially in cases where the degradation pattern varies spatially.

Recent advancements in machine learning have introduced vision transformers, a new architecture initially developed for natural language processing and later adapted for image analysis. Unlike CNNs, vision transformers operate on sequences of image patches rather than relying on convolutional filters. By doing so, they can capture global relationships within an image, enabling them to effectively handle complex structures and long-range dependencies. This characteristic is particularly advantageous for document images, where understanding spatial layout and relationships between elements (such as text and whitespace) is crucial for accurate reconstruction.

The emergence of vision transformers offers a promising solution to document image enhancement. By using an encoder-decoder structure, transformers can learn to represent the global context of an image, which is vital for reconstructing fine-grained details in both machine-printed and handwritten text. This project builds on these recent advances, applying a transformer-based approach to enhance document images in an end-to-end framework, ultimately seeking to address the limitations of traditional CNN-based methods and provide a more robust solution for diverse document types and degradation scenarios.

Chapter 3

CONCLUSION

This project presents a novel vision transformer-based approach to document image enhancement, addressing the persistent challenges of noise, blurriness, and other forms of degradation common in both machine-printed and handwritten documents. By utilizing an encoder-decoder architecture that processes pixel patches with positional information, the model bypasses traditional convolutional operations, demonstrating a new way to capture both local and global image contexts effectively.

Experimental results on several DIBCO benchmarks indicate that this transformer-based model achieves superior performance compared to state-of-the-art methods, reinforcing its potential for applications requiring high-quality document images. By restoring image clarity and improving accuracy in tasks like optical character recognition (OCR), this approach can significantly enhance the digitization process in areas such as archival work, academic research, and legal documentation.

Future work could explore further optimizations of the model and potential adaptations for other document types and more diverse degradation scenarios. The project's code and models will be made publicly available, encouraging continued research and collaboration in document image processing and enhancement.

Appendix

Bibliography

[1] DIBCO : Document Image Binarization Contest 2009-2018

[2] An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale

[3] Mohamed Ali Souibgui, Sanket Biswas, Sana Khamekhem Jemni, Yousri Kessentini, Alicia Fornés, Josep Lladós, Umapada Pal, "DocEnTr: An End-to-End Document Image Enhancement Transformer," [Computer Vision and Pattern Recognition], 2022.

PAPER ACCEPTANCE PROOF

REGISTRATION PROOF

SCOPUS INDEXED CONFERENCE PROOF