Model Evaluation and Fine-Tuning Using Data Augmentation

Team:RIL-LAB

GemBench Challenge

Abstract

In this report, we discuss an experiment that involves training a model on raw data and evaluating its performance across multiple levels. The model's evaluation results for levels 1 to 4 were initially 91.03%, 44.46%, 41.90%, and 0%, respectively. After applying data augmentation techniques and fine-tuning the model, we observe an improvement in the evaluation results, yielding scores of 92.56%, 46.79%, 42.86%, and 0% across the same levels. This report describes the experimental setup, evaluation methodology, and the impact of data augmentation on the model's performance.

1 Introduction

The rapid development of machine learning has made it increasingly essential to enhance model accuracy through various techniques. In many applications, data augmentation serves as a powerful tool for improving model performance, especially when data is limited or imbalanced. This report investigates the impact of data augmentation on the performance of a trained model, which was initially evaluated using raw data.

The experiment is conducted in two phases: 1. Training and evaluation of a baseline model using raw data. 2. Application of data augmentation techniques to fine-tune the model and comparison of results.

The aim of this study is to assess whether the application of data augmentation leads to a significant improvement in model accuracy across multiple levels of evaluation.

2 Methodology

2.1 Data Preparation

The dataset used in this experiment consists of raw data obtained from a robotic system. The data comprises various features, including 3D point clouds, RGB images, and ground truth action labels. The dataset is divided into training and validation sets, with the training set being used to train the model and the validation set for evaluation.

For the initial phase of the experiment, no data augmentation techniques were applied. The model was trained using the raw data only. For the second phase, a range of data augmentation techniques were applied to the training data to artificially increase the diversity of the dataset and reduce overfitting.

2.2 Model Training

The model is a deep learning architecture designed to predict the 3D poses and actions of robotic systems. The architecture consists of convolutional layers for feature extraction, followed by fully connected layers for prediction. The model is trained using a combination of supervised learning techniques, where the ground truth labels (actions) are compared against the model's predictions using a loss function.

In the initial phase of the experiment, the model is trained on the raw data for a fixed number of epochs. The model's performance is then evaluated across four levels, each corresponding to a different evaluation metric:

- **Level 1: Novel placements** - Performance when the model is required to handle novel placements of objects.
- **Level 2: Novel rigid objects** - Performance when interacting with objects that were not seen during training (novel rigid objects).
- **Level 3: Novel articulated objects** - Performance on objects that are articulated or have multiple parts, not seen during training.
- **Level 4: Long-horizon tasks** - Performance on tasks that require sequential actions or long-horizon reasoning, such as stacking cups or putting items in a drawer.

2.3 Data Augmentation Techniques

For the fine-tuning phase, several data augmentation techniques were applied to the training data to improve the model's robustness and generalization capabilities. These techniques included:

- **Rotation**: Randomly rotating the data points around the z-axis to simulate different perspectives.
- **Translation**: Randomly translating the data points to simulate variations in the robot's positioning.
- **Scaling**: Scaling the data points to simulate different object sizes and distances.
- **Noise Injection**: Adding small amounts of noise to the data to simulate sensor inaccuracies.
- **Random Sampling**: Randomly selecting subsets of data points to simulate variations in the input data.

These augmentations were applied during the training phase to improve the model's ability to generalize across different conditions. The model was then fine-tuned using the augmented data, and the evaluation process was repeated.

3 Results

The evaluation of the model was performed across four levels, and the results are summarized in Table 1. The model was initially evaluated on raw data, followed by an evaluation after data augmentation.

Evaluation Level	Raw Data (Baseline)	Augmented Data (Fine-Tuned)
Level 1: Novel placements	91.03%	92.56%
Level 2: Novel rigid objects	44.46%	46.79%
Level 3: Novel articulated objects	41.90%	42.86%
Level 4: Long-horizon tasks	0%	0%

Table 1: Model Evaluation Results Before and After Data Augmentation

As shown in Table 1, the model's performance improved across all levels after the application of data augmentation techniques:

- **Level 1: Novel placements**: The overall accuracy increased from 91.03% to 92.56%, indicating a general improvement in the model's ability to predict object placements.
- **Level 2: Novel rigid objects**: Performance on novel rigid objects improved from 44.46% to 46.79%, showing that the model learned to handle new objects.
- **Level 3: Novel articulated objects**: Although the improvement was marginal, the performance increased from 41.90% to 42.86%.
- **Level 4: Long-horizon tasks**: No improvement was observed at this level, as the model continued to struggle with long-horizon tasks like stacking cups or putting items in drawers.

4 Discussion

The results indicate that data augmentation techniques can lead to significant improvements in model performance, especially at the lower levels of evaluation. By introducing variations in the training data, the model becomes more robust and capable of generalizing across different conditions.

At **Level 1**, the overall accuracy showed a noticeable improvement, demonstrating that the model is better able to handle a wide range of inputs after fine-tuning with augmented data. This suggests that the augmentation techniques provided sufficient variability to allow the model to adapt to different scenarios.

At **Level 2**, the model's performance improved, indicating that the data augmentation techniques helped the model make finer distinctions between different input data points. However, the improvement was less pronounced at **Level 3**, where the model was still able to improve marginally but struggled to make significant progress.

At **Level 4**, no improvement was observed, indicating that the model continues to face significant challenges in predicting the most difficult samples. This could be due to the inherent complexity of the data at this level, which may require more advanced techniques, such as additional model architectures or more sophisticated augmentation strategies.

5 Conclusion

In conclusion, the application of data augmentation techniques resulted in noticeable improvements in the model's performance across most evaluation levels. The overall accuracy increased from 91.03% to 92.56%, and the performance on novel rigid objects and articulated objects also showed improvement. However, challenges remain, particularly in long-horizon tasks. Future work could explore more advanced augmentation techniques, as well as alternative model architectures, to further improve performance.