Evaluating Efficient Object Detection Models for Real-Time Robotic Applications

UIC UNDERGRADUATE RESEARCH FORUM 2023

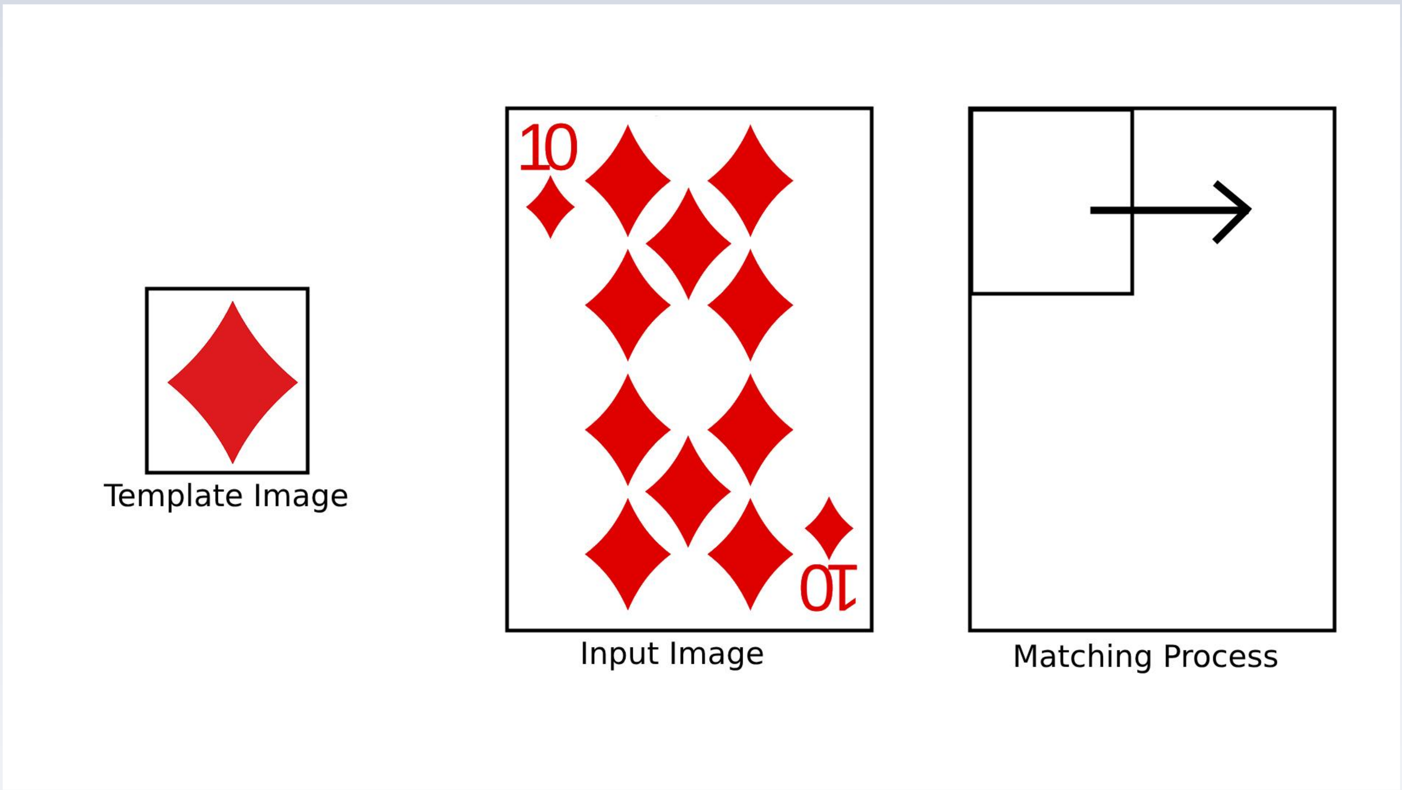
Ahmed Shahab, University of Illinois Chicago, Robotics & Motion Laboratory

Background

Deep learning models achieve excellent results in object detection but typically require powerful hardware, such as GPUs or TPUs, which are impractical for many robotics and embedded applications. This project evaluates three lightweight object detection methods — Template Matching, SVM with HOG, and YOLOv11n — to identify approaches that are both accurate and efficient for real-time robotic applications.

