Which performs better, a Convolutional Neural Network or a Deep Neural Network for Object Recognition

Introduction

The purpose of this paper is to analyze and assess which base form neural network variation performs best for object recognition. Object recognition is a process whereby technology is capable of assessing a list of related computer vision tasks to identify objects in digital photos(Mohammad, M.S., 2021) and thus in essence, object recognition is a type of image classification (Wang, X.,2011). There are multiple machine learning methods that can be implemented for a model to do object recognition. One of them is a Deep Neural Network (DNN). This is defined as a neural network with some level of complexity. There are usually at least two layers which process data in complex ways, and this is completed by using sophisticated math modelling (Mjalli, 2007). The second model type is a Convolutional Neural Network (CNN), a form of subclass to the deep learning neural network. This is defined as a neural network designed for computing structured lists of data, such as images (Wallach, 2015).

When addressing the topic relating to who performs object recognition the best or more successfully, a convolutional neural network outperforms a deep neural network in object recognition. This is because convolutional neural networks take advantage of local spatial coherence of images (Tammina, S., 2019). This is different from a deep neural network which, being a form of super class to the convolutional neural network, lacks this ability as it treats all elements of their vector inputs equally (Singhania, 2017).

Experimenting with a deep neural network, convolutional neural networks, a fully connected neural network and a deep convolutional neural network will identify the better model. After the completion of the experiments, it was proven that by using convolutional layers in the neural networks, a higher accuracy is more likely to be returned when compared to the deep neural networks. Even after many dense layer adjustments as well as removing and adding of hidden layers, the deep neural network only benchmarked around 35% accuracy with coarse labels. On the other hand, the convolutional neural network reached past the benchmark of around 50% accuracy with coarse labels.

Methodology

In order to attempt and solve this experiment and compare a Convolutional Neural Network design and a Deep Neural Network design. My proposed method is to experiment with different neural network model designs. As seen in the notebook, there are four models displayed, with the main two models being the deep neural network model and the convolutional neural network model. The purpose of the other 2 models, a fully connected neural network and a deep convolutional neural network, is to show alterations of their corresponding models and differences between how each of them function. By viewing and comparing the results of each model's test accuracy, train time and loss results. I am able to confirm which Neural Network model variation performs best in object recognition. No feature processing was used in this experiment, as these models do not require such type of algorithm or edits to the data in order to reach the benchmark of this assessment.

Feature Processing in both models:

In the models used in this experiment, no pre-processing is used like clustering, histogram generation or dimensionality reduction. In order to achieve the desirable image classification and prediction required to solve the goal of this specific object recognition experiment, clustering is of no use or benefit. This is because the purpose of these models is to identify objects in images instead of actively solving for a mathematical or algorithmic solution. In addition, since this experiment is about comparing the performance of both in a deep neural network model and convolutional network model, there appeared to be no need to apply either histogram generation or dimensionality reduction to the train data before passing to the models, as altering the input image data beforehand would likely result in both models shifting their accuracies unnecessarily.

Feature Extraction Mechanisms in Individual Models:

The way in which both model have been designed, in the notebook, extracting features from the input imagers, in order to train themselves, is in essence the key difference between both variations of Neural Networks. A key difference which helps explain the difference in image classification relates to the layers of the models. The Deep Neural network model only possesses neuron layers, while the Convolutional Neural Network model possesses convolutional layers as well.

The Deep Neural Network Model uses a specific mechanism which extracts features of an image by looking at each pixel in the image and then matches them to a neuron. Once this is completed, the model applies all the image pixels to give a value to the data before passing into its classifier. The CNN convolutional layer is unique in its mechanism. Compared to the Deep Neural network model, the Convolutional Neural Network model works by placing filters to the input or the image. This suggests that the CNN model takes the image and applies filters which allows it to analyze the pixels of an image more intensely and hence, obtain a clearer and accurate image before classification. The CNN model scans the image from top to bottom, which allows it to change the image and pixel values in a way that makes it more detail-oriented. Therefore, in essence, the CNN model pulls out the presence of a detected object in the input (Albawi, S., 2017), and applies convolutional layers to extract features and use max pooling layers to generalize features (Yu, and Wei, Z., 2014,). Furthermore, dropouts are used to help reduce overfitting as it removes random features from the model, when training, however despite some benefits, the accuracy of image clarification may drop(Srivastava, N., 2014). After careful analysis, the CNN used in this experiment which had two convolutional layers allowed me to reach past the benchmark.

