RECOGNITION of OBJECTS and THEIR DEFECTS

PhD Thesis Booklet

by

Amr Mohamed Abdelhameed Nagy Abdo

Supervisor: Dr. **László Czúni**

Department of Electrical Engineering and Information Systems

Doctoral School of Information Science and Technology

University of Pannonia

Veszprém, Hungary
2022

1 Introduction

In recent years, object recognition, as one of the most fundamental and demanding topics in computer vision, has received great attention. It is considered as one of the most important tasks that brings significant impacts to the society since many applications are built upon object recognition techniques, such as object recognition of traffic signs, shopping applications, human recognition in surveillance, road object recognition in autonomous driving, and daily object recognition in handheld devices and robotics. Unfortunately, 3D object recognition still remains one of the core problems in computer vision. 3D object recognition in case of real-life environments using handheld devices or robots is a difficult task due to changing viewpoints, varying 3D to 2D projections, possible different noises (e.g. motion blur, color distortion), and the limited computational resources and memory. Also, often video-stream acquisition is crucial for many computer vision applications such as unmanned aerial vehicle (UAV) applications, manufacturing industry or video surveillance. The aim of object detection is to develop computational models and techniques that provide one of the most basic pieces of information needed by computer vision applications: What objects do we see and exactly where are they? As one of the fundamental problems of computer vision, object recognition and detection forms the basis of many other computer vision tasks, such as visual inspection [2, 4], object tracking [6], instance segmentation [1], image captioning [5], etc. The next important step after recognizing the objects by machine is to detect anomalies on it. Visual inspection technology allows differentiating anomalies in objects mimicking human visual inspection. While it suggests monitoring with a minimum amount of human activity, applying the same solution to a wide variety of tasks or defect types is challenging. There is a wide field of such applications including traffic signs defects, steel surface defects, solar panels, automated product manufacturing, railway industry, casting or welding, and healthcare. In order to meet industrial expectations, there is a strong need to achieve high performance in automated visual inspection. This work is concerned with research methods that can recognize objects viewed from multiple directions with traditional and deep learning approaches and reliable solutions to recognize different types of defects efficiently.

2 Motivation

Technology required by industry 4.0 in manufacturing processes requires reliable, accurate, and fast object recognition and visual inspection technologies, often relying on continuous monitoring of optical information. Additionally, lightweight methods for object recognition may be very useful for a variety of applications such as drones or wearable computing. Moreover, as automation is wide-spreading in manufacturing processes, the need for automatic anomaly detection is growing. That is, if we trained our machine intelligence for a given task, there is always a non-zero probability that unseen events might happen that the system is not trained to handle. A part of this problem is few-shot learning, where new kinds of errors appear to be classified as soon as possible, typically with a very low number of training samples. We have to carry out incremental learning, since previously trained knowledge should not be forgotten. Moreover, the retraining of the whole architecture would be resource demanding: the large amount of time, memory, and processing power is typically not available on site or in time.

3 Summary of the Theses

This section presents the main original contributions of my PhD research.

Thesis I: Object Recognition Techniques using Deep Neural Networks and HMMs

Object recognition only in the last few years has made a significant improvement with the evolution of neural networks. While convolutional neural networks are very efficient in object recognition there is still need for improvements in many practical cases. For example if the performance from single images is not satisfactory, natural ambiguities (such as noise, occlusion and geometrical distortions) and the requirements of the computational and memory are high. Additionally, to interact with the objects of the environment, not only specific or generic object recognition is inevitable, but the determination of their pose is also essential.

