Automated Novel Object Discovery and Detection in Unlabeled Image Datasets

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

The proposal outlines our approach to tackle the object discovery challenge set by OBJ-DISC. We aim to develop an algorithm capable of identifying and grouping semantically coherent objects from large unlabeled datasets and then train an object detector for each identified cluster.

1 Introduction

Object discovery is vital for understanding and interpreting the vast amount of visual data generated daily, leading to advancements in AI's image recognition capabilities. The OBJ-DISC challenge addresses this by focusing on identifying and categorizing new objects within large, unlabeled image datasets. Our project aligns with these objectives, aiming to significantly advance AI's ability to recognize and understand diverse objects, thereby enhancing machine learning models' generalization and application in real-world scenarios.

2 Objectives

The objectives are to create a system that can group objects in images based on semantic similarity without needing predefined labels, and to validate these clusters by training and assessing object detectors on them, ensuring they can effectively identify these objects in real-world settings.

3 Motivation

Object recognition is very important in understanding visual data, which there are vast amounts of and which increases daily. Given the vastness of visual data, prelabeling objects to allow for training of algorithms becomes less feasible. Therefore there is significant incentive to automate this and reduce the need for human work in labeling image sets. This is why it is important to increase AI's ability to categorize unlabeled objects in datasets, as this will allow human work to focus on other areas. One area where object recognition is important is self driving cars, where the vehicle needs to know if it's approaching something that requires a stop like a person or animal or a non obstacle like a garbage bag. Some negative consequences from object recognition stem from if we have faulty object recognition, such as misidentifying something and causing harm as a result, which is all the more reason to improve object recognition.

4 Methodology

[TODO]This is base on some research may need to change later on

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

4.1 Object Discovery in Unlabeled Data

Utilize self-supervised learning techniques to discover object clusters within the unlabeled dataset. Techniques such as MOST (Multiple Object localization with Self-supervised Transformers) could be particularly useful here, as they allow for the discovery of multiple objects within images without prior training. The goal would be to segment these images into meaningful clusters that potentially represent different object classes.

4.2 Feature Extraction and Clustering

Extract features from these images using pre-trained models or self-supervised methods and then apply clustering algorithms (like K-means, DBSCAN, or hierarchical clustering) to group similar regions together. Each cluster would ideally represent a different object class. The number of clusters, M, would not be known a priori and might need to be determined based on the dataset's inherent structure, which can be evaluated using metrics like silhouette scores or the Davies–Bouldin index.

4.3 Labeling Known Objects

Leverage the labeled dataset of K known objects to label clusters that closely match these known categories. This could be done by training a supervised classifier on the labeled data and then applying this classifier to the centroids or representative samples of the clusters derived from the unlabeled data.

4.4 Training Object Detectors

For each of the M clusters, train an object detector using the images in the cluster as positive examples and images from other clusters as negative examples. This step might involve fine-tuning pre-trained object detection models (such as YOLO, SSD, or Faster R-CNN) on clustered datasets.

4.5 Evaluation

Evaluate the performance of each object detector on a held-out test set. This set should ideally contain both known and unknown objects to assess the real-world applicability of the discovered object classes and their corresponding detectors. Metrics such as precision, recall, F1 score, and mAP (mean Average Precision) could be used for evaluation.

4.6 Refinement and Iteration

Based on the evaluation, refine the clustering and object detector training processes. This could involve adjusting the number of clusters, changing the feature extraction process, or tuning the object detection models.

This process combines elements of unsupervised learning (clustering and feature extraction from unlabeled data) with supervised learning (using labeled data for initial classifier training and object detector refinement). The success of such a system depends on the quality of the unsupervised object discovery process and the effectiveness of the subsequent object detectors trained for each discovered class.

5 Datasets

Labeled dataset: The object detection dataset, PASCAL VOC 2007 split is considered as the labeled dataset for this challenge. refer to `http://host.robots.ox.ac.uk/pascal/VOC/voc2007/`

Discovery Set: The COCO 2014 train set, without any labels, is used as the discovery dataset. The remaining categories, not common with PASCAL-VOC, are considered the novel categories. refer to: `https://github.com/cocodataset/cocoapi`

Pre-training dataset: To train object detection datasets, ImageNet pre-training is a standard practice. refer to: `https://www.image-net.org/challenges/LSVRC/2012/2012-downloads.php`

Evaluation set: All systems will be evaluated on the discovery performance and object detection performance. For object discovery, results are reported on the COCO 2014 train set. For object detection on the 20 known classes and the newly discovered objects, results will be reported on the COCO minival set.

The license for these datasets states:
Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

6 Relevant Papers

This paper: Object Detection with Deep Learning: A Review covers a basic overview of how object detection has progressed over time as deep neural networks have advanced, as well as covering a bit of history of non deep learning object detection.
This paper Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review covers various methods for object detection that have been used over the years and the concepts underlying advancements in Object detection. It's useful in deepening our understanding of this subject to better implement our approach.

7 Expected Outcomes

The expected outcomes are to surpass the performance metrics set by the baseline code provided in the challenge and to apply theoretical knowledge from the course to maximize the efficiency and effectiveness of the algorithm, striving for optimal results in object discovery and detection.

8 Moonshot Plan

The moonshot plan is to get performance matching or surpassing that of a human (recognizing real objects that humans don't). This would require lots of fine tuning and time taken to see how adjustments affect testing performance. It would also require a very deep understanding of the images being evaluated, as it could allow for tuning the algorithm around those features. It would also be difficult to avoid overfitting when trying to acheive such precise results.

9 Timeline and Milestones

Proposals due March 15th
Check-ins week of April 11th
Summary slides due May 2nd
Report and presentation due May 9th
[TODO]Add intermediate deadline(progress check)

10 Team Structure and Responsibilities

[TODO]specific job for each member

11 Resources and Tools

[TODO]base on the algorithm

12 Evaluation and Testing

$$performance = \frac{\sum_1^N Purity_i \cdot C_i}{D}$$
$$N = \text{Total number of clusters}$$
$$C_i = \text{Number of elements in the cluster } i$$
$$Purity_i = \text{Purity of cluster } i$$
$$D = \text{Total number of instances in the dataset}$$

13 Conclusion

In conclusion, our project aims to advance the field of AI in object discovery and detection by developing a system that outperforms existing baseline models and applying our academic knowledge to optimize the algorithm. Our approach is systematic and grounded in robust methodologies, ensuring compliance with challenge standards while striving for innovation and improved performance in object recognition tasks.

References

[TODO]Need to be fill in

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?
   [Yes] We cover everything in the abstract throughout the rest of the proposal
   (b) Did you describe the limitations of your work?
   [Yes] We describe in section 7 which is expected outcomes what we expect to accomplish and in section 8 which is the moonhshot plan, the possible although unlikely best results we could get.
   (c) Did you discuss any potential negative societal impacts of your work?
   [Yes] We talk a bit about possible negative societal impacts in the motivation section, such as its possible use for increased surveillance
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them?
   [Yes] We have
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]

(b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators?
   [Yes] We cite the creators in section 5 datasets and in our references
   (b) Did you mention the license of the assets?
   [Yes] We put a copy of the license information in section 5 datasets
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating?
   [Yes] We cover how the license states we can use the datasets as long as credit is provided, which it will be.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content?
   [No]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

Optionally include extra information (complete proofs, additional experiments and plots) in the appendix. This section will often be part of the supplemental material.