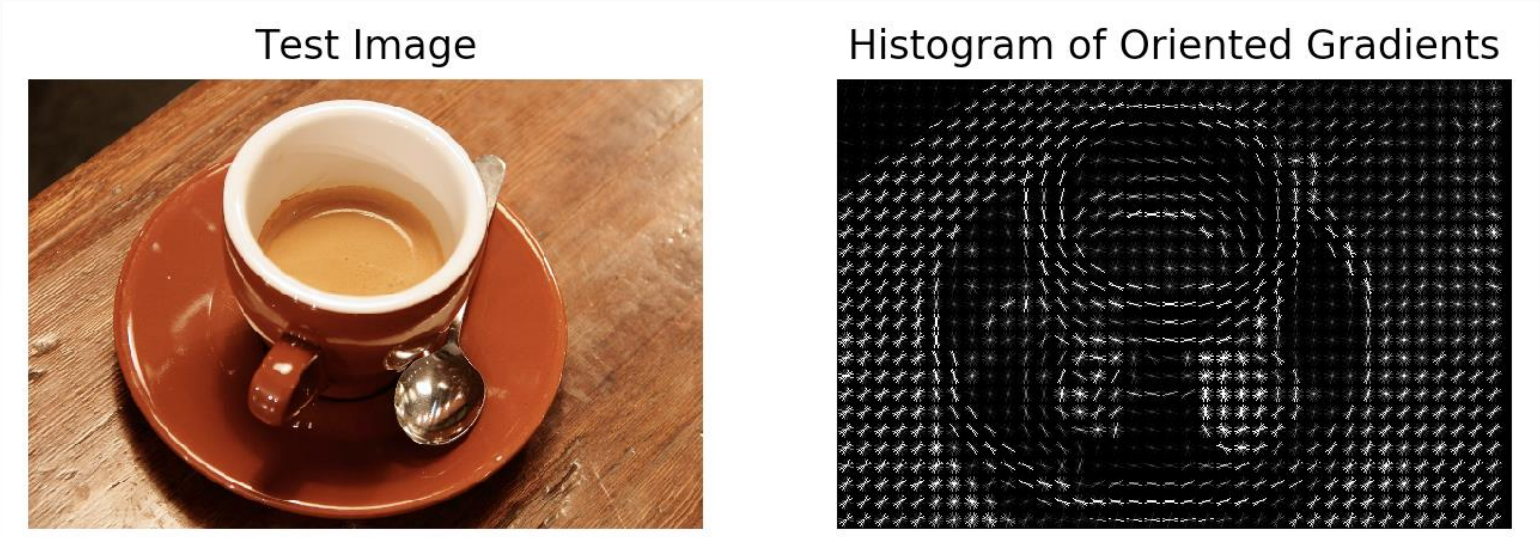
Methods

Template Matching detects objects by sliding a template over the input image and computing similarity at each location. The position with the highest match indicates the object. In this project, 20 cardboard box templates were used, and the one with the highest correlation determined the bounding box.



YOLO (You Only Look Once) is a deep learning model. In this case, the nano-sized model was trained on 621 images of cardboard and 815 images of office spaces without cardboard.

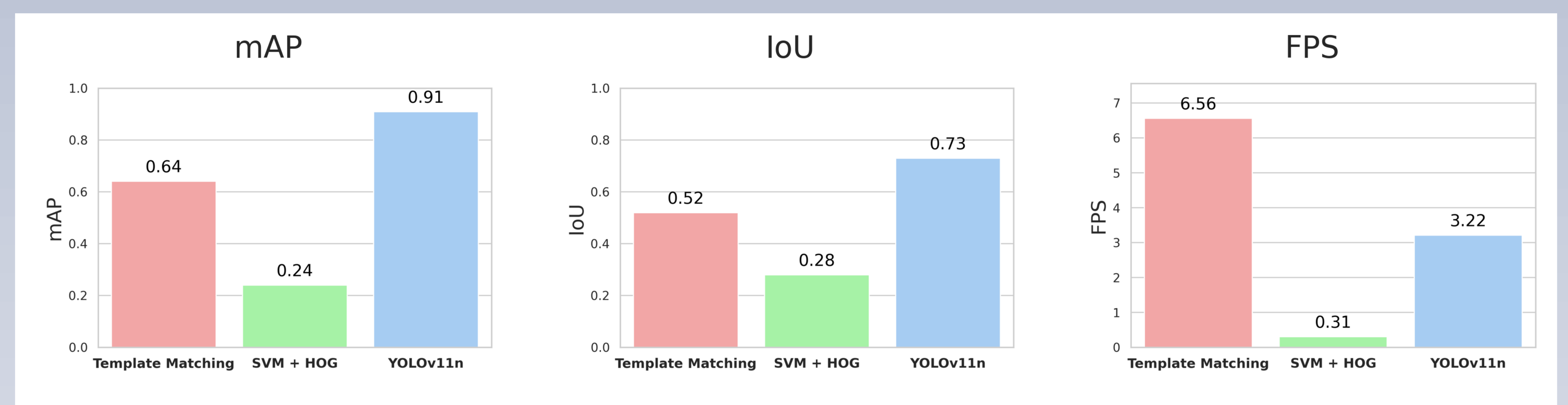
SVM with HOG detects objects by first extracting edge orientation histograms (HOG features), then classifying regions using a linear SVM. The model was trained on 1,125 cardboard images and 1,412 background images.



Results

All the models were tested on images taken in an office that emulated a laboratory environment. These images were then manually labelled and used to measure the model performance.

All models were evaluated on a laptop with a 13th Gen Intel(R) Core(TM) i9-13900H and 32 GB of RAM. The program was restricted to only using 1 out of the 14 cores available.