1. **I proposed a framework to show that HMMs can be used to combine the data of CNNs and IMUs for object recognition and pose estimation**. I showed that relative pose changes can give enough information for the HMMs to estimate the most probable state sequences and thus find the most probable object visible on the images and improve the performance. To show the efficiency of the proposed method, we compared the results with VGG16 network.
2. **HMM based knowledge can be improved by integrating it with an active vision approach, the proposed framework is called AV-HMM-CEDD**. I showed that how active perception and information fusion from sequential multiple shots can help the recognition of 3D objects even if the objects are occluded. Experimental results shows that our proposed method can handle much better the untrained occluded queries if we compare it with DNNs. Moreover, the proposed AV-HMM-CEDD framework is computationally lightweight, requires limited memory and can incorporate other classifiers, not only the presented CEDD.

This thesis is explained in detail in Chapter 2 and the related publications are **AM1**, **AM2**, **AM3**.

Thesis II: Detection of Traffic Sign Defects

There is no doubt that traffic signs are very important parts of the road infrastructure for vehicles. However, detection and recognition of traffic signs with high accuracy is still an unsolved problem, especially in real-life conditions. Beside unfavorable weather, lighting, and imaging conditions the unwanted defects of traffic signs may heavily affect the accuracy of such systems. It is inevitable to develop systems to monitor the state of traffic signs, by the detection of the different errors, supporting their maintenance.

1. **I proposed a new siamese neural network architecture to recognize the defects of a large number of classes of traffic signs**. The previously proposed networks were extended by several layers and the original features, besides computing the difference, were retained for fully connected layers. Experimental results show that this approach can be applied to recognize anomalies in images with better performance for well-known feature learning approaches. Additionally, it is possible to use it to recognize defects in untrained types of objects.
2. **To improve the performance, I proposed a new mechanism to combine the confidence values of siamese networks with SVM with the help of support set images**. The advantage of our approach, compared to the concept of ensemble of networks, is that only one network is to be trained and maintained.

This thesis is explained in detail in Chapter 3 and the related publications are **AM4**, **AM5**.

Thesis III: Classification, Zero and Fast Few-Shot Learning of Steel Surface Defects

Steel is the most important metal in terms of quantity and variety of applications in the modern world. Due to the manufacturing process and environmental conditions, steel surfaces can have a variety of defects. In most applications, beside the questions of detection and classification, the following key problems are to be answered: real-time problem, small target problem, small sample problem, unbalanced sample problem. Moreover, in incremental learning, new data (i.e. new shots of previously seen or unseen classes) arrive in phases over time and we have to extend our classification model to include these new classes or new samples of classes.

1. **I proposed a new architecture to combine EfficientNet with randomized networks (EffNet+RC) for classification and few-shot learning, as well as for continuous learning of steel surface defects**. This architecture achieved an effective training time at least an order of magnitude smaller than that of other classifiers. Moreover, it classifies defects with a good performance for a few shots.
2. **To deal with zero shot learning, I proposed a deep architecture of siamese network**. This network is useful for learning features to classify defects or to cluster unseen classes without a single training.

This thesis is explained in details in Chapter 4 and the related publications are **AM6**, **AM7**.

4 Applications of the Scientific Results

The proposed methods in this thesis can be useful for industry 4.0 technologies. Industry 4.0 embodies the growing trend towards automation and data exchange in technology and processes within the manufacturing industry, including the internet of things, smart manufacturing, robotics, machine learning and artificial intelligence.

The results of Thesis I could be applied to the drones era or wearable computing as it presents lightweight solutions for object recognition (small amount of data, low memory requirements). The proposed methods in Thesis II could help to monitor the condition of traffic signs which could be reflected in improved reliability of self-driving cars. For example, McAfee researchers recently easily tricked a Tesla into speeding while the car's intelligent cruise control feature was engaged [3]. They showed how they could get a Tesla to misread a speed limit and accelerate to 85 mph instead of 35 by simple and tiny modifications of a sign. We created a dataset for traffic sign defects and made it available online at the website of our laboratory. The results of Thesis III can be directly applied for the quality assessment of steel sheets and stripes during manufacturing, as we reached approximately 100% accuracy for the classification of their defects.

