Contents

Acknowledgements i

Abstract iii

Zusammenfassung v

1 Introduction 1

2 Datasets and Metrics 5
- 2.1 Introduction . . . 5
- 2.2 Datasets . . . 7
  - 2.2.1 Cityscapes . . . 8
  - 2.2.2 PASCAL visual object classes (VOC) 2012 . . . 9
  - 2.2.3 Adverse conditions with diverse correspondences (ACDC) . . . 10
- 2.3 Semantic Segmentation Metrics . . . 11
  - 2.3.1 Mean intersection-over-union (mIoU) . . . 11
  - 2.3.2 Mean performance-under-corruption (mPC) . . . 12
- 2.4 Denoising Metrics . . . 13
  - 2.4.1 Mean squared error (MSE) . . . 13
  - 2.4.2 Structural similarity index measure (SSIM) . . . 14
  - 2.4.3 Signal-to-noise ratio (SNR) . . . 15
  - 2.4.4 Peak signal-to-noise ratio (PSNR) . . . 16

3 Perception and Robustness 17
- 3.1 Introduction . . . 17
- 3.2 Perception . . . 18
  - 3.2.1 Camera-based perception . . . 19
  - 3.2.2 Multi-sensor perception . . . 20
- 3.3 Semantic Segmentation Task . . . 22
  - 3.3.1 Mathematical notations . . . 22
  - 3.3.2 Current state of research . . . 24
  - 3.3.3 State of the art . . . 24
- 3.4 Baseline Models . . . 30
  - 3.4.1 ICNet . . . 30
  - 3.4.2 DeepLabv3+ . . . 32
  - 3.4.3 Baseline performance . . . 35
  - 3.4.4 Baseline robustness . . . 36
- 3.5 Summary . . . 39

4 Data Impairments and State-of-the-art Robustness Methods **41**
4.1 Introduction . . . 41
4.2 Data Impairments . . . 42
4.2.1 Comparison of corruptions and attacks . . . 43
4.3 State of the Art Robustness Methods . . . 45
4.3.1 Overview . . . 45
4.3.2 Baseline robustness methods . . . 48
4.4 State-of-the-Art Adversarial Defense Methods . . . 50
4.4.1 Overview . . . 51
4.4.2 Baseline Adversarial Defenses . . . 53
4.5 Summary . . . 55

5 Wiener Filtering as a Novel Adversarial Defense Preprocessor **57**
5.1 Introduction . . . 57
5.2 A Novel Frequency Domain Perspective on Adversarial Attacks . . . 59
5.2.1 Mathematical notations . . . 59
5.2.2 Data analysis . . . 60
5.2.3 Periodic artifacts in neural networks . . . 64
5.2.4 Effect of upsampling methods in the frequency domain . . . 68
5.3 Wiener Filter as Defense Method . . . 72
5.4 Experimental Setup . . . 75
5.5 Experimental Results . . . 76
5.5.1 Selecting the right Wiener filter . . . 76
5.5.2 Robustness to attacks . . . 80
5.5.3 Robustness to corruptions . . . 85
5.6 Summary . . . 87

6 VQVAE as a Novel Adversarial Defense Pre-processor **89**
6.1 Introduction . . . 89
6.2 Method . . . 91
6.2.1 Mathematical notations . . . 91
6.2.2 Vector-quantized variational autoencoders (VQVAE) . . . 92
6.2.3 VQVAE as a denoising autoencoder . . . 94
6.2.4 Training strategy . . . 95
6.3 Experimental Setup . . . 97
6.3.1 Employed datasets and models . . . 97
6.3.2 Denoising autoencoders . . . 98
6.4 Experimental Results . . . 99
6.4.1 Codebook ablation experiments . . . 99
6.4.2 Robustness evaluation . . . 101
6.5 Summary . . . 107

7 A Novel Self-Supervised FMA Loss to Improve Corruption Robustness **109**
7.1 Introduction . . . 109
7.2 Method . . . 111
7.2.1 Intuition behind the FMA loss . . . 112
7.2.2 Feature map augmentation loss . . . 112
7.3 Experimental Setup . . . 113
7.3.1 Robust training methods . . . 114

7.4 Experimental Results 119
7.4.1 Robust training methods 119
7.4.2 Robust training combined with pre-processors 123
7.5 Summary 126

8 Conclusions **129**

Symbols and Acronyms **135**

Bibliography **141**

Publications **163**

Appendix **165**