Live Environment Object Recognition and Model Correspondence
Project Proposal

Jay Ho
hophuc@msu.edu

Samuel Hus
hussamue@msu.edu

Suraj Karthikeyan
karthik3@msu.edu

David Khankin
khankind@msu.edu

October 3, 2023

1 Problem Description

Being able to identify items in a 3D environment has been a perennial challenge in the realm of computer vision. Over the years, researchers and scientists have grappled with the complexities of transforming raw visual data into meaningful, actionable information. The nuanced interplay of algorithms, data processing techniques, and machine learning models has been meticulously explored and refined, culminating in a rich tapestry of research literature and breakthroughs. Pioneering studies have laid the groundwork for understanding the intricacies of 3D object recognition and have paved the way for the development of robust algorithms capable of deciphering the multidimensional nature of the physical world[1].

With this in mind, implementing the ability to have a readily available application that can help recognize items can be used on a large scale. Whether this is used in general analysis or for the purpose of maintenance, the use cases are extremely vast. The primary goal of this project is to take information from a live environment, such as an app that uses your phone camera, where the user is able to select objects by clicking on them, and the program will separate these objects from the background and find where this object may correspond to on an uploaded model. As a concept, this would ideally work with corresponding to a 3D model, but for the purpose of this implementation, we will remain within the constraints of corresponding to a 2D image/model.

The primary goal of this project is to take information from a live environment, such as an app that uses your phone camera, where the user is able to select objects by clicking on them, and the program will separate these objects from the background and find where this object may correspond to on an uploaded model. As a concept, this would ide-ally work with corresponding to a 3D model, but for the purpose of this implementation, we will remain within the constraints of corresponding to a 2D image/model.

To clarify a use case instance, having uploaded a 2D solution of a puzzle, one can use the app to click on an object in the environment, and all potential areas on the puzzle image that have a potential match will be highlighted or circled. Note: This does not imply that shape will be recognized as part of this project. The goal is to match details, features, and color to a larger image and recognize if something may not be a part of that.

Another example, we can take a look at the windows logo, where the 4 corners are red, green, blue, and yellow. If one were to select a yellow sheet of paper in the environment, the program would highlight/circle the yellow portion of the windows logo that was pre-uploaded, regardless of the shape of the paper.

Ideally, if one were to take something with more detail, this would also be able to be recognized and properly matched to an image.

2 Related Work

In the realm of object extraction and segmentation, the work on BASNet stands out as a pertinent reference point for its significant contributions in this field. BASNet, an acronym for Boundary-Aware Salient Object Detection Network, has made notable strides in precisely identifying and extracting salient objects within complex environments. This technique has demonstrated a deep understanding of the spatial relationships and boundaries of objects, thereby enhancing the accuracy of object delineation. The core principle of BASNet, which emphasizes the

importance of boundary information, resonates with the objectives of my research in object extraction. By leveraging the insights from BASNet's approach to extract important objects from their surroundings, our work seeks to build upon this foundation to further refine and advance the state-of-the-art in object extraction, with a particular focus on using these extracted objects in correspondence to an additional pre-processed model. In this manner, the knowledge and techniques developed in the context of BASNet provide a valuable reference and inspiration for the evolution of object extraction methodologies, contributing to the broader discourse in computer vision and image processing.[2]

In recent studies within the domain of computer vision and image processing, there has been a noticeable transition in the way images are represented. Historically, images have been predominantly represented in pixel-based formats, where each pixel denotes a specific color or intensity in a fixed grid. However, the advent of graph-based representations has transformed this conventional approach. By conceptualizing images as graph structures, where nodes might represent regions or features and edges capture the relationships or proximities between these features, a more flexible and potentially richer representation is achieved.

This paradigm shift towards graph-based representations offers a plethora of opportunities for advanced image analysis. One of the most compelling advantages is the integration of graph neural networks (GNNs) into image processing workflows. GNNs, which have demonstrated significant success in processing and analyzing structured data, can be naturally incorporated to address challenges related to image recognition when using graph-based image representations. Moreover, the inherent ability of GNNs to capture long-range dependencies and complex patterns can potentially enhance the robustness and accuracy of image recognition tasks. As the field continues to evolve, the synergy between graph-based image representations and graph neural networks promises to redefine the frontiers of image analysis and recognition techniques[3].

3 Project Milestones

Figure 1 below shows our project milestones and timeline to completing them. This is divided into four distinct milestones which represent key phases of our research endeavor, each with specific objectives and tasks. The following milestones are presented in chronological order: Data Compilation, Solution Development, Solution Deployment, and Iterative Enhancement.