We used the latest common software technologies to develop our solutions, such as Python, TensorFlow, Keras, and Java, so others can apply easily these solutions in real-life applications.

This work has been partly implemented within the TKP2020-NKA-10, OTKA K 135729, and OTKA K 120369 projects.

5 Publications

Publications of Amr Mohamed Nagy Abdo are listed below.

5.1 Publications Related to this Thesis

AM1. Czúni László and **Amr M. Nagy**: Improving object recognition of CNNs with multiple queries and HMMs. In Twelfth International Conference on Machine Vision (ICMV), pages 1143310, 2020.

AM2. Czúni László and **Amr M. Nagy**: Hidden Markov models for pose estimation. In 15th International Conference on Computer Vision Theory and Applications (VISAPP), pages 1143310, 2020.

AM3. **Amr M. Nagy**, Metwally Rashad and Czúni László: Active multiview recognition with hidden Markov temporal support. In Signal, Image and Video Processing, pages 315–322, 2021 (**2020-IF: 2.157**).

AM4. Amr M. Nagy and Czúni László: Detecting object defects with fusioning convolutional siamese neural networks. In 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), pages 157–163, 2021.

AM5. Amr M. Nagy and Czúni László: Deep neural network models for the recognition of traffic signs defects. In 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pages 725-729, 2021.

AM6. Amr M. Nagy and Czúni László: Classification and fast few-shot learning of steel surface defects with randomized network. Applied Sciences; 12(8):3967, 2022, (**2020-IF: 2.67**).

AM7. Amr M. Nagy and Czúni László: Zero-shot learning and classification of steel surface defects. In 14th International Conference on Machine Vision (ICMV), pages 386 - 394, 2021.

5.2 Publications not Related to this Thesis

A1. Alejandro R. Rodriguez, **Amr M. Nagy**, Zsolt Vörösházi, György Bereczky, László Czúni: Segmentation and error detection of PV modules, In 27th International Conference on Emerging Technologies and Factory Automation, 2022.

A2. Amr M. Nagy, Ali Ahmed and Hala Helmy Zayed: Hybrid iterated Kalman particle filter for object tracking problems. In 8th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), pages 375–381, 2013.

A3. Amr M. Nagy, Ali Ahmed and Hala Helmy Zayed: Particle filter based on joint color texture histogram for object tracking. In International Image Processing, Applications and Systems Conference, pages 1–6, 2014.

5.3 Other Presentations

P1. Czúni László, **Amr M. Nagy**, M. Rashad, "About the temporal support of active object recognition", Pannonian Conference on Advances in Information Technology (PCIT 2020)", Veszprem, Hungary.

P2. Czúni László, **Amr M. Nagy**, M. Rashad, "Low complexity 3D object recognition for mobile devices based on Markovian framework (HMM)", in: 12th Conference of Hungarian Association for Image Processing and Pattern Recognition (KÉPAF-2019), Debrecen, Hungary.

P3. Czúni László, **Amr M. Nagy**, M. Rashad, "Temporal models for 3D object recognition", in: World Congress on Artificial Intelligence and Machine Learning, (WCAIML-2019), Spain.

Bibliography

[1] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3150–3158, 2016.

[2] Nazrul Ismail and Owais A Malik. Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*, 9(1):24–37, 2022.

[3] Trivedi S. Povolny, S. Model hacking ADAS to save safer roads for autonomous vehicles, February 2020. [Online; posted 4-May-2022].

[4] Andrew DH Thomas, Michael G Rodd, John D Holt, and CJ Neill. Real-time industrial visual inspection: A review. *Real-Time Imaging*, 1(2):139–158, 1995.

[5] Qi Wu, Chunhua Shen, Peng Wang, Anthony Dick, and Anton Van Den Hengel. Image captioning and visual question answering based on attributes and external knowledge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1367–1381, 2017.

[6] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *ArXiv Preprint ArXiv:1905.05055*, 2019.