Assisting Upper Limb Prosthesis with a Computer Vision System for Material Detection

Dezmun Roper-Bryant
Department of Electrical and Computer Engineering
North Carolina State University
dgroperb@ncsu.edu

Joseph Berman
Department of Electrical and Computer Engineering
North Carolina State University
jmberman@ncsu.edu

He (Helen) Huang
Joint Department of Biomedical Engineering
North Carolina State University and University of North Carolina at Chapel Hill
hhuang11@ncsu.edu

2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782550

***Abstract*—The concept of utilizing computer vision to aid in the control of prosthetic limbs has been explored in recent studies. However, many computer vision systems are challenged by noisy and cluttered backgrounds in realistic scenarios. In this study, we investigated the feasibility of using a state-of-the-art object detection model (YOLOv8) to identify objects grasped by a myoelectric prosthetic hand. A custom model was trained to recognize various types of cups and was integrated with an EMG-based prosthesis controller. The classification accuracy of the model during an object grasping task with a human subject was approximately 93%. Predictions of grasped objects will be used in future work to adjust the parameters of a grip force controller for a prosthetic hand based on the detected object's properties.**

I. Introduction

The human hand has characteristically high levels of dexterity, allowing stable grasping of various types of objects to be achieved. In the planning phase of object grasping, visual perception of the properties of an object (e.g., shape, size, orientation, density, etc.) allows a person to optimally preshape their hand [1] and apply an anticipatory initial grip force [2]. Following a successful initial grasp, mechanoreceptors in the skin provide feedback of information such as pressure and slip [3], which enable continuous corrective adjustments in grip force. However, these capabilities are disrupted when an upper limb amputation takes place.

Several advancements in this field have been made toward reliable and intuitive prosthetic devices. Conventional robotic upper-limb prostheses are controlled by activating motors in the device proportional to electromyography (EMG) signals produced by residual muscles [4]. EMG-based pattern recognition, focused on decoding discrete movements such as hand open/close gestures [5, 6], represents the current state-of-the-art control of these devices. Furthermore, machine learning [7, 8] and neuromusculoskeletal modeling [9, 10] have been explored in recent studies for decoding intended joint kinematics. However, the latency and estimation errors of EMG-decoding algorithms make fine force control for object grasping tasks challenging.

In recent years, the concept of applying autonomous sensing and control to enhance myoelectric control has gained prominence. The integration of computer vision into upper limb prosthetics seeks to provide prosthetic arms with a sense of perception and enable visual context-based decision-making. Some past studies have incorporated cameras in prosthesis control and implemented machine learning classification models to predict grasp patterns from images of targeted objects [11-14]. The prediction of grasp patterns can be directly applied to prosthesis control schemes to reduce the physical and cognitive efforts of planning for object grasping. Furthermore, it may be possible to estimate additional information from the predicted object type such as surface material, object weight, and overall fragility. Although this concept is promising for improving robotic grasping capabilities, a significant challenge of many computer vision models is background clutter and noise. For example, a convolutional neural network designed to predict a single object class from each image may have unexpected behavior when multiple object types are simultaneously in view of the camera.

In this study, we investigated the capabilities of a modern object detection model for predicting intended grasping targets for an upper-limb prosthesis. The model is capable of detecting multiple objects in a single image and output the predicted class and location in the image of each. We first trained the model to detect various types of common cups and implemented a simple strategy for selecting the most probable target object class at the time of grasping. Next, we integrated the model with an EMG-based prosthesis controller in a complete grasped object classification system, which was implemented by mounting a webcam on a prosthetic hand. Finally, we evaluated the capability of the system in predicting the types of grasped cups with experiments with a human subject. This system will be used in our future work to select optimal parameters for a grip force controller based on predicted object material and fragility.

This work was supported by NSF under Grant 2221479 and by NIH under Grant F31HD111290.

II. Methods

A. Object Detection Model

Various object recognition systems have been developed in recent years. This study uses the You Only Look Once (YOLO) model, which is capable of object detection and classification. This model was chosen because of its ability to output classifications of multiple objects in a single image, apply bounding boxes to each detected object, and provide a confidence score for each detected object at a speed sufficient for real-time video processing. The main details of the model implementation can be found in the original publication [15]. In this study, we used YOLOv8 released by Ultralytics [16].

