DEAKIN SIMPSONS CHALLENGE 2023

Building a machine learning/deep learning model in answering Yes/No questions using Simpson images

Gavinn Fernando
Deakin University
Burwood VIC, Australia

Figure 1. Simpson Characters

ABSTRACT

Artificial intelligence(AI), Machine Learning (ML) and Deep Learning (DL) are hot topics in social media especially with the launch of Microsoft AI platform ChatGPT . AI is becoming increasingly popular for very many industries and individuals in almost all the countries around the world. AI promises to help human and human interactions in many industries, and it is already happening. In my report I am not considering arguing the opinions of the people in this rapidly growing industry but how I achieved my results in this challenge and the use of Recurrent Neural Network (RNN) Architecture and Convolutional Neural Networks (CNN) in image classification. My target was to achieve 80% accuracy, but the result ended lower, limiting my ability to attempt further due to other factors including time and interstate travel. The more time we spend could be the key to getting better results.

KEYWORDS

AI, ML, DL, Neural Networks, CNN, RNN, LSTM, Overfitting, Accuracy

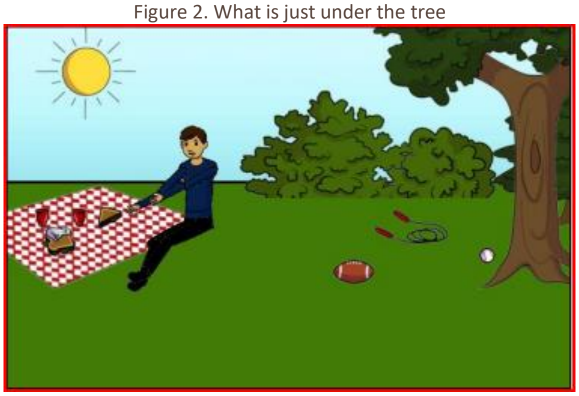
1. INTRODUCTION

Unlike any other time in history, now everyone talks about AI and Machine learning. A few industries including pioneers of AI such as Google and Microsoft have managed to absorb these amazing technologies into their workplaces such as manufacturing where I worked during the last 5 years. In this challenge I noticed how important it is for us to know what we should be doing to achieve the best results that we are after but not trial-and-error methods to apply until we reach some higher percentages compared to the previous results. At my workplace, collaboration activities such as emails and attachments were protected in terms of safety and security when we started using Multifactor Authentication (MFA) and Advanced Threat Protection (ATP) that are heavily dependent on Microsoft AI and ML driven by their engines that have no real time human interactions (Microsoft 2017).

This paper highlights the problem we need to solve ,the methods used to achieve highest possible accuracy using “Tensorflow“ framework in “Google Colab” .

In this challenge, I have focused on binary (yes/no) solutions but not open-ended queries similar to (Q: “What is just under the tree “) ; ref. below picture figure. 2, where all the relevant words in the vocabulary used including “white ball” exists within the answer. How do we know it is a ball ?, is it the shape we consider? There is another ball which is not just under the tree, and it is very much recognizable. What is interesting to observe in the picture below is , without ML or DL being applied to the image, we can reasonably easily identify the object quite accurately through our naked eye, since the picture has high resolution.

Figure 2. What is just under the tree



Source: (Agrawal et al 2016)

In contrast , looking at Figure 3. Which are low-resolution images, answering even a binary question with “yes/no” solution by looking at them is not easy. For example, the first image with a question “is there glass on the table “? ML and DL techniques available could be used to predict the answer very much accurately for such questions with very low-resolution images as below, but in this situation, we find all answers are incorrect even with right vocabulary in place. So, this is the challenge, how to get the best and accurate results using ML and DL?

Figure 3. Low resolution images for binary (yes/no) answers

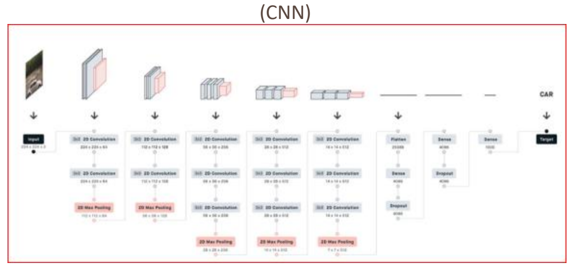
To address this complex challenge, various methods could be proposed to achieve the best results possible.

2. METHOD

Deep Neural Networks (DNN) are very much used in image classifications in Convolutional Neural Networks (CNNs) (Shorten and Khoshgoftaar 2019). Some of the popular CNN architectures are LeNet-5, AlexNet, VGG and ResNet. By preserving special characteristics of the images on the test , these networks could be used to employ parameterized lightly merged kernels.

They are also used in many other state-of- the-art detections and recognitions. In CNNs classification and feature extractions are combined.

Figure 4. Basic convolutional neural network (CNN)



Source: (Amor 2020)

3. DESIGN

Recurrent Neural Network (RNN) Architecture combined with CNN are used in this challenge. Similar to CNN, RNN utilizes training data to learn. In this challenge the RNN model we use is LSTM (long-short term memory) .

Fig. 5. Colab extractions

```
# This model is the deeper LSTM Q from Figure 8 in
# https://arxiv.org/pdf/1505.00468.pdf
def build_model(img_size, vocab_size, num_answers):
    # Define the VGG19 conv_base for image input
    img_input = keras.Input(shape=img_size + (3,), name="input_image")
    img = layers.Flatten()(img_input)
```

Source: (Bouadjenek 2023)

In traditional context, deep Neural Networks adapt when the inputs and outputs are mutually exclusive, however RNN outputs depend on preceding elements within the sequence (IBM 2020). For image captioning and natural language processing DL algorithms are generally used. Image captioning is a written description of an image the RNN works on the language modelling by decoding the encoded output given by CNN (Radhakrishnan 2017).

