報告番号	※甲　　　第　　　号

主論文の要旨

論文題目　Towards Efficient and Accurate Attention Learning for Fine-grained Image Classification
（詳細画像分類におけるアテンション学習の精度向上と効率化に関する研究）

氏　　名　　劉　迪超

論文内容の要旨

The studies on image classification can be divided into two Fine-grained image classification (FGIC) aims to recognize hard-to-distinguish object classes, such as different breeds of birds or models of cars. It is a very difficult task and capturing attention is key for solving the difficulty. The objective of this work is to explore how to efficiently capture attention information to improve the accuracy of FGIC. With this objective, we propose three novel attention-learning frameworks for FGIC. This paper has six chapters.

Chapter 1 gives the background of this research, discusses the motivation of this thesis as well as gives an overview of the proposed approaches.

Chapter 2 introduces the studies that are related to this research or the topic of fine-grained image classification.

Chapter 3 introduces a guided attention-learning framework, named as Attention-Guided Spatial Transformer Networks (AG-STNs), which focuses on capturing effective attention regions for FGIC. Traditional region-based attention learning approaches treat the localization and recognition of attention regions as two separate steps, during which the errors in each step can be accumulated. AG-STNs localize attention regions by deep neural networks, which can be optimized together with the recognition networks. Learning cropping attention regions is very difficult for deep neural networks, and AG-STNs solve the training difficulty by utilizing hard-coded attention as the guiding signal to initialize the localization network. Moreover, AG-STNs can generate multiple scales of attention regions, a fusion of whose predictions

further improves the accuracy.

Chapter 4 introduces a multi-task attention-learning framework, named Contrastively-reinforced Attention Convolutional Neural Network (CRA-CNN). CRA-CNN is inspired by the human behavior of using the knowledge learned from one task to help learn another related task. During the training, CRA-CNN has two networks. One is the major network used for the task of categorizing the given input image. The other is the subordinate network used for the task to make the deep features of the major network correspond more to the attention regions. After training, the subordinate network can be removed, and only the major network is kept for utilization. In this way, CRA-CNN does not require extra overhead for utilization and has no loss of information from input images.

Chapter 5 introduces a recursively-refined multi-scale attention framework, named Recursive Multi-scale Channel-spatial Attention Module (RMCSAM). Different from the approaches proposed in Chapters 3 and 4, RMCSAM is an insertable module that has small weights and can be embedded into various backbone networks. RMCSAM explores both channel-wise and spatial-wise attention from deep features, and recursively refines the learned attention information for more accurate attention. RMCSAM is lightweight and has strong versatility, and it can be combined with the Progressive Multi-Granularity Training (PMG), which is the state-of-the-art approach in the FGIC task, to further improve the accuracy. RMCSAM is also possible to combine with other training frameworks.

Chapter 6 gives the summary and prospect of this paper.