SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

BACHELOR THESIS

Range Loss Based Generative Adversarial Network (RLGAN) on Face Recognition and Generation

Author:
Zhiyuan FANG

Supervisor:
Dr. Qi HAO

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Engineering

in the

UAV Lab
Department of Computer Science and Engineering

April 24, 2017

Declaration of Authorship

I, Zhiyuan FANG, declare that this thesis titled, "Range Loss Based Generative Adversarial Network (RLGAN) on Face Recognition and Generation" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"Say 'Hello' to the Era of Artificial Intelligence !"

Jacob

Southern University of Science and Technology

Abstract

Qi Hao
Department of Computer Science and Engineering

Bachelor of Engineering

Range Loss Based Generative Adversarial Network (RLGAN) on Face Recognition and Generation

by Zhiyuan FANG

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Contents

List of Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

For My Devotion.

Chapter 1

Vision and Convolutional Neural Network

1.1 Statistical Learning Based Vision

My general area of research is the study of Statistical Learning and Computational Vision - including the Deep Neural Network and Representation Learning. The ultimate goals for my research are to build intelligent systems based on machine vision techonology that inspired by the understanding of human representation of visual information. In particular, I focuse on theoretical research and engineering methods in the following topics for at this stage:

- Visual Object Recognition and Detection (eg, Face Recognition, Lesion Detection on Human Spine Column)
- Deep Neural Networks (Eg, Convolutional Neural Networks, Siamese Networks, Generative Adversarial Networks)
- Representational Theories

To give an intuitive explaination of vertebrate visual system, vision is the process of discovering from images what is present around and where it is. In fact, to better coordinate the sensorium and motorium, representing the visual information (geometric shape, color and texture information) is indispensable and crucial significant. The same truth also applies to the construction of intelligent machine's (Intellectual Machinery Arm, Image Guiding Device, Autonomous Car): without a mature computational vision system, artificial intelligence's interactive modes would be largely limited and shallowed. Based on that intention, I wish my futuer research career focused on the integration of learning and vision system for strong and large sacale AI systems applied in various fields.

1.1.1 Study Scopes

Computational Vision is an interdisciplinary field that deals with how to make computers gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do[1–3]. In 19^{th} century, David Marr put forward that computational vision can be interpreted as an information processing system on three complementary levels of analysis (Marr's Tri-Level Hypothesis)[5]: computational level, representational level, and implementational

level. Among these directions, I am espacially interested in the representation and implementation foundations of vision. My research emphasis is the intersection of statistical learning, discriminative learning techniques. I will briefly introduce the deep neural networks, one typical learning method and my understanding to its future in the following chapter.

1.1.2 Deep Neural Networks

Deep Neural Network can be viewed as a computational model consists of multiple processing layers (Hierarchical Structre) to learn representation of data. These methods have dramatically improved the state-of-the-art in visual object recognition, object detection and many other domains. Convolutional neural network (CNN), consists of multiple layers of receptive fields. Each receptive field contains neuron collections that process portions of the input image. To obtain a better representation of the original image, features retrieved from each collection will be tiled, and this repeated for every such layer. CNN have achieved great improvement on both object detection and recognition tasks in recent years because of its extraordinary ability in learning discriminative features of objects with different identities.

1.1.3 Conv Net Based Recognition and Detection

Basically, object detection is a computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as animals, face, or buildings) in digital images or videos. Considering CNN's good performance in feature extraction, it is applied on detection and recognition as a feature based method: to find feasible matches between object features and image features. Traditional feature based method includes Interpretation trees, Pose consistency, Scale-invariant feature transform (SIFT) and so forth. One of the common grounds of the above handcraft feature extraction methods lie in that things can be said analytically or with a gaurantee of performance, but often the conditions are quite limited and do not apply to general situations in the real world (Song-Chun Zhu, 2014). While successful as CNN, there also exits drastic disputes among this domain:

1. Lack of explainability of the deep features in layers.
2. Extreme dependence on mass data samples (Data-driven model).
3. Lack of precise spatial relations and semantic information. (Hinton)

1.1.4 Future Work on Deep Model Based CV

Deep learning, with its popular meaning, is very much like the method practiced by Chinese herbal clinics over the past three thousand years[4]. Though sounds ironically, it does reflect the dilemma that most deep learning engineers are facing: just like ancient people who own little knowledge of medication that they mix different ingredients with weights and boiled to black and bitter soup as drugs, most deep learning engineers would try different model structure, compositions, weights to acquire most performanced deep models. It's true that herbal clinics can cure illness without understanding

either the biologic functions or the mechanism of the drugs. But as for deep neural network, when encountering complicated data distribution in the real word or lack of enough training samples, to try different components and ensemble models is not a cost we can afford, nor is this scientism for us.