References

[1] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," Proceedings Shape Modeling Applications, 2004 Genova, Italy, 2004, pp. 145-156, doi: 10.1109/SMI.2004.1314502.

[2] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan and M. Jagersand, "BASNet: Boundary-Aware Salient Object Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 7471-7481, doi: 10.1109/CVPR.2019.00766..

[3] Xu, N., Nikolentzos, G., Vazirgiannis, M., Boström, H. (2022). Image Keypoint Matching Using Graph Neural Networks. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds) Complex Networks Their Applications X. COMPLEX NETWORKS 2021. Studies in Computational Intelligence, vol 1073. Springer, Cham.
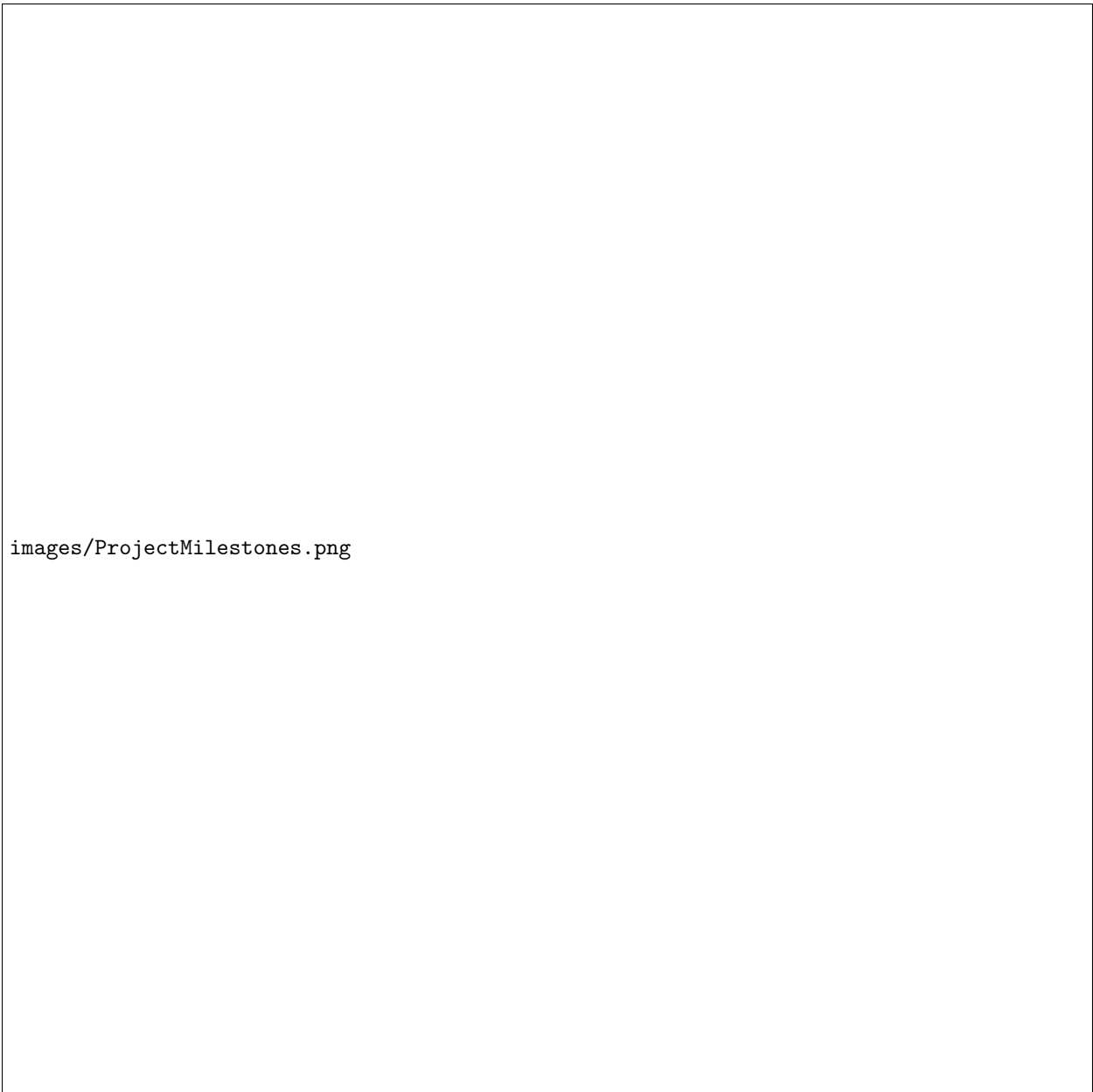
Figure 1: Milestones and Expected Completion Dates.