

CONTENTS

Halaman Pernyataan	2
ABSTRACT	10
I INTRODUCTION	11
1.1 Research Background	11
1.2 Research Problem	13
1.3 Research Scope	13
1.4 Research Objective	13
1.5 Research Advantages	14
1.6 Writing System	14
II LITERATURE REVIEW	15
III THEORETICAL BASIS	24
3.1 Machine Learning	24
3.1.1 Deep Learning	24
3.1.2 Computer Vision	24
3.2 Face Detection	25
3.2.1 Multi Orientation Face Detection	25
3.3 MobileNetV2	27
3.4 RetinaFace	27
3.4.1 Feature Pyramid Network (FPN)	27
3.4.2 Context Head Module	28
3.4.3 Anchor Settings and Matching Strategy	28
3.5 Receptive Field Block (RFB) Module	28
3.6 Data Augmentations	29
3.7 Intersection over Union (IoU)	30
3.8 Multi Task Loss	30
IV METHODOLOGY	32
4.1 General Research Description	32
4.2 Problem Analysis	32
4.3 Data Collection and Acquisition	33

4.3.1	WIDER FACE Dataset	33
4.3.2	AFLW2000 Dataset	34
4.3.3	HPID Dataset	34
4.4	Algorithm and Model Design	35
4.4.1	Algorithm Overview	35
4.4.2	Random Crop	37
4.4.3	Distortion	37
4.4.4	Image Expansion	38
4.4.5	Random Flip	39
4.4.6	Pad to Square	39
4.4.7	Resize and Normalize	40
4.5	Deep Learning Model Design and Architecture	41
4.5.1	MobileNetV2 Backbone	41
4.5.2	Receptive Field Block (RFB) Integration with RetinaFace	42
4.5.3	Loss Functions and Optimization	44
4.6	Evaluation Metrics and Benchmarking	45
4.6.1	Evaluation Metrics	45
4.6.2	Baseline Comparison	45
4.7	Experimental Setup	45
4.7.1	Training Configuration	45
4.7.2	Validation and Testing	46
4.8	Report Writing	46
V	RESEARCH IMPLEMENTATION	47
5.1	Research Environment	47
5.2	Project Structure	47
5.3	Libraries	48
5.4	Data Augmentation Class	48
5.5	Load Training Data	50
5.6	Prior Box Generation	51
5.7	MultiBox Loss Implementation	53
5.7.1	Initialization	53
5.7.2	Forward Pass	54
5.7.3	Target Preparation	55
5.7.4	Landmark Loss Calculation	56

5.7.5	Localization Loss Calculation	57
5.7.6	Confidence Loss Calculation	57
5.7.7	Loss Normalization	59
5.8	RetinaFace	60
5.8.1	Backbone Feature Extractor	60
5.8.2	Feature Pyramid Network (FPN)	61
5.8.3	Single-Stage Headless (SSH) Modules	62
5.9	RFB Module implementation	64
5.9.1	Branches	65
5.9.2	Feature Aggregation	66
5.9.3	Shortcut Connection	66
5.10	Training Procedure for a Single Epoch	68
5.11	Main Training Function	69
5.11.1	Dataset Preparation	70
5.11.2	Loss Function and Model Initialization	70
5.11.3	Optimization and Learning Rate Scheduling	70
5.11.4	Training Loop	71
5.11.5	Final Model Saving	71
5.12	WIDER FACE Evaluation	73
5.12.1	Inference Function	73
5.12.2	Image Resizing	73
5.12.3	Main Evaluation Loop	74
5.13	AFLW2000-3D Evaluation	77
5.13.1	Inference Function	77
5.13.2	Face Detection	78
5.13.3	Ground Truth Computation	79
5.14	HPID Head Pose Database Evaluation	81
5.14.1	Batched Inference	82
5.14.2	Batched RetinaFace Detection	82
5.14.3	Bounding Box Conversion	83
5.14.4	Image and Annotation Loading	84
VI	RESULTS AND DISCUSSION	86
6.1	Training Parameters	86
6.2	RetinaFace Model training	87

