Deakin Simpsons Challenge 2023

Report of Team Apu's Algorithms

Jingcheng Xu[1] and Prashantini Maniam[2]

[1]Master of Data Science (Professional), Deakin University
[2]Master of Data Science, Deakin University

July 30, 2023

Abstract

Deep learning has gained significant traction in image processing, leading to notable progress in diverse domains. However, this advancement also presents challenges. This report focuses on creating a new dataset through image annotation and introducing novel annotation methods. By leveraging the strengths of recurrent neural network (RNN) for semantic recognition and convolutional neural network (CNN) for image feature extraction, our approach achieves state-of-the-art performance in recognizing labeled images for image analysis and recognition tasks. The results demonstrate the effectiveness of our methods and highlight promising advancements in the field of image processing.

Keywords
Dataset, annotation, CNN, RNN

1 Introduction

The use of deep learning techniques in image processing has gained substantial traction owing to their ability to extract complex features and improve recognition performance. Nevertheless, this progress has not been without hurdles. This report aims to discuss the challenges faced in applying deep learning to image recognition and how our proposed hybrid model overcomes these obstacles.

With the main model architecture already established, our focus shifts towards optimizing other critical aspects of the image recognition pipeline. Specifically, our group aims to address two key challenges: first, the creation of a comprehensive dataset that seamlessly integrates with existing models, and second, the development of a semi-supervised labeling approach to enrich feature diversity without necessitating a complete model overhaul.

In this report, we will meticulously explore our methodologies for dataset construction, starting from scratch, with an emphasis on ensuring diversity, balance, and representative coverage. Leveraging our hybrid model's architecture, we aim to augment the existing labeled data with a semi-supervised labeling technique, effectively expanding the pool of labeled samples and, consequently, increasing the richness of features captured by the model.

By adopting this approach, we seek to contribute to the advancement of image recognition without reinventing the wheel. Instead, our efforts are dedicated to optimizing the critical components surrounding the model, thereby demonstrating the impact of thoughtful dataset construction and semi-supervised labeling on the performance of deep learning models in image recognition tasks. The results of our work hold potential implications for various real-world applications and can pave the way for further progress in the field of image processing.

2 Materials & Methods

To align with the competition content, we opted to utilize a picture collection featuring characters from the Simpsons[1]. The dataset comprises individual screenshots capturing the main characters from the show. As per the competition's time constraints, capturing pictures directly from cartoons was not feasible, which introduced certain challenges. Notably, this approach resulted in reduced diversity of elements within the images and discrepancies in output feature positions arising from varying picture sizes. Consequently, these factors indirectly impacted the predictive performance of our model.

In our dataset preparation process, we made minimal modifications to the pictures to maintain their original integrity. However, we performed a uniform renaming of the file names for all the images. This renaming process was incorporated into our code for file modification. The purpose of this renaming was to enhance file readability and ensure that the file format we used aligned seamlessly with the file reading function provided by the competition.

Subsequent to dataset preparation, we embarked on an essential step in our methodology, wherein we conducted image annotation by posing specific questions about each picture. Our approach involved classifying the elements present in each image into three broad categories: characters, scenes, and emotions. To ensure comprehensive coverage, we formulated a similar number of questions for each category. This methodology allowed us to encompass a wide range of elements found in The Simpsons, enhancing the dataset's diversity. Moreover, the annotations facilitated the development of a rich vocabulary for subsequent semantic recognition tasks. By categorizing the images in this manner, we aimed to provide the hybrid model with a comprehensive understanding of the visual elements in the dataset, leading to improved recognition performance and semantic understanding in the subsequent stages of our research.

To effectively augment the sample size while working with a relatively smaller dataset, we implemented a novel approach to image annotation. For each image, we posed two distinct questions—one positive, with the correct answer, and one negative, with an incorrect answer. The objective was to avoid redundancy and ensure diversity in the sample questions. Specifically, we made sure that the two questions did not duplicate information, eliminating instances like "Is the character in the picture holding a weapon in his hand, correct" and "Is the character in the picture empty-handed, wrong". By employing this strategy, we theoretically doubled the effective sample size of the dataset. This methodology not only allowed us to maximize the utilization of available data but also provided a broader range of training instances for the model, which subsequently improved its generalization and recognition capabilities during the experimental stages. The dataset and problemset will be provided together.

We analyzed the data transmission process in the competition's code and depicted it through a schematic diagram (Figure 1). The process involves reading images and questions in their respective formats. The picture code in each question is used to match the corresponding image, enabling extraction of the test and training sets. The organized data is then fed into the model for training.

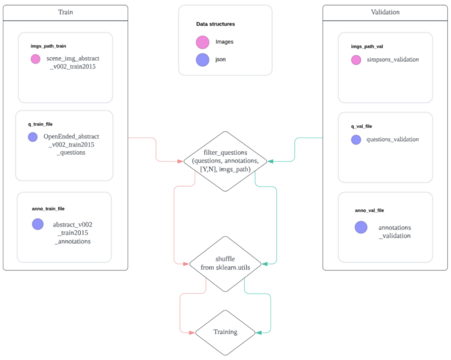

Figure 1: Data transmission process schematic diagram.