B. Dataset

To implement the proposed system, we focused on a foundational set of everyday objects: cups. The chosen dataset consisted of plastic cups, paper cups, foam cups, metal cans, and water bottles. The image data used to train the model were collected from two sources: 1662 images were gathered through manual capture (1188) and web scraping from Google (474). This dataset was duplicated and random adjustments, including gamma correction (with gamma values of 0.3 or 3) and random tinting, were applied to enhance dataset robustness, minimizing color reliance. The total dataset, post-augmentation, totaled to 3324 images. The dataset was split into an 80/20 ratio for training (2724 images) and validation (600 images). Annotation focused on the original dataset, utilizing Intel's Computer Vision Annotation Tool for manual annotation. The augmented dataset annotations were auto-generated preserving the original bounding box locations.

The YOLOv8m model was implemented using Ultralytics YOLOv8.0.199, Python 3.11.3, and torch 2.1.2 with CUDA support on an NVIDIA RTX 3070 Laptop with 8092MiB VRAM. The training process spanned 200 epochs with a batch size of 16, taking 5.33 hours to completion.

C. Selection of Target Object

As a new frame was loaded into the model for object detection, each prediction underwent thresholding at a confidence level of 0.9. For predictions over the threshold, the object's bounding boxes and corresponding centers were also created. For the prediction with the center of its bounding box was closest to the center of the image, a counter for that object type would be incremented. Once a material counter reached 15 frames (0.5 seconds), the computer would set the predicted target object to that object type and reset counters for all other objects to zero. If no object was detected for 60 frames (2 seconds) all counters would reset and the predicted target object would be set as "unknown."

D. Integration with Myoelectric Prosthetic

1) Hardware and Setup: To demonstrate how the object detection model can be used to predict the type of object last grasped by a robotic hand, we utilized a 1 degree of freedom prosthetic gripper (Ottobock, Germany). A miniature load cell (LLB130, FUTEK Advanced Sensor Technology, Inc., U.S.) was mounted to the thumb of the hand with a 3D-printed enclosure, and was used to detect object contact signify the start of object grasping and regulate constant grip force on grasped objects. A webcam (C920, Logitech) was mounted on the underside of the wrist and streamed images in at 30 frames per second to be used as input to the object detection model. The webcam was connected to a laptop running the object detection model via USB. An adapter was used to mount the prosthesis to the upper limb of a subject. This setup can be seen in Fig. 1.

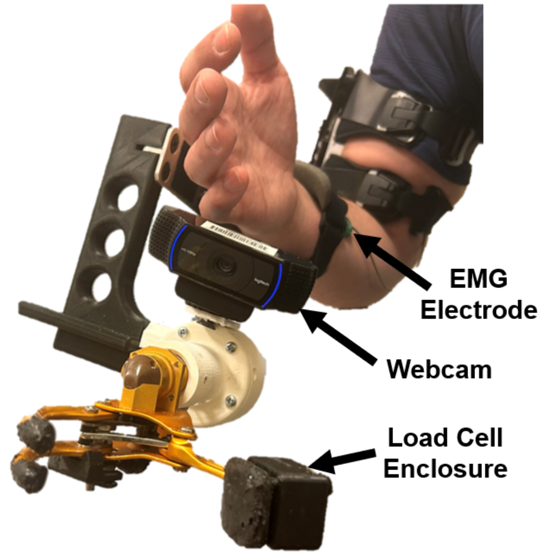

Fig. 1. The hardware used for live testing

2) Integration of Object Detection with Prosthesis Control: To allow a human subject to control the prosthetic hand for an object grasping task, a simple "on/off" myoelectric control scheme was implemented. In this controller, two dry bipolar surface electrodes are placed on the forearm over muscles that activate during the flexion and extension of the fingers respectively. Raw EMG signals are recorded at 1000Hz on an Arduino Due. Muscle activations are then estimated by calculating the mean absolute value of EMG signals using a sliding window with a length of 200ms and increment of 10ms and normalizing by the maximum voluntary contractions (MVC), producing values in the range [0 1]. Three possible control commands (open hand, close hand, and no movement) were selected from using the following conditions. If the flexion muscle activation was simultaneously greater than the extension muscle activation and 25% of the MVC, the hand was closed at a constant speed. If the extension muscle activation was simultaneously greater than the flexion muscle activation and 25% of the MVC, the hand was opened at a constant speed. Otherwise, the hand remained stationary.