Classification for both models:

Once both the Convolution neural network and Deep Neural network have applied their feature extraction methods to their training data, they pass their results to their classifiers. In truth, both models use the same classification method. However, unlike the Deep Neural Network model, there is one more step for the CNN before classification. The CNN has to flatten its data, meaning it reduces its dimension of the data into a single dimension, in order to model the input layer and build its neural network for classification(Yeh, C.W., 2016). Once the CNN has flattened its data, both models classify in the same way by taking their extracted feature result and matching it to an appropriate label that the model tries to do based on its output layer's activation function and its neuron layer. Depending on how well both models successfully match each object with its appropriate label, thus, determines the accuracy of each model.

How both models are trained and tested:

Both models in this experiment are tested and trained on the same dataset with coarse labels instead of fine labels. This is mainly because using fine labels takes much longer for both models, and the difference in accuracy of both models does not give off a clear difference between themselves.

Results.

Table of key findings:				
Model:	Neural Network	Deep Neural Network	Convolutional Neural Network	Deep Convolutional Neural Network
Estimated Time taken to train in 15 epochs:	24 minutes	27 minutes	20 minutes	50 minutes
Test Data Accuracy Results (%):	23.13%	36.75%	50.59%	55.23%
Test Data Loss Results:	2.5519	2.239	1.6336	1.5036

As seen from the table above, each model was compared to one another under the following criteria:

1. Estimated time taken to train in 15 epochs
2. Test data accuracy results (%)
3. Test data loss results

The reason these three criteria were used was to show which type of model performs best for object recognition. It should be noted and emphasized that all models are given the same number of epochs in order for a fair comparison. Despite comparing all four models to each other, the two models emphasized in this paper are the Deep Neural Network and Convolutional Neural Network. Therefore, the main analysis will surround these two models. The other two models results in the table are being displayed in order to show how they perform against their corresponding models, which does make it easier to understand the main models.

When looking at the table, it is clear that the Convolutional Neural Network out-performs the Deep Neural Network in all three comparison criterias, and hence, is more successful at object recognition. As seen by the estimated time taken to train in 15 epochs, the CNN requires less time than the DNN. This emphasizes that the CNN is capable of knowing how to perform the task faster than the DNN. In terms of test data accuracy, the CNN is almost 20% more accurate than the DNN. This suggests that when creating the image, the CNN is more likely to have a clearer and more accurate image than the DNN. This analysis is further emphasized in Appendix 3 and Appendix 4. As seen in the two appendices and their corresponding graphs, the CNN's training accuracy curve and validation accuracy curve are very similar. This is also seen in the loss curves of the CNN where the training loss and validation loss curves are very similar. From these two curves, it is evident that the CNN, after it is trained, is able to perform similarly to what is required of them. When comparing the DNN accuracy and loss curves, it is clear that the DNN is unable to perform in the same manner that it is trained to do. The validation accuracy is way lower than the training accuracy, which suggests that despite undergoing training, it is still unable to accurately perform the way it should. Similarly, the loss that is expected after training is greater than what it should be, overfitting can be the cause for one of the reasons. Finally, the CNN is less likely to lose data compared to the DNN which further emphasizes how when obtaining the image, it is more likely to be accurate when done using the CNN model.

As seen by looking at Appendix 1 and Appendix 2 (the model's confusion matrix diagrams), we notice both models struggle to tell the difference between vehicle 1 and vehicle 2 images; however, the convolutional model performs better in knowing the difference. This is seen by the fact that Appendix 2 has clearer and more specific boxes. Furthermore, as seen by the values in each box, the CNN and its use of convolutional layers filters result in higher values which suggest a more accurate reporting.

Conclusion.

As depicted in the discussion above, it is evident that the Convolutional Neural Network out-performs the Deep Neural Network in terms of image classification. From the results in the notebook and the appendices, it is clear that the CNN is more likely to provide a more clearer and accurate image compared to any other model. When comparing accuracy, training time, and data lost, the CNN provides better results, which suggests that by using the CNN, researchers are more likely to have an accurate image in a shorter training time. The smaller loss of data suggests that the CNN is capable of reporting an image with lower data loss, which further emphasizes how the image is more likely to be accurate when done by the CNN model compared to the DNN model.

The method described in this paper provided me with the opportunity to determine which model is more accurate, however, this is done on a broader level. In order to ensure that the results are much more accurate, some more specific analysis needs to be done as well as repetitive testing of the data. Furthermore, some additional testing could include doing more complicated image classification. This is because by having more complicated images, researchers would be able to determine the extent in which either model is capable of accurate classification.