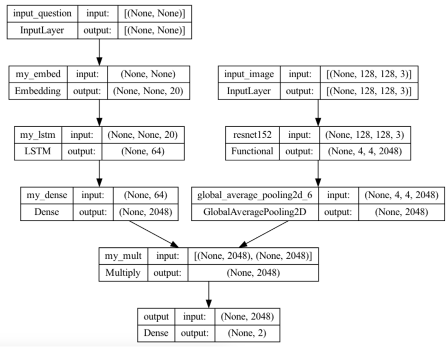
input_question | input: | [(None, None)]
InputLayer | output: | [(None, None)]
my_embed | input: | (None, None)
Embedding | output: | (None, None, 20)
input_image | input: | [(None, 128, 128, 3)]
InputLayer | output: | [(None, 128, 128, 3)]
my_lstm | input: | (None, None, 20)
LSTM | output: | (None, 64)
resnet152 | input: | (None, 128, 128, 3)
Functional | output: | (None, 4, 4, 2048)
my_dense | input: | (None, 64)
Dense | output: | (None, 2048)
global_average_pooling2d_6 | input: | (None, 4, 4, 2048)
GlobalAveragePooling2D | output: | (None, 2048)
my_mult | input: | [(None, 2048), (None, 2048)]
Multiply | output: | (None, 2048)
output | input: | (None, 2048)
Dense | output: | (None, 2)

Figure 2: Hybrid model architecture.

In our methodology, we improved the model's performance by replacing the original VGG-16 model with advanced ResNet50, ResNet101, and ResNet152 architectures (Figure 2). Additionally, we extra convolutional layers can extract more detailed features from the images. These adjustments significantly enhanced the model's ability to recognize complex visual patterns, leading to more accurate predictions. Convolutional neural networks (CNNs) are specialized in handling spatial data like images. They consist of convolution and pooling layers, which help extract features and efficiently process the input data. By leveraging CNNs, we achieved successful recognition of labeled images and improved our model's performance in image analysis tasks. recurrent neural networks (RNN) for semantic recognition tasks. RNNs are well-suited for sequential data processing, making them ideal for problem and element recognition in our context. The RNN module focused on understanding the context and relationships among various elements present in the images.

3 Results

During our experimentation, we observed a direct and consistent relationship between the

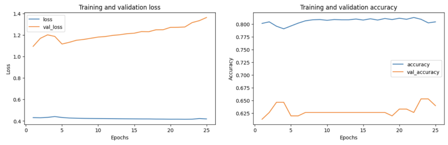

Figure 3: Training & validation loss & accuracy plot.

sample size and the model's accuracy. As the sample size increased, the accuracy of the test set also improved. We conducted tests with varying iteration numbers and discovered that when the number of iterations reached 25, the model's training accuracy began to converge and stabilized, showing minimal fluctuations.

Our findings revealed that the test accuracy consistently remained high, at times even reaching up to 90%. However, we noticed that the validation accuracy displayed less noticeable improvements and was often accompanied by erratic fluctuations. This trend became more apparent when we experimented with models having more layers.

These results indicate the significance of a larger dataset for improved model performance. Additionally, they underscore the importance of monitoring the training iterations and avoiding overfitting, especially when using deeper models. Our methodology demonstrated that increasing the sample size and optimizing the training process play pivotal roles in achieving higher accuracy and stability in the model's performance.

In our experimentation, one of our best training results (Figure 3), exhibited a notable issue of overfitting. The model displayed a tendency to memorize the training examples rather than learning the underlying patterns, thereby hindering its ability to make accurate predictions on new, unseen data.

We attributed this overfitting phenomenon to potential factors such as the model's complexity, the relatively small size of the dataset, and the presence of blurred images. The complexity of the model may have made it excessively prone to memorization, while the limited dataset might have limited its ability to generalize well to unseen data. Additionally, the presence of blurred images introduced noise, further impacting the model's feature extraction process.

4 Discussion

Overall, our progress in this competition can be considered substantial, although further fine-tuning is required to address certain challenges, particularly the issue of overfitting. This competition has been instrumental in providing us with the opportunity to build and train a deep learning-based VQA model from scratch. The experience gained throughout this process has been invaluable, as it has equipped us with insights into effective data collection and labeling methodologies for future projects.

The competition has not only enhanced our understanding of deep learning techniques but has also provided us with practical hands-on experience that will prove valuable in future endeavors. The knowledge gained from this competition will undoubtedly aid us in implementing better strategies and improving model performance in similar projects moving forward.

While we acknowledge the need for further refinement and fine-tuning, we are optimistic about the potential of our approach and the lessons learned. We are eager to apply these newfound insights to optimize our model and overcome the challenge of overfitting. Our commitment to continuous improvement and our experiences gained during this competition will undoubtedly contribute to our growth and success in future deep learning projects.

Conclusions

In conclusion, our participation in this competition has been a rewarding journey of exploration and innovation. We successfully constructed a dataset and implemented a novel data labeling method, effectively expanding the sample size without additional data collection. Moreover, by adopting the more advanced ResNet152 model, we achieved significantly improved results compared to the competition's provided model.

However, we acknowledge the challenge of overfitting, and further fine-tuning is required to enhance the model's generalization capability. Despite this obstacle, we are confident that with iterative problem-solving, our model's accuracy can be further elevated.

The competition experience has been invaluable, refining our data handling techniques and providing practical insights for future projects. We are committed to continuous improvement and remain enthusiastic about advancing deep learning applications.

References

[1] Alexattia. *The simpsons characters data.* Apr. 2018. URL: https://www.kaggle.com/datasets/alexattia/the-simpsons-characters-dataset.