

**BUILDERFORMER: A PANOPTIC SEGMENTATION
APPROACH TO AUTOMATIC FLOOR PLAN
ANALYSIS USING VISION TRANSFORMERS**

Mestiyage Don Pathum Pramuditha Goonathilake

218525N

Master of Science in Artificial Intelligence

Department of Computational Mathematics
Faculty of Information Technology

University of Moratuwa
Sri Lanka

July 2024

BUILDERFORMER: A PANOPTIC SEGMENTATION APPROACH TO AUTOMATIC FLOOR PLAN ANALYSIS USING VISION TRANSFORMERS

Mestiyage Don Pathum Pramuditha Goonathilake

218525N

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

Department of Computational Mathematics
Faculty of Information Technology

University of Moratuwa
Sri Lanka

July 2024

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 01/07/2024

The above candidate has carried out research for the Master's thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. ALARR Thanuja

Signature of the Supervisor:

Date: 01/07/2024

DEDICATION

I would like to express my heartfelt dedication of this thesis to my parents, who have served as the pioneering figures and unwavering pillars that have provided support and inspiration throughout my educational journey. Furthermore, I am deeply grateful to extend this dedication to all the lecturers of the Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa.

ACKNOWLEDGEMENT

I convey my gratitude to everyone who has contributed to completing this research project. Their constant support, guidance and encouragement have greatly benefited me.

Firstly, I want to express my heartfelt gratitude to my supervisor, Dr. ALARR Thanuja, Senior Lecturer in the Department of Computational Mathematics at the Faculty of Information Technology, University of Moratuwa. I thank him for his expert advice, exceptional mentoring, and continuous support. He was instrumental in shaping the direction and quality of this research through his deep knowledge, insightful feedback, and constant commitment to academic excellence.

In addition, I would like to extend my deepest gratitude to my family and friends for their unwavering support. Their continuous encouragement, understanding, and support are determined as a consistent source of motivation and inspiration throughout the research journey. They have been instrumental in my belief that I can do it, and their constant support at times of failure has been indispensable. I am deeply grateful for the loyalty they have shown me.

Finally, I would also like to thank BuilderBid founder Mr. Bo Dillon for accepting and developing my idea to integrate this research into their production-grade software, which will help them streamline the takeoff process of floor plans. I have been able to focus on the correct direction thanks to his immense feedback throughout the research project.

Throughout this research, I referred to numerous research publications. I sincerely appreciate the authors of these works' notable contributions to the field, which greatly enhanced the scope and quality of this project.

ABSTRACT

Automatic floor plan analysis, rooted in computer vision and pattern recognition, seeks to extract meaningful insights from architectural and interior design drawings. Its significance spans across various sectors, including architecture, interior design, real estate, and construction. However, conventional methods, whether rule-based or learning-based, encounter constraints such as manual annotation requirements and the challenge of achieving precise segmentation. This research project endeavors to evaluate the efficacy of vision transformers in tasks crucial to floor plan analysis, including identifying unique objects like doors and windows, extracting line segments to represent room layouts, and segmenting object areas such as room spaces. A novel panoptic segmentation approach is proposed to achieve a holistic understanding of scenes by integrating object detection with image segmentation. This approach involves identifying and delineating individual object instances within regions of interest and subsequently segmenting images into meaningful regions. Key to this methodology is the utilization of the Segment Anything Model for segmentation, guided by input prompts generated through object detection models. The proposed approach introduces a mask classification mechanism within the segmentation stage, diverging from conventional methods of assigning semantic labels to individual pixels. The effectiveness of this approach is demonstrated through promising results obtained on existing datasets. Real-time object detection models like YOLOv8 are employed alongside end-to-end object detection models like RT-DETR to provide visual prompts for the Segment Anything Model. This innovative approach is prepared to revolutionize automatic floor plan analysis in real-world applications by bridging the gap between object detection and image segmentation. As the final step of this research project, the proposed methodology integrates into a web application, enabling users to visualize results for various floor plan types, thus providing a practical tool for professionals in architecture, interior design, real estate, and construction to analyze floor plans with enhanced accuracy and efficiency.