4. IMPROVING ACCURACY OF IMAGE RECOGNITION MODELS

This challenge was to increase the accuracy of the training dataset and generate a model to be used that goes to the competition with this report. To improve the accuracy of the model, I used the techniques below.

4.1 DATA AUGMENTATION

In this process I managed to create new datasets specifically adjusting the brightness, flipping, rotating, and scaling to improve the model that provides higher accuracy.

4.2 GETTING MORE DATA

Increased the number of images to a larger matrix that fits in to an array of more images than three. This also gave good results in improving the model's accuracy.

Figure 6. 30 images

4.3 HYPER PARAMETER TUNING

Various hyper parameters including batch size, number of layers (Dense and RNN increased to 128) , learning rate , different no of units per layer were attempted that made some changes to the performance of the model.

4.4 ENSEMBLE METHODS

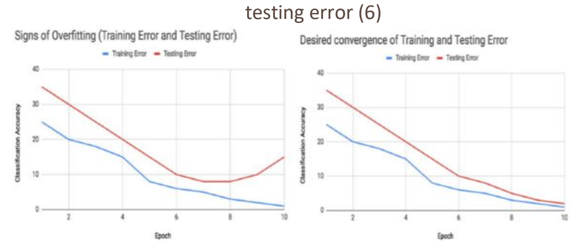
As a result of generating many models I was able to combine them to yield much more accurate results without depending on a single model.

4.5 APPLYING REGULARIZING TECHNIQUES

In AI modeling, an obvious issue that I also faced was , "overfitting". I attempted some techniques such as dropouts and early stopping to inhibit the overfitting of the model to the training data to get the final result.

Below is a good example of the "overfitting" issue and a comparison of such a situation with desired relationship with testing and training errors.

Figure 7. Overfitting plot vs desired relationship of training and testing error (6)



Source: (Shorten and Khoshgoftaar 2019)

5. RESULTS

With multiple attempts in improving accuracy of image recognition models, the final model with an accuracy of 55.6 was submitted. The following result was achieved.

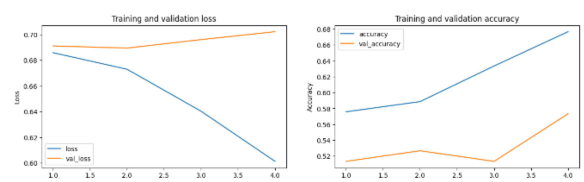


Figure 8. Training and validation

6. CONCLUSION

To conclude this paper, I have summarized the building process of a "Deep Learning Model" with technical briefing of CNN and RNN/LSTM and how they network in digital model designing. In this paper I have discussed what I applied to increase the accuracy of the model. I only managed to submit 79 entries, however, if I had more time to spend on this challenge, I would have submitted more entries by applying better evaluation matrices using precision, recall, F1 Score etc. (Korstanje 2021), to get a more accurate prediction of a model.

It was also observed that when building useful Deep Learning Models, the “validation errors” must be reduced continuously against the “training errors”.

ACKNOWLEDGEMENT

I am indebted to Dr Mohamed Reda Bouadjenek for his patience and support extended to me to complete this challenge. I also thank the prestigious Deakin University and its staff for giving me this opportunity to participate in this challenge. As a candidate for this challenge, I am also privileged to learn about modern technologies such as Machine Learning, Deep Learning, and their worldwide applications ,supporting various industries.

REFERENCES

Agrawal A, Lu J, Antol S, Mitchell M, Zitnick L, Batra D and Parikh D (27 Oct 2016), ‘VQA: Visual Question Answering’, accessed 20 August 2023.
www.visualqa.org

Amor E (20 Feb 2020), ‘4 CNN Networks Every Machine Learning Engineer Should Know’ [image], TOPBOTS, accessed 19 August 2023.
https://www.topbots.com/important-cnn-architectures/

Bouadjenek, M.R. (2023), Notebook for the Deakin Simpsons Challenge 2023, accessed 1 August 2023. Available under the Apache License, Version 2.0.
https://colab.research.google.com/github/rbouadjenek/deakin-ai-challenge2023/blob/main/deakin_ai_challenge_training.ipynb#scrollTo=xrPViXGqxqWu

IBM (2020) Recurrent Neural Networks, IBM, accessed 30 September 2022.
https://www.ibm.com/cloud/learn/recurrent-neural-networks

Korstanje J (31 Aug 2021), ‘The F1 score’, Towards Data Science, accessed 15 August 2023.
https://towardsdatascience.com/the-f1-score-bec2bbc38aa6

Kroese DP, Botev ZI, Taimre T and Vaisman, R (2020) *Data Science and Machine Learning,* CRC Press, Boca Raton, FL.

Microsoft (2017) Windows Defender ATP machine learning: Detecting new and unusual breach activity, Microsoft, accessed 27 September 2022.
https://www.microsoft.com/security/blog/2017/08/03/windows-defender-atp-machine-learning-detecting-new-and-unusual-breach-activity/

Radhakrishnan P (29 September 2017) ‘Image Captioning for Deep Learning’, Towards Data Science, accessed 1 October 2022.
https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2

Shorten C and Khoshgoftaar T M (2019) A comparison of overfitting and desired convergence [image], Journal of Big Data, accessed 02 October 2022.
https://doi.org/10.1186/s40537-019-0197-0

Shorten C and Khoshgoftaar T M (2019) ‘A survey on Image Data Augmentation for Deep Learning’, Journal of Big Data 6,60:1-48, https://doi.org/10.1186/s40537-019-0197-0