Secondly, as we have referred above, according to Marr's Tri-Level Hypothesis, we can understand and construct a vision system on three levels: computational level, representation level and implementation level. While in DNN structure, it is encouraged to build end-to-end networks based on the imitation to mammal's mentally development. All of the three levels are compressed in one end-to-end netwotk, thus posing a negative influence: when the model fails our expectation, it's hard for us to locate the primary cause on a specific stage.

For the future, much scope remains to be improved. In [7] Prof. Leonardis proposed his Deep Compositional Networks, that is a novel analytic model of a basic unit in a layered hierarchical model with both explicit compositional structure and a well-defined discriminative cost function in 2016. Such model takes design inspiration from CNNs' lack of explicit structure in features, and explicit shapes and features can be visualized semantically. In [10], we investigated the possibility to better utilize the imbalanced data in deep model's training for face-recognition problem. New loss function is designed especially for long-tail distributed training samples. Such work is motivated by the intention to relieve the general deep model's highly dependence on data.

In my view, tentative but pioneering works like this are quite indispensable and highly meaningful for the whole academic circle. More than that, as a burgeoning interdiscipline, inspirations may once again be found from neuroscience and cognitive pshycology.

1.1.5 Light From Neuroscience and Cognitive Psychology

The understanding of human vision and computer vision, are strongly interconnected (Shimon Ullman, 2010).

This is a word picked from the book: Vision, A Computational Investigation into the Human Representation and Processing of Visual Information. by David Marr, 1979, who is been viewed as the father of Neuroscience in contemporary era. Convolutional Neural Network is originally inspired by biological network structures in the human brain visual cortex and the vision system's ventral stream[6]. In the computer implementation, the convolution layer's node is linked with several spatial adjacent neurons in the previous layer, just like the optic nerve in human beings. Biologist also states that the eye might provide alone an example of a convolution neural network (as made popular for object classification tasks). The topology of the network is in a sense self-structured but a relatively slow process as the blueprint is a product of evolution and the practical implementation takes place during the development of the organ. Inversely, in [9], this paper investigates how the goal-driven CNN approach can be used to delve deeply into understanding the development and organization of sensory cortical processing. In [8],

Based on a variety of these facts, I realized that the relationship between the computational vision's development and the study of neuroscience is

to interact and stimulate each other. More groundbreaking and innovative work must be proposed based on the inspiration of them.

To develop a robustive and highly intelligent vision system is a challenging peak people need to conquer in the pursuit of truely artificial intelligence. Heuristic may come from Statistical Learning, Neuroscience and any other interdiscipline subjects. I fully understand this is an outward journey, a difficult but enjoyable exploring process into unknown territory. But the ultimate measure of a scientist is never where he stands in moments of comfort and convenience, but where he stands at times of challenge and unknown.

1.2 Convolutional Neural Network

1.2.1 Convolutional and Deconvolutional Layer

1.2.2 Pooling Layer

1.3 Residual Neural Network

1.4 Range Loss for Deep Face Recognition

Chapter 2

Generative Adversarial Network

2.1 Unsupervised Learning in Deep Learning

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

2.2 Generative Adversarial Network

2.2.1 Wasserstein GAN

2.2.2 Conditional GAN

2.2.3 Deep Convolutional GAN

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

2.3 DC-GAN for Image Generation and Face Super Resolution

Chapter 3

Range Loss Based Generative Adversarial Network

3.1 Range Loss and Bound Equilibrium

3.2 RL-GAN for Face Generation

Bibliography

[1] Dana H Ballard and Christopher M Brown. "Computer vision, 1982". In: *Prenice-Hall, Englewood Cliffs, NJ* (1982).

[2] Michael RW Dawson. *Understanding cognitive science*. Blackwell Oxford, 1998.

[3] TS Huang. "Computer Vision: Evolution and Promise". In: *CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH-REPORTS-CERN* (1996), pp. 21–26.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[5] David Marr and Tomaso Poggio. "From understanding computation to understanding neural circuitry". In: (1976).

[6] Josh Morgan and Rachel Wong. "Development of cell types and synaptic connections in the retina". In: (2007).

[7] Domen Tabernik et al. "Towards Deep Compositional Networks". In: *arXiv preprint arXiv:1609.03795* (2016).

[8] Juan Wang et al. "A Multi-Resolution Approach for Spinal Metastasis Detection using Deep Siamese Neural Networks". In: *Computers in Biology and Medicine* (2017).

[9] Daniel LK Yamins and James J DiCarlo. "Using goal-driven deep learning models to understand sensory cortex". In: *Nature neuroscience* 19.3 (2016), pp. 356–365.

[10] Xiao Zhang et al. "Range Loss for Deep Face Recognition with Long-tail". In: *arXiv preprint arXiv:1611.08976* (2016).