UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Structured Models for Representation Inference and Data Generation

by

Pingbo Pan

Doctor of Philosophy

Sydney, Australia

2020

Certificate of Authorship/Originality

I, Pingbo Pan declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 20 September 2020

Acknowledgements

I would like to thank my supervisor, Prof. Yi Yang, for his kind and patient guidance and encouragement in my Ph.D. study. He could always give me useful advice when I met problems in my research. I'm lucky to have him as my supervisor who teaches me a lot of things about how to be a good researcher.

I would also like to thank my team members, *e.g.*, Xiaojun Chang, Hehe Fan, Qianyu Feng, Qingji Guan, Yang He, Yanbin Liu, Peike Li, Ping Liu, Yawei Luo, Fan Ma, Jiaxu Miao, Xiaohan Wang, Yunchao Wei, Yu Wu, Zhongwen Xu, Yan Yan, Zongxin Yang, Hu Zhang, Linchao Zhu, Minfeng Zhu, Zhong Zhun, Zhedong Zheng, Liang Zheng, Xiaolin Zhang, and many others. I'm fortunate to work with them and have discussions with them during my Ph.D. study.

Pingbo Pan

Sydney, Australia, 2020.

List of Publications

Journal Papers

J-1. **P. Pan**, Y. Yan and Y. Yang, "Continual Learning with Functional and Weight Regularization at Boundary Points," under review of *IEEE Transactions on Neural Networks and Learning Systems*.

Conference Papers

C-1. **P. Pan**, P. Liu, Y. Yan, T. Yang and Y. Yang, "Adversarial Localized Energy Network for Structured Prediction," *Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 5347-5354, Feb. 7-12, 2020.

C-2. M. Zhu, **P. Pan**, W. Chen and Y. Yang, "EEMEFN: Low-Light Image Enhancement via Edge-Enhanced Multi-Exposure Fusion Network," *Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 13106-13113, Feb. 7-12, 2020.

C-3. M. Zhu, **P. Pan**, W. Chen and Y. Yang, "DM-GAN: Dynamic Memory Generative Adversarial Networks for Text-To-Image Synthesis," *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5802-5810, Jun. 16-20, 2019.

C-4. **P. Pan**, J. Feng, L. Chen and Y. Yang, "Online compressed robust PCA," *The International Conference on Neural Networks (IJCNN)*, pp. 1041-1048, May. 14-19, 2017.

C-5. G. Li, **P. Pan** and Y. Yang, "UTS CAI submission at TRECVID 2017 Video to Text Description Task," *TRECVID*, 2017.

Dedication

I dedicate my dissertation work to my parents, Mr. Xuejun Pan and Mrs. Yuejuan Pan for they always encourage me never to give up and provide me with timely help when I meet troubles in my life.

Contents

List of Figures

Abbreviation

DVN - Deep Value Network

CNN - Convolution Neural Network

GAN - Generative Adversarial Network

IS - Inception Score

FID - Fréchet Inception Distance

PCA - Principal Component Analysis

RPCA - Robust Principal Component Analysis

DNN - Deep Neural Network

EBM - Energy-Based Model

SSVM - Structured Support Vector Machine

Nomenclature and Notation

Bold capital letters denote matrices.

Bold lower-case alphabets denote column vectors.

$(.)^T$ denotes the transpose operation.

I_n is the identity matrix of dimension $n \times n$.

0_n is the zero matrix of dimension $n \times n$.

$\mathbb{R}$ denotes the field of real numbers.

$\mathcal{A}$ denotes the linear compression operator

ABSTRACT

Structured Models for Representation Inference and Data Generation

by

Pingbo Pan

Many problems, *e.g.*, image segmentation, image generation involve generating complex structured outputs that consist of several correlated variables. Developing structured models to effectively capture the correlation among these variables is essential to solve these problems. This dissertation mainly focuses on two different types of structured models: Principle Component Analysis (PCA) and deep generative models. We explore to utilize these two models to predict structured outputs in machine learning problems.

In real-world applications, collected data may contain random noise and data compression is necessary when one transfers high-dimensional and large-scale data. This dissertation presents an online compressed robust PCA model to efficiently recover the structured low-rank component of the high-dimensional data from the compressed data and remove the random noise. Though severe information loss occurs in the data compression process, our method can asymptotically recover the low-rank component under mild conditions. The proposed method is memory efficient since it processes data in an online fashion.

This dissertation presents a special generative model, *i.e.*, an energy network for structured output prediction. The energy network is an implicit model that measures the quality of outputs by assigning different energy values to different output configurations. Previous energy-based methods suffer from substantial computation costs due to enormous amounts of gradient steps in the inference process. Our method addresses this issue by learning an inference network to estimate good initializations and reduce the searching space for the inference process. We propose

a novel framework analogous to the adversarial learning framework to learn the inference network and the energy network. In the proposed framework, the inference network is treated as a generator and the energy network is treated as a discriminator. These two networks can benefit each other mutually in the whole training process. On the one hand, the inference can generate training samples for the energy network. On the other hand, the energy network can evaluate the quality of the generated output from the inference network and provides a guide for the training of the inference network.

This dissertation also presents two works for image generation. The first work is generating realistic images from text descriptions. We propose a novel Dynamic Memory Generative Adversarial Network (DM-GAN) to address two main problems in recent text-to-image methods: (1) The quality of generated high-resolution images heavily depends on the quality of initial low-resolution images; (2) Different words in the text description contributes differently when depicting image contents. Our method introduces a dynamic memory module to refine the initial images and select the important text information based on the initial image content. The second work is generating normal-light images from extremely low-light images. We first generate initial images by utilizing a multi-exposure fusion network to combine well-exposed areas of images with different exposure time. Then, we utilize an edge enhancement module to refine the initial image with the help of the edge information.

Dissertation directed by Professor Yi Yang