6.3	RetinaFace Model with RFB modules training	88
6.4	Results on WIDER FACE	88
6.5	Results on AFLW2000	89
6.5.1	Precision Increases with Smaller Scales	90
6.6	Results on HPID	91
6.6.1	Performance at Original Scale (Longer Side: 640 Pixels)	91
6.6.2	Performance at Reduced Scales (Longer Side: 128, 64, 56, and 48 Pixels)	92
6.6.3	Observations on HPID Results	100
VI CONCLUSION AND FUTURE RESEARCH		103
7.1	Conclusion	103
7.2	Future Research	103

LIST OF TABLES

2.1	Comparison of the different face detection models	23
5.1	Specifications of the machine used	47
6.1	MobileNet V2 Configuration	86
6.2	Average Accuracy	88
6.3	Face Detection Performance on Modified AFLW Dataset Variants. . .	89

LIST OF FIGURES

2.1	Output heatmaps visualized (Farfade et al., 2015)	16
2.2	SSH architecture (Najibi et al., 2017)	18
2.3	Rapidly Digested Convolutional Layers (RDCL) (Zhang et al., 2017)	19
2.4	Dual Shot Architecture (Li et al., 2019)	20
2.5	RetinaFace Architecture (Deng et al., 2020)	21
3.1	The three axis of orientation (Arcoverde et al., 2014)	26
3.2	Construction of the RFB Module (Liu et al., 2018)	29
3.3	Illustration of IoU calculation from two images (Rosebrock, 2016)	30
4.1	Examples from the WIDER Face dataset (Yang et al., 2016)	33
4.2	Algorithm Overview	36
4.3	A grid of images: The original image (top left) and three examples after the distortion process.	38
4.4	An example of a flipped image	39
4.5	A photo from the WIDER FACE dataset (Yang et al., 2016) (left), and an example image after undergoing the pad to square process	40
4.6	A photo from the WIDER FACE dataset (Yang et al., 2016) (left), and an example image after the resizing and normalizing	41
4.7	Implementation of Receptive Field Block (RFB) modules (Liu et al., 2018) in the RetinaFace architecture (Deng et al., 2020)	44
5.1	Project directory structure	47
5.2	Augmentation class code	49
5.3	WiderFace dataset helper class	51
5.4	PriorBox class to generate priorboxes for a given image	53
5.5	MultiBoxLoss Class Initialization	54
5.6	Forward Pass Entry	55
5.7	Target Preparation and Matching	56
5.8	Landmark Loss Calculation	57
5.9	Localization Loss Calculation	57
5.10	Confidence Loss Calculation with Hard Negative Mining	59
5.11	Loss Normalization and Return	59
5.12	FPN Module	62

5.13 SSH Module	63
5.14 RetinaFace class	64
5.15 RFB Module implementation	67
5.16 train_one_epoch function to carry out training for one epoch	68
5.17 Main function in train.py	72
5.18 Inference function	73
5.19 resize_image function	74
5.20 Model Configuration and Device Setup	75
5.21 Image Processing and Inference	76
5.22 Post-Processing and Results Decoding	77
5.23 Inference Function	78
5.24 Face Detection Function	79
5.25 Ground Truth Computation	80
5.26 Evaluation and Metrics	81
5.27 Batched Inference Function	82
5.28 Batched RetinaFace Detection Function	83
5.29 Bounding Box Conversion Function	84
5.30 Image and Annotation Loading Function	85
6.1 F1 Score on HPID scaled to 128 pixels (longer side) without RFB module	93
6.2 F1 Score on HPID scaled to 128 pixels (longer side) with RFB module	94
6.3 F1 Score on HPID scaled to 64 pixels (longer side) without RFB module	95
6.4 F1 Score on HPID scaled to 64 pixels (longer side) with RFB module	96
6.5 F1 Score on HPID scaled to 56 pixels (longer side) without RFB module	97
6.6 F1 Score on HPID scaled to 56 pixels (longer side) with RFB module	98
6.7 F1 Score on HPID scaled to 48 pixels (longer side) without RFB module	99
6.8 F1 Score on HPID scaled to 48 pixels (longer side) with RFB module	100
6.9 Examples of failed detections, red boxes are ground truth and green boxes are detection boxes	101