2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan) | 979-8-3503-8684-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCE-Taiwan62264.2024.10674087

Few Shot Ultra Fine-Grained Visual Classification for Defect Detection

Zheng-Wei Chen, and Chia-Yu Lin
Department of Computer Science and Information Engineering
National Central University, Taoyuan, Taiwan
Corresponding Author: Chia-Yu Lin (sallylin0121@ncu.edu.tw)

***Abstract*—The challenge in defect detection lies in accurately classifying various defect types, each minute in size, and similar defects. Moreover, as defect images are scarce, defect detection becomes a few-shot sample task falling under fine-grained visual classification (FGVC). However, in practical scenarios, discerning defects for model differentiation is complicated by the need to capture subtle features and nuances among different defect classes. Consequently, effective defect detection demands models with robust feature extraction capabilities for few-shot sample tasks to resolve the challenges of FGVC. To address this, we integrate the Self-supervised Knowledge Distillation (SKD) with the few-shot learning model to enhance the feature extraction capabilities of models in handling few-shot FGVC. According to our experiments, our proposed method significantly enhances average accuracy by 5.5%. Hence, this approach enables the rapid training of highly accurate models using limited samples.**

***Index Terms*—Defect detection, fine-grained visual classification, few-shot learning**

I. Introduction

FGVC [1] is one of the challenging issues in image classification, which can generally be divided into three categories: basic-level, fine-grained, and instance-level. The basic-level category can be considered as general image classification, such as classifying cars, airplanes, and birds. Fine-grained classification requires distinguishing finer features, such as classifying species of birds. Instance-level classification not only requires distinguishing finer features but also entails subtle differences within each category, making classification tasks more difficult, for example, classifying different types of computer screens. In Fig. 1, we see the difference between the fine-grained and instance-level categories. Therefore, models may perform well on fine-grained tasks but may need to maintain the same level of performance on instance-level tasks.

Few-shot image classification is also a challenging issue nowadays and is more closely related to real-world scenarios. In few-shot training, models not only need to learn features with limited training data but also need to avoid overfitting.

Defective images are extremely rare in defect detection, leading to a few-shot image classification problem. Besides, defect detection is not just a typical fine-grained problem; it requires capturing even finer features. Additionally, different types of features may be very similar, which could cause the model to confuse and degrade performance, thus belonging to the instance-level category.

In the current field of FGVC, most methods focus on subtle differences in local features within categories while over-looking the importance of global features. High-temperature refinement and background suppression (HERBS) [2], a state-of-the-art method in FGVC, stands out for its unique approach to emphasizing global features. This strategy allows the model to grasp essential contextual information, leading to outstanding performance. Nonetheless, we have found that HERBS's performance at the instance level can be affected.

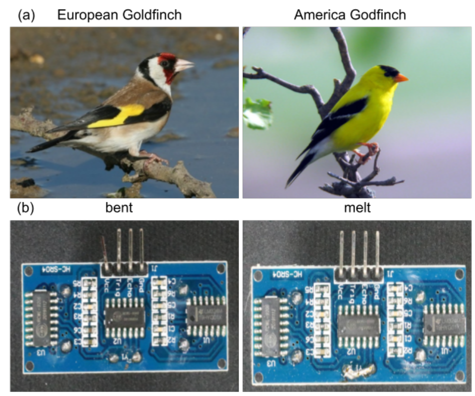

Fig. 1. The difference between the normal fine-grained Visual classification task (a) and defect detection task (b). (a) The left side is the European Goldfinch in the CUB-200-2011 [5], and the right side is the American Goldfinchin the CUB-200-2011. (b) The left side is the defect of the bent in the VisA [4], and the right side is the defect of the melt in the VisA.

In this paper, we aim to develop a model that simultaneously considers both local features and global information. Therefore, we utilize self-supervised knowledge distillation for few-shot learning (SKD) [3] to enhance the model's ability to extract fine-grained features. Subsequently, we employ HERBS for fine-tuning to enable the model to learn global information.

We use the PCB class from the VisA dataset [4] as defect detection images. From the experimental results, it can be seen that our proposed method significantly improves overall accuracy.

