COVID-CT: Model Search

Alfred R. Premkumar, Braun Z. Lippe, Hephzibah Rajan, Suraj Kacham, Uswa F. Hanif

[1]Maryland Innovation and Security Institute (MISI)
621 East Pratt Street, suite 610
Baltimore, MD 21202

Abstract. *This project was done by Interns as part of the MISI Internship program. Interns were divided into six teams in order to work on the project. Our group was responsible for finding possible models to apply with transfer learning that have not yet been tried, as well as train and validate said models.*

1. Introduction

As COVID-19 continues to wreak havoc on our lives, taking action is a key component in combating the global pandemic. By using a database of Lung CT-Scans, we were able to fork a model from the University of California San Diego to create a reliable machine learning model that can determine if a patient has COVID-19.

In order to improve this model, our group was assigned to find, train, and validate possible models to apply to the initial machine learning model. Through trial and error, we decided that using Bilinear Convolutional Neural Networks (CNN) was the most effective method at observing training data images. Using a Graphics Processing Unit (GPU) on our Amazon Web Services (AWS) workspaces, we were able to train and validate a Bilinear CNN as per the methodology proposed by the University of Massachusetts, Amherst. [1]

This project was done through our Artificial Intelligence Project at the Maryland Innovation and Security Institute Summer Internship Program.

2. Bilinear Convolutional Neural Networks

For our task, we chose to implement the Bilinear CNN methodology proposed by the University of Massachusetts, Amherst. Bilinear models are a type of recognition architecture that consists of two element extractors whose outputs are multiplied using outer product at each location of the image and joined to obtain an image descriptor. [1] In other words, a bilinear model is a function of two or more independently, linear, variables. Essentially, becoming a dot product of two vectors. As for better fine-grained categorization, this architecture can display local pairwise highlight communications in a translationally even way. Not only does it model local pairwise feature interactions, but also is able to generalize many odorless texture descriptors such as the Fisher vector, O2P, and VLAD. [1]

The research group's experiments involving linear models where the feature extractors are based on convolutional neural networks (CNN) simplifies gradient computation and allows end-to-end training using image labels only for both networks. [1] After utilizing the networks from the Kaggle dataset, the research group obtained 64.01% accuracy of the SARS-CoV-2 CT scan dataset. Our experiment presents the effects of fine-tuning and the choice of two networks on speed and accuracy of the models. Their model compares unfavorably to the existing model. Moreover, the model is fairly efficient running at 8 frames/sec on a NVIDIA Tesla K40 GPU. We chose to implement a PyTorch version of this promising methodology due to its innate capabilities in fine-grained visual classification problems.

3. Training

In training the Bilinear CNN, three steps were carried out, which are as follows:

First, the fully connected layer of the CNN was fine-tuned on the Caltech-UCSD Birds 200 dataset, which gave a test accuracy of 75.47%. [2] Second, all layers of the CNN were fine-tuned on the Caltech-UCSD Birds 200 dataset, which gave a test accuracy of 84.41%. [2] Third, all layers of the CNN were fine-tuned once again on the COVID-CT dataset, which gave a test accuracy of 64.01%.

These steps were carried out utilizing 'blinear-cnn-faster', a faster PyTorch implementation of the Bilinear CNN methodology. [1]

4. Models

The results gave a test accuracy of 64.01%. The model performed worse than the current state-of-the-art self-trans methodology provided with the COVID-CT dataset, which gives a test accuracy of 86% on the same dataset. As with the nature of neural networks, the reasons for this disparity could be linked to the purpose and method of the Bilinear CNN methodology. [1]

The Bilinear CNN methodology is meant for classification problems with incredibly fine-grained features to distinguish between, such as the differences between bird species or airplane models. [1] Distinguishing between low-resolution lung scans may not fill that fine-grained classification description, as the image size that the Bilinear-CNN is set to train is 448x448, while the COVID-CT dataset is 224x224, so a large amount of quality disparity exists between the two implementations. [1] If more refined COVID-19 CT scan data exists, it is possible the Bilinear-CNN methodology may prove more useful.

5. Conclusion

Our team successfully trained and validated the Bilinear Convolutional Neural Networks (CNN) using the GPU instance on our AWS workspaces. Using 'bilinear-cnn-faster', the model was trained in three separate steps (the Caltech-UCSD Birds 200 dataset twice followed by the COVID-CT dataset) to fine-tune it and make it more efficient. In our final step, we received an accuracy of 64.04% when we used the SARS-CoV-2 CT scan dataset. However, as explained in the model section, the Bilinear CNN model may not be the best fit for our purposes. Either the COVID-CT scan dataset need to be more refined or we need to search for a different option. Our team has demonstrated the result of fine-tuning a model, and although we weren't able to achieve a positive outcome, we know the problems it poses in this case.

References

[1] Lin, T., RoyChowdhury, A., & Maji, S. (n.d.). *Bilinear CNN Models for Fine-grained Visual Recognition*. Amherst: University of Massachusetts.

[2] Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., & Perona, P. (n.d.). *Caltech-UCSD Birds 200 [PDF].* San Diego: California Institute of Technology and University of California San Diego.