The Arduino Due simultaneously sampled the values output from the load cell at 1000Hz. Contact with an object, and thus the start of object grasping, was detected by a measured force on the load cell exceeding a threshold of 1N. In the timestep that object contact was initially detected, the Arduino Due sent

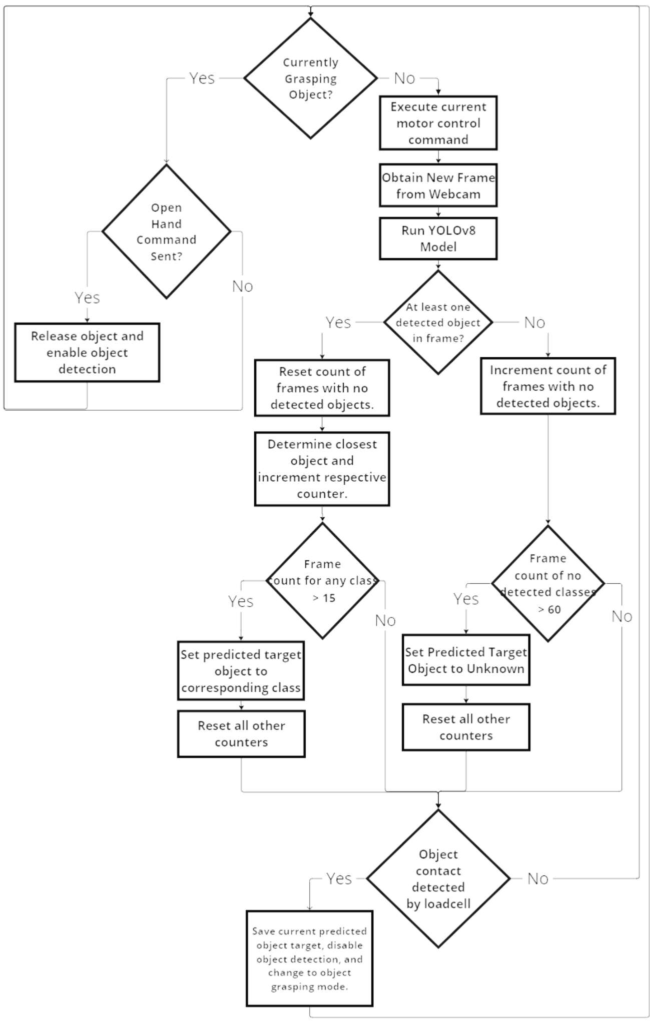

Fig. 2. Flowchart highlighting the controller used.

Fig. 3. Setup of the experiment. Cups used for the grasping task were placed on the black table. The number and location of background objects behind the table were randomized in each trial.

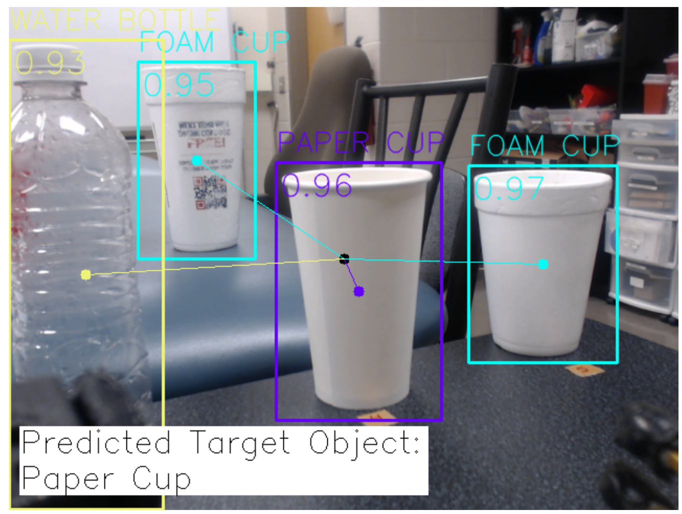

Fig. 4. Example frame of what is captured by the webcam during the test. The numbers under the label indicate the confidence level of that prediction. The lines represent the distance to each detected object from the center of the frame. The predicted target object is paper cup.