II. Methods

A. SKD

SKD [3] is a method used in the field of few-shot learning to enhance feature extraction capabilities. SKD combines augmentation methods and knowledge distillation. Initially, SKD applies rotation augmentation to images to generate training data from various angles. Subsequently, the augmented and original data train a model, denoted as Gen-0. Gen-0 then serves as a teacher to train a new student model, Gen-1, using knowledge distillation. Gen-1 learns from the teacher's knowledge through this process and enhances its feature extraction capabilities. Additionally, SKD utilizes rotated augmentation to improve the model's robustness.

B. HERBS

HERBS [2] is the state-of-the-art model for addressing fine-grained image classification issues. Unlike other methods, it focuses on subtle differences between categories and enhances discriminative features. HERBS emphasizes the relationship between global image features and contextual information, not neglecting the importance of backgrounds.

The HERBS framework consists of two main modules: the high-temperature refinement module and the background suppression module. It adopts the architectures of FPN [6] to emphasize the significance of backgrounds. The high-temperature refinement module adjusts the temperature parameter T for predicted categories. This module increases the temperature T to learn feature maps, capturing more global and contextual information. Lowering the temperature refines the feature maps to capture finer details, thus enhancing feature extraction capabilities.

Subsequently, the background suppression module computes scores for feature maps in each layer. Those with lower scores are considered background and the remaining feature maps are considered foreground. Then, the background suppression module combined the foreground to produce a new set of feature maps deemed essential for training. The purpose is to increase the gap between the background and the foreground. Overall, the background suppression module discards unimportant features while training the significant ones, aiming to enhance feature extraction capabilities.

Compared to HERBS, SKD focuses more on learning detailed features. While HERBS emphasizes global features and the relationship between contextual information, it may overlook finer details. Therefore, we first employ SKD to train the backbone and then utilize HERBS for fine-tuning, enabling the model to consider local features and global information simultaneously.

III. Experiment

In SKD training dataset, we use the office high-shot setting in the VisA dataset [4]. There are 12 categories in total. Each category is divided into normal and anomaly. Therefore, the training dataset is 24 classes.

In HERBS training dataset, we use the PCB dataset from VisA, which has four subsets: PCB1, PCB2, PCB3, and PCB4. We classify the images that were initially labeled as anomaly based on the type of defect, resulting in five categories for each subset of PCB1, PCB2, and PCB3 and eight categories for PCB4. To ensure that there is enough data, we have included at least ten images per class in each training set and at least five images per class in each test set. We have also augmented the data to have 40 images per class for training purposes. Each experiment is repeated at least three times to ensure accuracy.

TABLE I
The precision(%) and recall(%) on the PCB from VisA [4]. (The left side is HERBS, and the right side is HERBS with SKD pre-trained model.)

Method	PCB1	PCB2	PCB3	PCB4	Average
HERBS	86	82.2	**88.3**	80	84.1
SKD+HERBS	**94**	**84.4**	**88.3**	**91.7**	**89.6**

Table I shows the experimental results. Our method has shown better performance than others in every subset of the dataset, resulting in an average accuracy increase of 5.5%. This indicates the effectiveness of our approach in enhancing the performance of defect detection tasks, and it also highlights its potential for fine-grained classification tasks.

IV. Conclusion

Our method utilizes SKD to enhance the model's feature extraction capabilities, leveraging rotation augmentation and knowledge distillation to improve performance. Then, we employs HERBS for fine-tuning, enabling the model to simultaneously consider local features and global information while addressing the challenges of few-shot and fine-grained visual classification. Experiments conducted on VisA dataset demonstrate that HERBS with SKD significantly enhances average accuracy by 5.5%.

Acknowledgements

This work is sponsored by the National Science and Technology Council (NSTC) under the projects NSTC 110-2222-E-008-008-MY3 and NSTC 112-2622-8-A49-021.

References

[1] Xiu-Shen Wei, Yi-Zhe Song, Oisin Mac Aodha, Jianxin Wu, Yuxin Peng, Jinhui Tang, Jian Yang, and Serge Belongie, "Fine-grained image analysis with deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 8927–8948, 2021.

[2] Po-Yung Chou, Yu-Yung Kao, and Cheng-Hung Lin, "Fine-grained visual classification with high-temperature refinement and background suppression," *arXiv preprint arXiv:2303.06442*, 2023.

[3] Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Mubarak Shah, "Self-supervised knowledge distillation for few-shot learning," *arXiv preprint arXiv:2006.09785*, 2020.

[4] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer, "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 392–408.

[5] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie, "The caltech-ucsd birds-200-2011 dataset," in *California Institute of Technology*, 2011.

[6] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.