DEVELOPING A TRANSFORMER-BASED APPROACH FOR FUSING INFRARED AND VISIBLE IMAGES FOR IMPROVED OBJECT DETECTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

AYTEKIN ERDOGAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

AUGUST 2023

Developing A Transformer-Based Approach for Fusing Infrared and Visible Images for Improved Object Detection

submitted by **AYTEKIN ERDOGAN** in partial fulfillment of the requirements for the degree of **Master of Science in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Director of Institute _____
Dean, **Graduate School of Informatics**

Prof. Dr. Head of Department _____
Head of Department, **Information Systems**

Assoc. Prof. Dr. Supervisor _____
Supervisor, **Department, School**

Assoc. Prof. Dr. Co-supervisor if Exists _____
Co-supervisor, **Department, School**

Examining Committee Members:

Prof. Dr. Committee Member 1
Department, School _____

Assoc. Prof. Dr. Committee Member 2
Department, School _____

Assoc. Prof. Dr. Committee Member 3
Department, School _____

Assist. Prof. Dr. Committee Member 4
Department, School _____

Assist. Prof. Dr. Committee Member 5
Department, School _____

Date: 28.08.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Aytekin Erdogan

Signature :

ABSTRACT

DEVELOPING A TRANSFORMER-BASED APPROACH FOR FUSING INFRARED AND VISIBLE IMAGES FOR IMPROVED OBJECT DETECTION

Erdogan, Aytekin

M.S., Department of Information Systems

Supervisor: Assoc. Prof. Dr. Supervisor

Co-Supervisor: Assoc. Prof. Dr. Co-supervisor if Exists

August 2023, **??** pages

English abstract here

Keywords: A keyword, another keyword, some other keywords

ÖZ

TÜRKÇE BAŞLIK

Erdogan, Aytekin
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Doç. Dr. Supervisor
Ortak Tez Yöneticisi: Doç. Dr. Co-supervisor if Exists

Ağustos 2023, **??** sayfa

Türkçe öz buraya

Anahtar Kelimeler: Bir anahtar kelime, başka bir anahtar kelime, başka anahtar kelimeler

To the memories of my beloved friends Murat Tekin and Ragip Enes Katran

ACKNOWLEDGMENTS

Acknowledgments here

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IF	Image Fusion
VIF	Visual and Infrared Image Fusion
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
GAN	General Adveserial Networks
SR	Sparse Representation
MST	Multi-Scale Transformation
LRR	Low-Rank Representation

CHAPTER 1

INTRODUCTION

Image Fusion is a computer vision task that has been taken place for many years. Gathering all the complementary usefull informations into single image is called image fusion, *a.k.a* IF. Visual and Infrared Image Fusion, *henceforth will be referred to as VIF*, is a subfield of iamge fusion. Since the first study [**?**] in 1989, VIF is actively studied. In the era of AI, new methods such as CNN, GAN, auto-encoder, transformers are also applied to the VIF problem.

1.1 Research Questions

1.2 Contributions of the Study

1.3 Organization of the Thesis

CHAPTER 2

RELATED WORK

Considering the available VIF methods, we can classify the methods into two parts as traditional methods, highly used before the era of AI, and learning based methods. Regardless of the classification, all methods consist of three main parts as image feature extraction, fusion of multiple images' features and reconstruction of the image from fused features. In feature extraction part, features from multiple images are extracted. In fusion part of the algorithm, extracted features are compared and complementary features are tried to be inserted into single feature map or set. In the Reconstruction part, from the fused set of features, image is reconstructed. All related studies try to improve one or more part of the this process.

For the traditional algorithms, there are competitive methods but still they suffer from several shortcomings such as handcrafted steps, time complexity and generalizability. To be more specific sparse representation (SR) based methods such as [**?**] and [**?**] requires dictionary learning which increase time complexity quadratically and they includes handcrafted steps. Multi-scale transformation (MST) based methods such as [**?**] and [**?**], low-rank representation (LRR) based methods such as [**?**, text], saliency-based methods such as [**?**, text] suffer from generalizability. In summary, these studies are frequently utilized to capture various characteristics of images at different scales. The extracted features are merged together using a suitable technique, and the final combined image is reconstructed by reversing the multi-scale process. It's clear that the success of these fusion algorithms heavily relies on the quality of the feature extraction method applied.

2.1 Related Work Section I

CHAPTER 3

USER EXPERIMENT

In this chapter, the details of the user experiment are presented.

3.1 Research Method and Experiment Design

CHAPTER 4

USER EXPERIMENT

In this chapter, the details of the user experiment are presented.

4.1 Research Method and Experiment Design

CHAPTER 5

CONCLUSION AND FUTURE WORK

APPENDIX A

TABLES FOR RELATED WORK CHAPTER

A.1 Summary of the Studies

APPENDIX B

EXTRA MATERIAL

APPENDIX C

INSTRUMENTS AND ETHICAL CLEARANCE