Khanh Huynh

ITAI 1378

Prof. Patricia McManus

12/2/2024

Midterm

1. Report

Data Selection: PASCAL VOC 2007

PASCAL VOC 2007 is the dataset selected for this project and is among the best-known benchmark datasets for various object detection tasks. The PASCAL VOC 2007 dataset engages 20 distinct objects such as people, animals, vehicles, and other commonly found items. It is the adaptation that makes the dataset versatile and representative of object detection tasks. As a result, the dataset has been tested extensively at various object detection competitions, establishing it as a benchmark in the field.

Justification

The selection of a PASCAL VOC 2007 dataset is based on its wide recognition in the domain. PASCAL VOC is a widely recognized and heavily used dataset in computer vision research. The results can be compared against those of previous studies to identify overlaps with the approaches used. The 20 object categories span a wide range of real-world items that will allow the model to learn a wide range of features and flexible/back-end insolubility for complex Object Detection challenges.  
 Consistency with Instruments: The dataset is readily available through TensorFlow Datasets (TFDS), thus facilitating simple data loading and direct usage within the TensorFlow framework. The data set, therefore, serves beneficial for transfer learning as various pre-trained models exist from the PASCAL VOC data set.  
 Scale and Manageability: In view of constraints such as limited resources, employing 20% of the dataset helps rather smartly balance training time for the model with data required for ensuring satisfactory performance.

1. Risk Management

-Data issues

Some of the major risks associated with the use of datasets such as PASCAL VOC 2007 are concerns arising from data quality and integrity. Notable challenges can be posed during data importation or processing even with what may be considered a reliable and organized dataset. For instance, there can still be instances of an invalid data entry or corrupted images resulting in errors during the training process. Factors that give rise to these issues include files failing to load correctly or the wrong alignment of annotations resulting in import errors or the invalid image error during data preprocessing. Other errors may occur due to differences in the format of image files and label files as well as image size and aspect ratio. A comprehensive data validation check before data processing eliminates these issues. Such a preprocessing pipeline to identify corrupted files, fix mislabels, and standardize images is a great way to reduce errors during training. Furthermore, both manual inspection and automated scripts assist in upstream identification and rectification of data points that could be problematic.  
 Another crucial factor is the data imbalance. PASCAL VOC 2007 may, however, be balanced overall, but there may be an under representative class of objects, which may in turn undermine the model's performance. The models may become inherently biased towards the higher-classified objects, ultimately impairing their very efficient ability to detect other categories with less frequency.

-Computational Limits

Faster R-CNN is obviously powerful but much slower and requires significantly more resources than the seemingly straightforward architecture of SSD MobileNet V2. This change in the architecture may require longer training runs. The runtime for Faster R-CNN was over 2 hours and no output was produced, suggesting that the model is unable to process the data under the current time and resource constraints effectively. The slow performance may be due to the very much more complex construction of Faster R-CNN and larger requirements in terms of memory and time of processing.

To reduce the load of computational requirements, it will be pertinent to refine the model and ascertain that the available hardware will be sufficient to accomplish the desired task. It is apparent now that one option may be to decrease the batch size, simpler models, or even use very light models during the process. Furthermore, one advantage of gradient checkpointing is the memory conservation attributed to it. Early stopping prevents the model from going through unnecessary epochs. If computation still outsizes available resources, using cloud-based infrastructures that would offer better GPU support, or using distributed computing frameworks would be the other option. An initial strategy would be to downscale and implement simpler architectures during initial experimentation, before moving on to more complex structures after feasibility is confirmed.

Occasionally, errors when importing items may arise from insufficient memory or conflict of resources when loading large data sets. The problems arise mainly when the system cannot effectively load important libraries and data files as memory becomes scarce; consequently, these ImportError issues are encountered. Loading such libraries and datasets one at a time into memory may be a good solution to this, and, while performing this, resource usage can still be monitored to ensure it does not exceed the threshold.

1. Reflection

Khanh: Engagement with the Dataset Selection and Preparation subsection was very fortunate, giving me glimpses of the practical side of dealing with real-world data. Opting for PASCAL VOC 2007 as the dataset was an easy decision to make; however, my first impression regarding data quality and the process of dataset preparation for model training soon changed.   
The biggest problem I faced was to overcome class imbalance as an issue in the overall balance of the dataset. Some categories of objects for which the dataset was really biased were very sparse; hence, I started implementing some data augmentation methods to curb this bias.   
  
This section established the relevance of data cleansing and augmentation, components quite often underestimated but indeed essential to any machine learning pipeline. I had practical experience in augmenting the dataset using TFDS, which further facilitated the management and loading of the dataset. I also learned that what happens to the data in the preparation step can create quite a large difference in how the model performs or works.

Enrique: One of the most important things that I learned about during my time was computational limits. While SSD MobileNet V2 was ready for efficiency, to work with Faster R-CNN would require more efficient hardware in addition to longer training times. I spent a lot of time researching ways to optimize the model, including trial-and-error modifications of reduced batch sizes and exploring unsuspected model-pruning techniques to reconcile with the balance of performance and speed. Besides that, I ran into a barrage of long training times and import errors during this whole process, especially when the dataset was huge or when the system resources were pushed to the extremes.  
 Through this experience, I saw how architecture affects both accuracy and efficiency, as well as the importance of optimizing computational resources to create a synergy between both. I learned how critical it is to cast a wider net while testing various models before settling with one, for, although Faster R-CNN was more accurate, it could bring the computational requirements down to its knees.

Jake: In the Other Techniques section, I have dealt with specialized methods such as ensemble techniques, post-processing, model quantization, and transfer learning. These techniques, I tended to discover, are not only interesting but also very critical methods for improving the model performance. My initial explorations of ensemble methods studied integrating predictions from various models to achieve higher overall accuracy. The goal was to take advantage of different models and blend them in a way that reduces the biases and errors of each model. This really set me down the path of gaining first experience in blending multiple models into one for better predictions. But I did face a large challenge when it came to appropriate model selection to ensure that the models could be effectively combined without falling prey to overfitting.  
 On the side of model quantization, I have explored such methods as cared for various methods for minimizing model size with enhanced inference speed. I had to balance quality and model size quite cautiously to make for good performance, especially real-time performance.  
 Finally, transfer learning seemed relevant in the process of adoption of a pre-trained model onto the PASCAL VOC 2007 dataset. I took very large datasets, trained from the COCO dataset, and refined onto our smaller dataset via pre-trained model weights. This minimization reflected not just on reducing the time to train, but definitely a convergence enhancement. The real-world use of transfer learning has opened my eyes to how much information can be passed across domains and thus cut the time and computational resources in the process.  
 This section gave me a chance to explore the techniques one considers essential for the enhancement of model performance but are often neglected in most initial tutorials. I understood the way ensemble methods and model quantization help take part in the quick modeling and provides more accuracy, not to mention the jeu d'esprit implied between processing and the model architecture for the best performance.

1. Innovations

We are interested in exploring the innovative approach of extending object detection to encompass video data rather than limiting it to static images. By looking at the temporal dimension of the scene, I aim to make a system that tracks an object from frame to frame and detects changes in its state over time. This is very pertinent for applications like surveillance, action recognition, and autonomous driving, where the understanding of the motion and progress of an object is quite significant.