a signal to the laptop, which triggered the currently predicted target object to be saved and the prosthesis controller to switch to grasping mode in which constant force was applied by the hand. The object detection model was temporarily disabled during grasping mode. The prosthesis switched back to myoelectric control and re-enabled the object detection model when an "open hand" command was selected by the subject via EMG signal modulation. A flow chart outlining the control procedure can be seen in Fig. 2.

E. Human Subject Experiment

Our system was tested with one human subject (male, age 27, right-hand dominant). The experimental protocol was approved by the Institutional Review Board at North Carolina State University and informed consent was obtained from the subject for their participation. At the start of the experiment, the two electrodes were placed on the forearm over an agonist/antagonist muscle pair located using muscle palpation, and EMG signals were recorded during the MVC for normalization. The prosthesis was then mounted to the subject's upper limb using the adapter.

The subject then performed five trials of an object grasping task. Before each trial, five random cups (one from each of the previously described classes) were placed on a table in front of the subject. Additionally, to create a cluttered background to test the robustness of the object detection system, other cups were placed in the background behind the table. The number, type, and location of cups in the background were randomized before each trial. During trials, the subject was asked to briefly grasp and release cups with the prosthesis. For each trial, each cup was grasped three times for a total of 15 grasps per trial and 75 grasps total for all five trials. The order in which cups were grasped was randomized and read to the subject by a researcher during each trial. The setup of the experiment is shown in Fig. 3. The predictions of grasped cups made by the object detection model were saved and compared to the actual grasped cups to quantify the performance of the system.

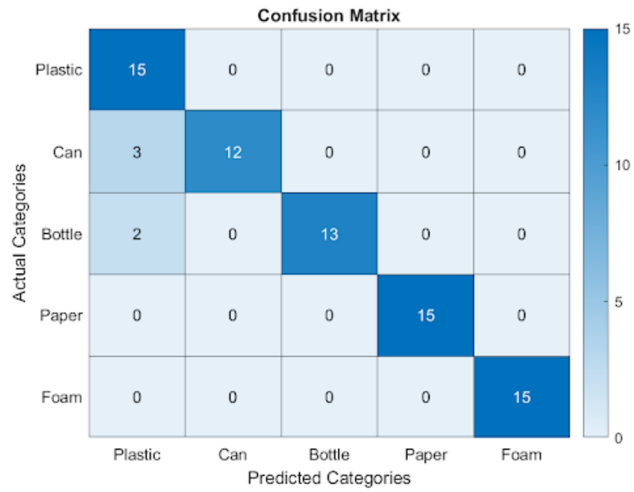

Fig. 5. Confusion matrix showing the predictions made during the cup grasping task.

III. Results/Discussion

An example frame captured by the webcam with predictions made by the object detection model is shown in Fig. 4. The custom YOLOv8 model trained on our cup dataset was able to accurately detect the types and locations of cups in a realistically cluttered environment. Notably, it can be seen that the model was capable of distinguishing between foam and paper cups of similar shapes and colors. This is due in part to the data augmentation procedures described previously. Additionally, while both the actual targeted cup and cups placed in the background are accurately detected, the bounding boxes produced by the object detection model enable a simple distance-based method of method of selecting correct predictions.

The ability of the system to predict the types of grasped cups across all trials is shown by the confusion matrix in Fig. 5. The overall real-time classification accuracy was approximately 93%. This result is comparable to the classification accuracy reported by a related study which used a convolutional neural network to predict one of four possible grasping patterns for a myoelectric hand from webcam images of household objects [11].