Keywords: Automatic Floor Plan Analysis, Panoptic Segmentation, Real-Time Object Detection, Visual Prompt Engineering, Vision Transformers

TABLE OF CONTENTS

Declaration	i
Dedication	ii
Acknowledgement	iii
Abstract	iv
Table of Contents	v
List of Figures	ix
List of Tables	xii
List of Abbreviations	xiii
List of Appendices	xiv
Chapter 1	1
Introduction	1
1.1 Prolegomena	1
1.2 Aim and Objectives	3
1.3 Background and Motivation	3
1.4 Problem in Brief	5
1.5 A Novel Approach for Automatic Floor Plan Analysis	5
1.6 Resource Requirements	6
1.7 Structure of Thesis.....	6
1.8 Summary	6
Chapter 2	7
Literature Review.....	7
2.1 Introduction	7
2.2 Evolution of Automatic Floor Plan Analysis	7
2.2.1 Rule-based Methods for Automatic Floor Plan Analysis	7
2.2.2 Learning-based Methods for Automatic Floor Plan Analysis.....	9
2.2.3 Machine Learning in Automatic Floor Plan Analysis.....	9
2.2.4 Deep Learning in Automatic Floor Plan Analysis	11
2.2.5 Discriminative-based Models in Automatic Floor Plan Analysis.....	12
2.2.6 Generative-based Models in Automatic Floor Plan Analysis	14

2.3	Recent Advancements in Computer Vision	15
2.3.1	Vision Transformers	16
2.3.2	Panoptic Segmentation.....	18
2.4	Challenges in Automatic Floor Plan Analysis	20
2.5	Existing Datasets	21
2.6	Research Findings	22
2.7	Problem Definition	23
2.8	Summary	24
Chapter 3		25
Technologies Adopted		25
3.1	Introduction	25
3.2	Computer Vision	25
3.2.1	Segment Anything.....	26
3.2.2	You Look Only Once (YOLO)	27
3.2.3	YOLOv8.....	28
3.2.4	Real-Time Detection Transformer (RT-DETR).....	29
3.3	Roboflow	30
3.4	Programming Languages.....	31
3.5	Machine Learning Frameworks.....	31
3.5.1	PyTorch.....	31
3.6	Cloud Computing	32
3.7	User Interface	33
3.8	Summary	33
Chapter 4		34
A Novel Approach for Automatic Floor Plan Analysis.....		34
4.1	Introduction	34
4.2	Hypothesis	34
4.3	Input.....	34
4.4	Output.....	34
4.5	Process.....	35
4.6	Features	36
4.7	Users	36

4.8	Summary	36
Chapter 5		37
Design of Builderformer		37
5.1	Introduction	37
5.2	High-Level Design of the System	37
5.3	Floor Plan Data Preparation	38
5.4	Pre-Processing Module.....	38
5.5	Detection Module	39
5.6	Post-Processing Module	39
5.7	Segmentation Module.....	40
5.8	Inference Module.....	40
5.9	Web App.....	41
5.10	Summary	41
Chapter 6		42
Implementation of Builderformer		42
6.1	Introduction	42
6.2	Data Collection and Preparation.....	42
6.2.1	Floor Plan Dataset Format	42
6.2.2	Data Preprocessing.....	43
6.3	Detection Module Implementation.....	46
6.4	Post-Processing Module Implementation.....	50
6.5	Segmentation Module Implementation	52
6.6	Model Registry	54
6.7	Web App Implementation	55
6.8	Summary	56
Chapter 7		57
Evaluation of Builderformer		57
7.1	Introduction	57
7.2	Experimental Datasets	57
7.3	Evaluation Metrics	57
7.4	Experimental Results.....	58
7.4.1	Results From Experiment 1.....	60

7.4.2	Results From Experiment 2.....	62
7.4.3	Results From Experiment 3.....	64
7.4.4	Results From Experiment 4.....	66
7.4.5	Results From Experiment 5.....	68
7.4.6	Results From Experiment 6.....	69
7.5	Discussion on Experiments	71
7.6	Summary	74
Chapter 8		75
Conclusion and Future Work		75
8.1	Introduction	75
8.2	Overall Conclusion.....	75
8.3	Objective Wise Achievements	76
8.4	Limitations.....	76
8.5	Future Work	77
8.6	Summary	77
References		78
Appendix - A.....		81
Code for Web App		81

