Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 169 (2020) 164–167

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Using convolutional neural networks for recognition of objects varied in appearance in computer vision for intellectual robots

Sergey Kulik[a], Alexander Shtanko[a,*]

[a]*National Research Nuclear University MEPHI, Kashirskoe shosse 31, Moscow 115409, Russia*

Abstract

The paper describes an effort to train a convolutional neural network capable of reliably recognizing complex objects that are highly varied in their shapes and appearances in images. Neural networks show very good results on objects that have constant appearances but may have trouble recognizing abstract objects that appear in different shapes, art-styles and lack solid structure, for example, national flags. In an image, a flag may appear waving on a pole, as an element of clothes, in a form of stickers, etc. Due to these differences in appearance computer vision systems may show unsatisfactory results on these types of objects. However, detecting such objects is a necessary task in computer vision, especially for intelligent robots in order to understand the environment. The aim of the research is to apply convolutional neural networks for the detection of flags. In this research, we prepared training and testing sets of objects, trained a neural network for detection task, conducted testing experiments and measured the neural net's performance. These results can be applied in cognitive and robotics technologies as well as general computer vision tasks.

Keywords: Machine learning; Object detection; Convolutional neural networks.

* Corresponding author. Tel.: +7-915-640-3476.
E-mail address: shtanko-mephi@yandex.ru

1877-0509 © 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures.
10.1016/j.procs.2020.02.129

1. Introduction

Intelligent image processing is an important topic in many application fields of the real world. It's used in robots to understand the environment around them, in automatic quality assessment systems to control for defects in production and so on.

The demand for intelligent processing systems is increasing. As such, for example, agent-based systems have been developed [1, 2] to reduce a large amount of loosely connected information and consolidate it into well-structured data [3]. Risk assessment can be a complicated task so intelligent evaluation methods are required [4, 5]. Intelligent systems will also be useful in personnel competencies analysis [6] and, of course, any predictive analytics [7].

For the past five years deep neural networks, especially convolutional neural networks, became a powerful tool of pattern recognition that allowed various predictive systems to be developed, for example, cracks detection [8], earthquake location [9], military mines detection and classification [10, 11].

Object detection is one of the oldest tasks in computer vision and there has been a lot of research done on it. It's closely related to object tracking which has also been heavily researched [12]. Usually, most objects of interest have a more or less consistent appearance like cars or traffic signs. However, there are informational objects that can vary in appearance and have a differently stylized realization. For example, signs, emblems, flags and so on may convey crucial data about the scene and its contents. These objects pose additional difficulty because small details can change their meaning completely. Because of the nature of these types of objects computer vision systems may have trouble detecting them.

The first real-time object detection method was breakthrough research by Viola and Jones [13]. Their algorithm could detect human faces with a fixed orientation. Modern approaches took advantage of advances in neural networks and now can learn to detect various objects. Generally, there're two approaches to solving object detection task: multi-stage (R-CNN [14]) and one-stage (SSD [15], YOLO [16]) detectors.

In this work, we're going to use YOLO detection system. YOLO [16] (later improved to its second [17] and third [18] version) is a detection system based on convolutional neural networks, developed by J. Redmon.

The target of our research is specifically robots. Human-like robots became a fascinating topic, significant efforts have been put into designing necessary movement algorithms, scene mapping and imitating human-like cognitive functions [19, 20].

In the research, we use YOLO to detect objects varied in appearance, specifically the flag of China, in photographs and other images. We collected and annotated images to make two datasets. One – with images of the Chinese flag and second one – with a more typical target for detection – animal faces, so the results can be compared. We trained separate neural networks to detect these objects, measured their performance and analyzed results.

2. Typical system structure

Since the target of this research is robots it's worth to outline the generalized structure of such a system. The main components of the recognition subsystem are presented in Fig. 1.

A robot exists in an outside environment, receives signals from it through sensors and interacts with it using actuation devices. The outside environment contains various objects. To adequately react to environment changes a robot must recognize these objects, their condition, position, and orientation. As such, a typical robot would have an image input sensor to receive visual information. Data from the sensor may go through some preprocessing and is inputted into the recognition unit. The recognition unit then processes image frames using a convolutional neural network and outputs the results into the decision-making unit. This unit based on visual and other data would determine appropriate response and make a request to the actuation control unit to make necessary actions using available actuation devices.

The target of this research is the convolutional neural network responsible for detection objects on images that come from the outside environment.