This study was intended as a proof of concept for a computer vision system that can be used in myoelectric prostheses. The limitations of the methods presented in this paper include the low number of cup types that our custom object detection model was trained on. However, the original paper introducing YOLO included 20 classes [15]. Thus it is likely that it will be possible to implement a model capable of detecting a larger number of types of grasped objects going beyond the cups included in this study, but will require a much larger image data collection effort.

Conclusion

In this study, we evaluated the feasibility of a state-of-the-art computer vision object detection model for predicting the object grasped by a myoelectric prosthetic hand. In our future work, we intend to use the predictions of currently grasped objects in the grip force controller for the myoelectric hand. One possible approach is setting an anticipatory initial grip force based on the material of the object detected. This robust approach signifies a jump toward seamless human-machine interaction, making prosthetics more efficient, intuitive, and versatile.

References

[1] L. F. Schettino, S. V. Adamovich, and H. Poizner, "Effects of object shape and visual feedback on hand configuration during grasping," (in eng), *Exp Brain Res,* vol. 151, no. 2, pp. 158-66, Jul 2003.

[2] C. Crajé, M. Santello, and A. M. Gordon, "Effects of visual cues of object density on perception and anticipatory control of dexterous manipulation," (in eng), *PLoS One,* vol. 8, no. 10, p. e76855, 2013.

[3] Y. Roudaut, A. Lonigro, B. Coste, J. Hao, P. Delmas, and M. Crest, "Touch sense: functional organization and molecular determinants of mechanosensitive receptors," (in eng), *Channels (Austin),* vol. 6, no. 4, pp. 234-45, Jul-Aug 2012.

[4] A. Fougner, O. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: terminology and proportional myoelectric control-a review," (in eng), *IEEE Trans Neural Syst Rehabil Eng,* vol. 20, no. 5, pp. 663-77, Sep 2012.

[5] L. Resnik, H. H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, "Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control," (in eng), *J Neuroeng Rehabil,* vol. 15, no. 1, p. 23, Mar 15 2018.

[6] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering,* vol. 50, no. 7, pp. 848-854, 2003.

[7] J. M. Hahne *et al.*, "Linear and Nonlinear Regression Techniques for Simultaneous and Proportional Myoelectric Control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 2, pp. 269-279, 2014.

[8] Q. Zhang, R. Liu, W. Chen, and C. Xiong, "Simultaneous and Continuous Estimation of Shoulder and Elbow Kinematics from Surface EMG Signals," (in eng), *Front Neurosci,* vol. 11, p. 280, 2017 .

[9] M. Sartori, G. Durandau, S. Došen, and D. Farina, "Robust simultaneous myoelectric control of multiple degrees of freedom in wrist-hand prostheses by real-time neuromusculoskeletal modeling," (in eng), *J Neural Eng,* vol. 15, no. 6, p. 066026, Dec 2018.

[10] D. L. Crouch and H. Huang, "Lumped-parameter electromyogram-driven musculoskeletal hand model: A potential platform for real-time prosthesis control," *Journal of Biomechanics,* vol. 49, no. 16, pp. 3901-3907, 2016.

[11] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of Neural Engineering,* vol. 14, no. 3, p. 036025, 2017/05/03 2017.

[12] J. DeGol, A. Akhtar, B. Manja, and T. Bretl, "Automatic grasp selection using a camera in a hand prosthesis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 16-20 Aug. 2016 2016, pp. 431-434, doi: 10.1109/EMBC.2016.7590732.

[13] D. T. G. Andrade, A. Ishikawa, A. D. Munoz, and E. Rohmer, "A hybrid approach for the actuation of upper limb prostheses based on computer vision," in *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*, 8-11 Nov. 2017 2017, pp. 1-6, doi: 10.1109/SBR-LARS-R.2017.8215308.

[14] C. Shi, D. Yang, J. Zhao, and H. Liu, "Computer Vision-Based Grasp Pattern Recognition With Application to Myoelectric Control of Dexterous Hand Prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 9, pp. 2090-2099, 2020.

[15] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[16] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics." https://github.com/ultralytics/ ultralytics, 2023. Accessed: February 30, 2023.