Key takeaways:

- YOLO achieves the highest accuracy across all the models with an intersection over union (IoU) of 0.73 and mean average precision (mAP) of 0.91
- Template matching was the fastest method, over 103% faster than YOLO, but had lower accuracy
- SVM and HOG performed the worst overall with low speed and accuracy

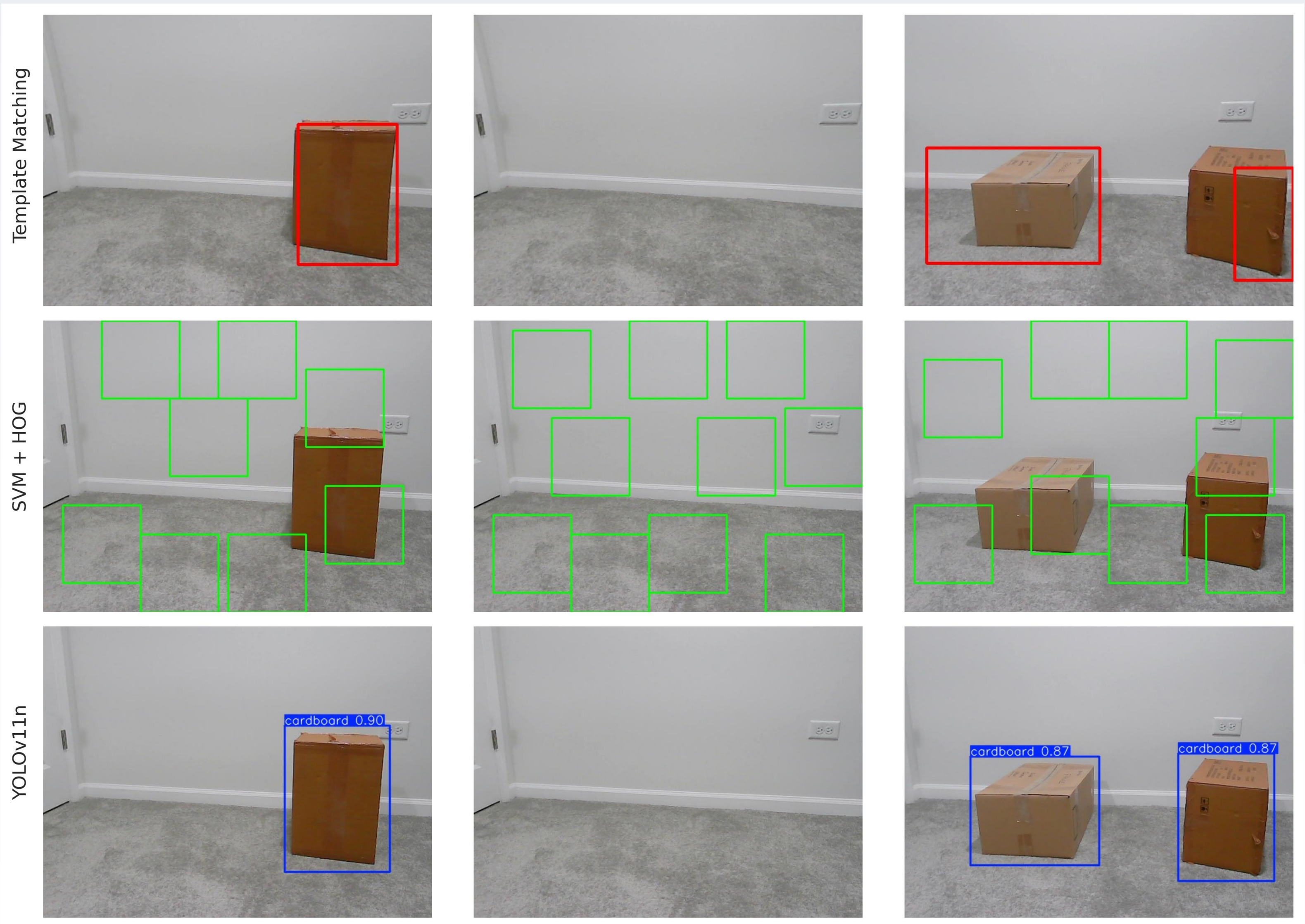


Figure showing model performance on the testing set

Future Work

The next step would be to deploy the YOLO model onto a quadruped robot for real-time object detection. Using a depth camera and a motion-planning algorithm, the object detection results can be used to help the robot navigate its environment. Additionally, the model can be improved by training it on a larger dataset and tuning the hyperparameters.

Picture of the Unitree Go1 quadruped robot

Acknowledgements

I would like to thank Daniel Torres (Ph.D. Student) for his mentorship and guidance throughout this project. This work was conducted in the Robotics & Motion Laboratory under the supervision of Dr. Pranav Bhounsule at the University of Illinois Chicago.

References

1. Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.02767. https://arxiv.org/abs/1804.02767
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934. https://arxiv.org/abs/2004.10934
3. Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 886–893.
4. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.
5. Kälin, M., & Fua, P. (2012). Learning People Detection Models from Few Training Samples. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
6. Brunelli, R. (2009). Template Matching Techniques in Computer Vision: Theory and Practice.