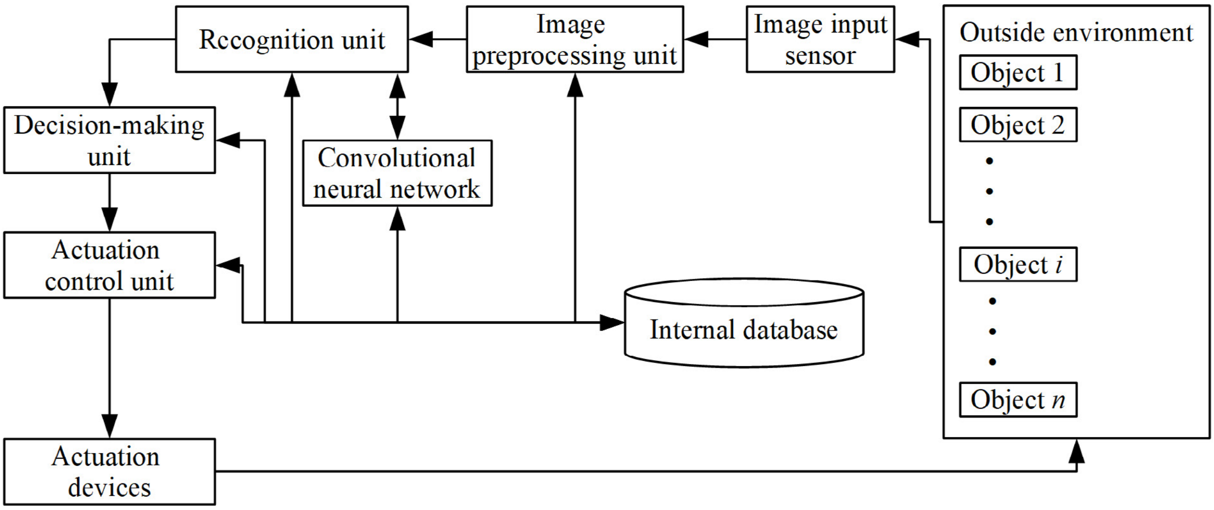

Fig. 1. The general structure of a robot.

3. Experiments and results

In order to analyze the neural network's performance, two datasets have been composed. In the first set the target objects are typical for detection tasks – animal faces. In the second set, the target object is the Chinese flag which is an example of an object varied in appearance. This dataset contains not only simple images with the flag spread out in full view but also their folded versions and stylized artwork. Both datasets contain around 1000 images with 80% selected for training, 10% for validation and 10% for testing.

For training, the tiny version of YOLOv3 was used. Both networks trained for 3000 steps. The training process is presented in Fig. 2.

As can be seen from the graph neural networks can reach very good results (90% mAP and higher) for objects like faces even on small datasets. Results for the flag are significantly lower (around 80% mAP).

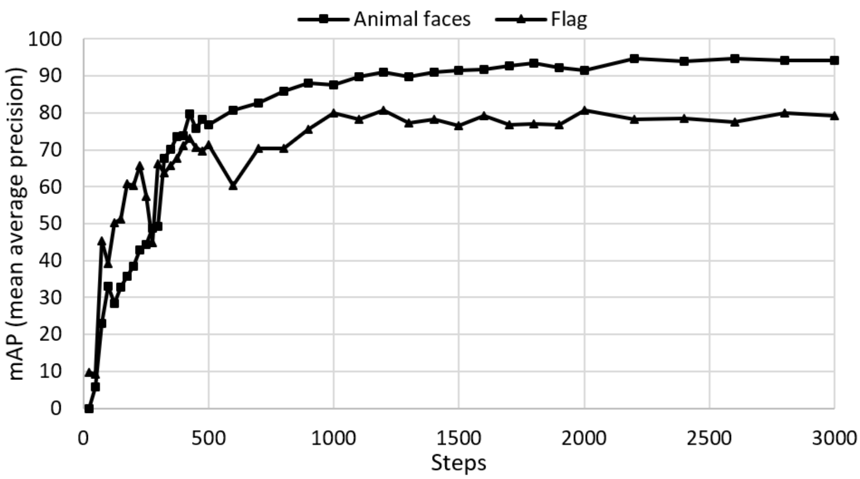

Fig. 2. Mean average precision on validation dataset depending on steps trained.

