UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Defence

Thursday July 28th, 2022 @ 9am via Zoom

Enhancing Learning Capability of Convolutional Neural Networks for Fundamental Vision Problems

Mingjie Wang

Chair: Dr. Joe Sawada
Advisor: Dr. Minglun Gong
Advisory: Dr. Simon Yang [School of Engineering]
Non-Advisory Member: Dr. Stefan Kremer
External Examiner: Dr. Guanghui (Richard) Wang [Toronto Metropolitan University]

ABSTRACT:

During the past decade, Convolutional Neural Networks (CNNs) have been dominating the realm of computer vision and becoming a de facto standard for modern data-driven algorithms. They have exhibited state-of-the-art performances in many vision tasks, thanks to their strong representative capacities. Hence, further enhancing CNNs' learning capabilities has very broad impacts. This thesis by articles presents several such enhancements through novel mechanisms, structures, and paradigms. These enhancements are evaluated using two main types of learning tasks, namely classification and regression.

The thesis starts with the image classification problem since it triggers the surge of various downstream vision tasks. Inspired by the success of feature reuse in DenseNet and a pool of drop-based stochastic regularization, Stochastic Features Reuse is presented to boost capacity and generalization of DenseNet through randomly dropping reused features. Simultaneously, a Multi-scale Convolution Aggregation module is also explored to facilitate learning scale-invariant representations. Albeit promising, the resulting algorithm still inherits DenseNet's limitations on the large model scale and superfluous feature reuse. To extract highly discriminative representations with more compact models, a layer-wise attention condenser, named Adaptively Dense Convolutional Neural Networks, is designed to form a strong variant.

To study the impacts on regression problems, the second part of the thesis focuses on the crowd counting problem since it expects a single, non-constrained value as output, making it more arduous and representative than other regression tasks. Motivated by the ideas of diverse receptive field and stochastic regularization, a Stochastic Multi-Scale Aggregation Network is designed to enrich the scale diversity of feature maps and to combat overfitting. To further boost the capacity of handling drastic scale variations, a Single-column Scale-invariant Network is presented, which extracts sophisticated scale-invariant features via the fabric-like combination of interlayer scale integration and a novel intralayer Scale-invariant Transformation. To further dig into fine-grained group convolutions and effects of multi-task supervision on network's capacity, an innovative Scale Tree Network is presented to parse scale information hierarchically and efficiently incorporating a tree-like structure. It also proposes a Multi-level Auxiliator to facilitate the recognition of cluttered backgrounds. Finally, a weakly-supervised counting framework, referred to as CrowdMLP, is presented to model global-range receptive fields for regression problems, which is characterized by count-level annotations and multi-granularity MLP.

Extensive experiments on widely-used benchmark datasets demonstrate the effectiveness of the proposed strategies or design principles in enhancing learning capabilities for two fundamental vision problems, thereby achieving superior performances in classification and counting accuracy. Ablation studies and visualizations are also performed to shed light on the impacts and behaviours of individual components.