VISION-BASED ROBOTIC ARM CONTROL WITH INCREMENTAL DEEP LEARNING FOR PICK-AND-PLACE TASKS

Mariusz Kubanek [1], ***Łukasz Kuczyński*** [2]

Department of Computer Science, Czestochowa University of Technology,
Czestochowa, Poland
[1]*mariusz.kubanek@pcz.pl*, [2]*lukasz.kuczynski@pcz.pl*

Keywords: **deep learning, robotics, pick-and-place, visual servoing, continual learning**

The rapid development of industrial and service robotics places increasingly high demands on manipulator control systems. Classical methods based on inverse kinematics and PID controllers ensure movement repeatability, but are sensitive to calibration errors, environmental variability, and unpredictable properties of grasped objects [1][2]. This paper proposes a concept of a robotic arm control system based on incremental deep learning supported by camera-based visual feedback, capable of gradually improving the precision of pick-and-place tasks without the need to retrain the model from scratch [3].

The proposed system consists of three main modules. The first is a visual module based on a lightweight convolutional neural network (CNN), whose task is to localize the target object in the image space and estimate its orientation. Based on a single image frame, the coordinates of the object's centroid are determined and then transformed into the robot's workspace using camera–robot calibration [5]. The second module is responsible for trajectory planning and generating control signals for the servo mechanisms of the Arduino-based robotic arm. The third module implements an incremental learning mechanism — after each grasping attempt, the system measures the positioning error and updates the neural network weights using an online learning algorithm with constrained memory (Elastic Weight Consolidation, EWC), which prevents the phenomenon of catastrophic forgetting [3].

A key feature of the proposed approach is the closed-loop visual feedback loop. After each movement, the camera captures the result of the operation: the system evaluates whether the object was correctly grasped and placed at the target position. The error information is encoded as a reward signal and passed to the learning module. In subsequent iterations, the neural network generates increasingly precise control commands, adapting to the specific set of objects and lighting [5].

The neural network architecture is intentionally simplified to enable operation in a computationally constrained environment. The CNN model for object detection is based on a structure similar to MobileNetV2, whose output feeds a two-layer fully connected network responsible for target coordinate regression [6]. The network parameters are stored and updated on a PC acting as the master unit, while the Arduino is responsible solely for executing the generated PWM sequences for the servo motors. This distributed architecture allows for low hardware costs while maintaining full machine learning functionality [4].

The application of the EWC mechanism ensures that incremental learning on new objects does not degrade the grasping performance on previously learned ones. According to the current state of knowledge, regularization-based methods such as EWC represent an effective solution to the stability–plasticity dilemma in continual learning systems [3]. Theoretical analysis of the learning convergence indicates that the positioning error decreases with the number of iterations, provided that sufficient diversity of training data is available. The system is scalable — it can be extended with additional degrees of freedom of the arm or with stereo vision to improve scene depth estimation. The conceptual diagram is shown in Fig. 1.

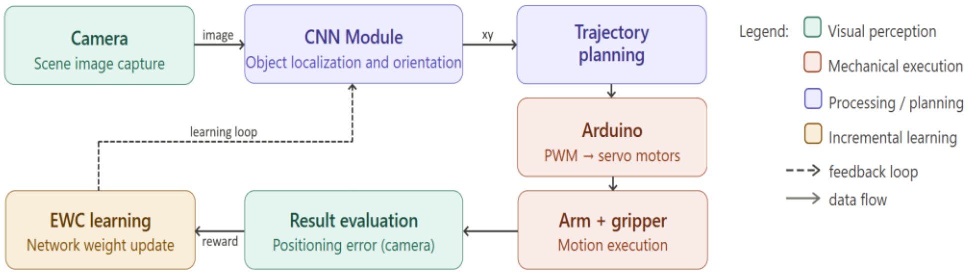

Fig. 1. Block diagram of the proposed system.

The proposed solution opens new possibilities in designing low-cost, adaptive pick-and-place systems for small and medium enterprises, robotics education, and domestic applications. The combination of a lightweight convolutional neural network, incremental learning with elastic memory, and visual feedback loop is an example of the effective application of deep learning in low-budget robotics, without the need for expensive force sensors or precise mechanical calibration [1][2].

References

[1] Imtiaz M.B., Qiao Y., Prehensile and Non-Prehensile Robotic Pick-and-Place of Objects in Clutter Using Deep Reinforcement Learning. Sensors 23(3), 1513, 2023.

[2] Duan J., Zhuang L., Zhang Q., Qin J., Zhou Y., Vision-based robotic grasping using faster R-CNN–GRCNN dual-layer detection mechanism. Proc. IMechE, Part B: J. Engineering Manufacture, 2024.

[3] Van de Ven G.M., Continual Learning and Catastrophic Forgetting. arXiv:2403.05175, 2024.

[4] Szymoniak S., Depta F., Karbowiak Ł., Kubanek M., Trustworthy Artificial Intelligence Methods for Users' Physical and Environmental Security: A Comprehensive Review. Applied Sciences 13, 12068, 2023.

[5] Carrasco V., Martínez J.L., Real-Time Deep Learning Approach to Visual Servo Control and Grasp Detection for Autonomous Robotic Manipulation. Robot. Auton. Syst. 139, 103745, 2021.

[6] Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C., MobileNetV2: Inverted Residuals and Linear Bottlenecks. CVPR, 4510–4520, 2018.