LIST OF FIGURES

Figure	Description	Page
Figure 2.1	Methodology for reconstructing a 3D building layout	8
Figure 2.2	VGG model architecture	10
Figure 2.3	Deep learning methods investigated in studies	12
Figure 2.4	U-Net model	13
Figure 2.5	Pix2Pix model	15
Figure 2.6	Current sota vision transformers in computer vision	18
Figure 2.7	Panoptic segmentation	19
Figure 3.1	Segment anything model	26
Figure 3.2	YOLOv8 architecture	29
Figure 3.3	RT-DETR architecture	30
Figure 3.4	Roboflow product overview	31
Figure 4.1	Pipeline of our approach	34
Figure 5.1	Conceptual design of the system	35
Figure 6.1	Annotated floor plan image	42
Figure 6.2	Annotated floor plan image data	43
Figure 6.3	Dataset-1 overview	44
Figure 6.4	Dataset-2 overview	45
Figure 6.5	Dataset-3 overview	45
Figure 6.6	YOLOv8 model summary	46
Figure 6.7	RT-DETR model summary	46
Figure 6.8	YOLOv8 model training	47
Figure 6.9	RT-DETR model training	47
Figure 6.10	Detection module predictions	49

Figure 6.11	Remove unnecessary fields function	50
Figure 6.12	Update image dimensions function	50
Figure 6.13	Roboflow supervision detection plot function	51
Figure 6.14	Plotted detection objects	51
Figure 6.15	Loading sam model	52
Figure 6.16	Prompt encode function	52
Figure 6.17	Roboflow supervision segmentation plot function	53
Figure 6.18	Plotted segmentation masks	53
Figure 6.19	Roboflow deploy function	54
Figure 6.20	Roboflow visualize models dashboard	54
Figure 6.21	BuilderFormer-1	55
Figure 6.22	BuilderFormer-2	55
Figure 6.23	BuilderFormer-3	56
Figure 7.1	Experiment 1 confusion matrix	60
Figure 7.2	Experiment 1 evaluation metrics graphs	61
Figure 7.3	Experiment 1 testing set visualization	61
Figure 7.4	Experiment 2 confusion matrix	62
Figure 7.5	Experiment 2 evaluation metrics graphs	63
Figure 7.6	Experiment 2 testing set visualization	63
Figure 7.7	Experiment 3 confusion matrix	64
Figure 7.8	Experiment 3 evaluation metrics Graphs	65
Figure 7.9	Experiment 3 testing set visualization	65
Figure 7.10	Experiment 4 confusion matrix	66
Figure 7.11	Experiment 4 evaluation metrics graphs	67
Figure 7.12	Experiment 4 testing set visualization	67

Figure 7.13	Experiment 5 confusion matrix	68
Figure 7.14	Experiment 5 evaluation metrics graphs	68
Figure 7.15	Experiment 5 testing set visualization	69
Figure 7.16	Experiment 6 confusion matrix	70
Figure 7.17	Experiment 6 evaluation metrics graphs	70
Figure 7.18	Experiment 6 testing set visualization	71
Figure 7.19	Dataset-1 object detection and segmentation visualization	72
Figure 7.20	Dataset-1 sam visualizations with prompts	72
Figure 7.21	Dataset-2 object detection and segmentation visualization	73
Figure 7.22	Dataset-2 sam visualizations with prompts	73
Figure 7.23	Dataset-3 object detection and segmentation visualization	74
Figure 7.24	Dataset-3 sam visualizations with prompts	74

LIST OF TABLES

Table	Description	Page
Table 2.1	Comparison of Research Findings	22
Table 5.1	Available Datasets	38
Table 7.1	Experimental Results	58
Table 8.1	Objective Wise Achievements	76

LIST OF ABBREVIATIONS

Abbreviation	Description
API	Application Programming Interface
CAD	Computer-Aided Design
CNN	Convolutional Neural Network
DL	Deep Learning
GAN	Generative Adversarial Network
GNN	Graph Neural Network
GPU	Graphics Processing Unit
ML	Machine Learning
RAG	Room Adjacency Graph
RT-DETR	Real-Time Detection Transformer
SAM	Segment Anything Model
SG	Shape Grammar
SOTA	State Of The Art
VIT	Vision Transformer
YOLO	You Only Look Once

LIST OF APPENDICES

Appendix	Description	Page
Appendix - A	Code For Web App	81