4. Conclusion

We analyzed how well convolutional neural networks can detect objects varied in appearance (emblems, symbols and so on) for intelligent robots. The generalized structure of a robot that uses visual sensors to understand its environment was presented. We composed training, validation and testing datasets, trained networks and measured their performance. As expected, these types of objects proved to be more difficult to detect than typical objects for detection. We will continue the research in the future.

References

[1] Artamonov, Alexey, Boris Onykiy, Anastasia Ananieva, Kristina Ionkina, Dmitry Kshnyakov, Valeriya Danilova, and Mikhail Korotkov. (2016) "Regular Agent Technologies for the Formation of Dynamic Profile." *Procedia Computer Science* **88**: 482-486.

[2] Artamonov, Aleksey A., Kristina V. Ionkina, Alexandr V. Kirichenko, Ekaterina O. Lopatina, Evheniy S. Tretyakov, and Andrey I. Cherkasskiy. (2018) "Agent-based search in social networks" *International Journal of Civil Engineering and Technology* **9 (13)**: 28-35.

[3] Artamonov, Alexey A., Anastasia G. Ananieva, Evheniy S. Tretyakov, Dmitry O. Kshnyakov, Boris N. Onykiy, and Larisa V. Pronicheva. (2016) "A three-tier model for structuring of scientific and technical information." *Journal of Digital Information Management* **14 (3)**:184-193.

[4] Senkov, A.V., A.S. Zaharov, and V.V. Borisov. (2016) "Accident Risks Assessment by Temporal Fuzzy Bayesian Network." *International Journal of Applied Engineering Research* **11 (22)**: 10731-10736.

[5] Senkov, Aleksey, and Vadim Borisov. (2016) "Risk assessment in fuzzy business processes based on high level fuzzy petri net." *International Journal of Applied Engineering Research* **11 (16)**: 9052-9057.

[6] Kireev, V., A. Silenko, and A. Guseva. (2017) "Cognitive competence of graduates, oriented to work in the knowledge management system in the state corporation "ROSATOM"." *Journal of Physics: Conference Series* **781 (1)**: 012060.

[7] Kireev, Vasiliy S., Anna I. Guseva, Pyotr V. Bochkaryov, Igor A. Kuznetsov, and Stanislav A. Filippov. (2018) "Association Rules Mining for Predictive Analytics in IoT Cloud System." *Advances in Intelligent Systems and Computing* **848**: 107-112.

[8] Cha, Young-Jin, Wooram Choi, and Oral Büyüköztürk. (2017) "Deep learning-based crack damage detection using convolutional neural networks." *Computer-Aided Civil and Infrastructure Engineering* **32(5)**: 361-378.

[9] Perol, Thibaut, Michaël Gharbi, and Marine Denolle. (2018) "Convolutional neural network for earthquake detection and location." *Science Advances* **4(2)**: e1700578.

[10] Kafedziski, Venceslav, Sinisha Pecov, and Dimitar Tanevski. (2018) "Detection and Classification of Land Mines from Ground Penetrating Radar Data Using Faster R-CNN" *26th Telecommunications Forum (TELFOR)*: 1-4.

[11] Besaw, Lance E., Philip J. Stimac. (2015) "Deep convolutional neural networks for classifying GPR B-scans." *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX (Proc. of SPIE)* **9454**

[12] Borisov, Vadim V., and Oleg I. Garanin. (2018) "A Method of Dynamic Visual Scene Analysis Based on Convolutional Neural Network." *Russian Conference on Artificial Intelligence, Communications in Computer and Information Science* **934**: 60–69.

[13] Viola, Paul, and Michael Jones. (2001) "Rapid object detection using a boosted cascade of simple features." *CVPR (1)* **1**: 511-518.

[14] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. (2014) "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*: 580-587.

[15] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. (2016) "Ssd: Single shot multibox detector." *European conference on computer vision*: 21-37.

[16] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. (2016) "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*: 779-788.

[17] Redmon, Joseph, and Ali Farhadi. (2017) "YOLO9000: better, faster, stronger." *Proceedings of the IEEE conference on computer vision and pattern recognition*: 7263-7271.

[18] Redmon, Joseph, and Ali Farhadi. (2018) "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767*

[19] Samsonovich, Alexei V. (2013) "Emotional biologically inspired cognitive architecture." *Biologically Inspired Cognitive Architectures* **6**: 109-125.

[20] Samsonovich, Alexei V. (2018) "On semantic map as a key component in socially-emotional BICA." *Biologically Inspired Cognitive Architectures* **23**: 1–6.