

Contents

Acknowledgements	v
Abstract (English/Français)	vii
Contents	xiii
List of Figures	xvi
List of Tables	xvii
1 Introduction	1
1.1 Applications	2
1.2 Contributions	6
1.3 Outline	7
2 Background	9
2.1 Probabilistic Modeling and Inference	9
2.1.1 Markov Random Fields	10
2.1.2 Conditional Random Fields	11
2.1.3 Deep Probabilistic Models	12
2.2 Approximate Inference	12
2.2.1 Markov Chain Monte Carlo	13
2.2.2 Variational Inference	13
2.2.3 Discussion	15
2.3 Mean-Field Variational Inference	15
2.4 Stochastic Optimization for Variational Inference	16
2.5 Related Paradigms	18
2.5.1 Structured Prediction	18
2.5.2 Energy-Based Models	18
2.6 Discussion	19
3 Variational Human Detection in Depth Images	21
3.1 Related Work	22

3.2	Method	24
3.3	Formulation	25
3.3.1	Generative Model	25
3.3.2	Inference	27
3.3.3	Numerical Model	28
3.3.4	Computing Segmentation Masks	29
3.3.5	Implementation Details	29
3.4	Evaluation	30
3.4.1	Datasets	30
3.4.2	Baselines	31
3.4.3	Overall Performance	32
3.4.4	Drones	34
3.5	Discussion	35
4	Efficient Variational Inference in Discrete Random Fields	37
4.1	Related Work	39
4.1.1	Conditional Random Fields	39
4.1.2	Mean-Field Inference	40
4.1.3	Proximal Gradient Descent	42
4.2	Method	43
4.2.1	Proximal Gradient for Mean-Field Inference	43
4.2.2	Fixed Step Size	45
4.2.3	Adaptive Step Size	46
4.2.4	Momentum	46
4.2.5	ADAM	47
4.3	Evaluation	48
4.3.1	Baselines and Variants	48
4.3.2	Experimental Setup	48
4.3.3	Comparative Results	50
4.4	Discussion	53
5	Multi-Human Scene Understanding	55
5.1	Related Work	57
5.2	Method	59
5.2.1	Overview	59
5.2.2	Joint Feature Representation	59
5.2.3	Dense Detections	60
5.2.4	Inference for Dense Detection Refinement	62
5.2.5	Matching RNN for Temporal Modeling	63
5.3	Evaluation	65
5.3.1	Datasets	65

5.3.2	Baselines	65
5.3.3	Implementation Details	66
5.3.4	Multi-Person Scene Understanding	67
5.3.5	Multi-Person Detection	69
5.4	Discussion	70
6	Face Modeling with Compositional Variational Autoencoders	71
6.1	Related Work	72
6.1.1	Parametric Face Models	72
6.1.2	Deep Learning for 3D Face Reconstruction	73
6.1.3	Deep Generative Models	73
6.2	Method	74
6.2.1	Mesh Representation	75
6.2.2	Linear Face Models as Autoencoders	75
6.2.3	Convolutional Mesh VAE	76
6.2.4	Compositional Mesh VAE	77
6.3	Model Fitting	79
6.4	Evaluation	81
6.4.1	Dataset	81
6.4.2	Implementation Details	82
6.4.3	Quantitative Evaluation	83
6.4.4	Qualitative Results	84
6.4.5	Exploring the Latent Space	86
6.5	Conclusion	89
7	Concluding Remarks	91
7.1	Summary	91
7.2	Limitations and Future Work	92
A	Appendix for Chapter 3	95
B	Appendix for Chapter 4	97
B.1	Proximal Gradient Mean-Field Inference	97
B.2	Proving Convergence	98
B.3	Adaptive Steps	100
C	Appendix for Chapter 5	103
D	Appendix for Chapter 6	107
	Bibliography	122