Furthermore, for a future example improvement in results for both models, applying data augmentation will allow the models to have make more accurate conclusions. This mechanism helps researchers increase the size of the dataset and introduce variability in the dataset, without actually collecting new data. The neural network treats these images as distinct images anyway. In addition, data Augmentation helps reduce over-fitting (Wong, S.C., 2016).

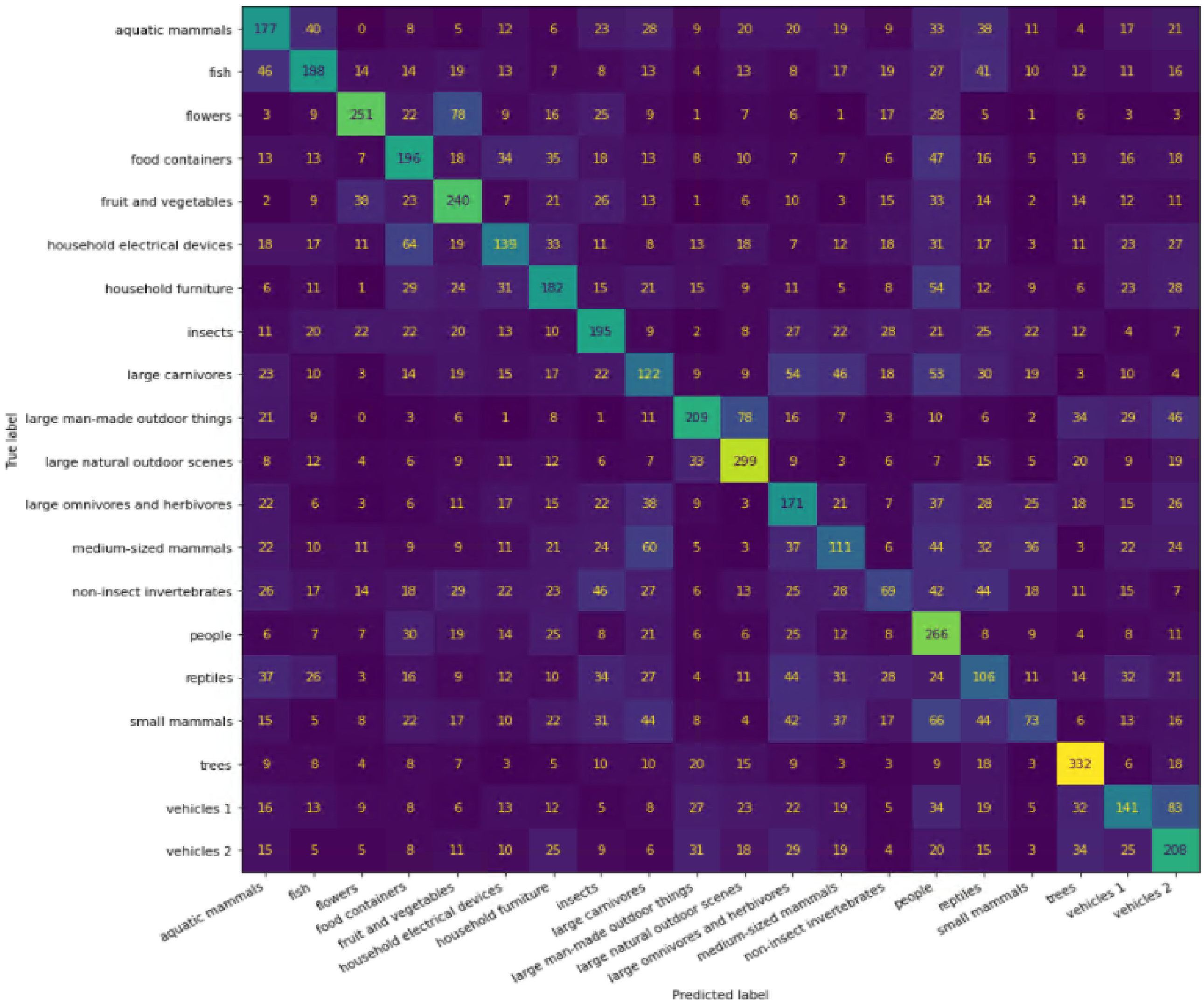
References.

Include correctly formatted references where appropriate. References are not included in the page limit.

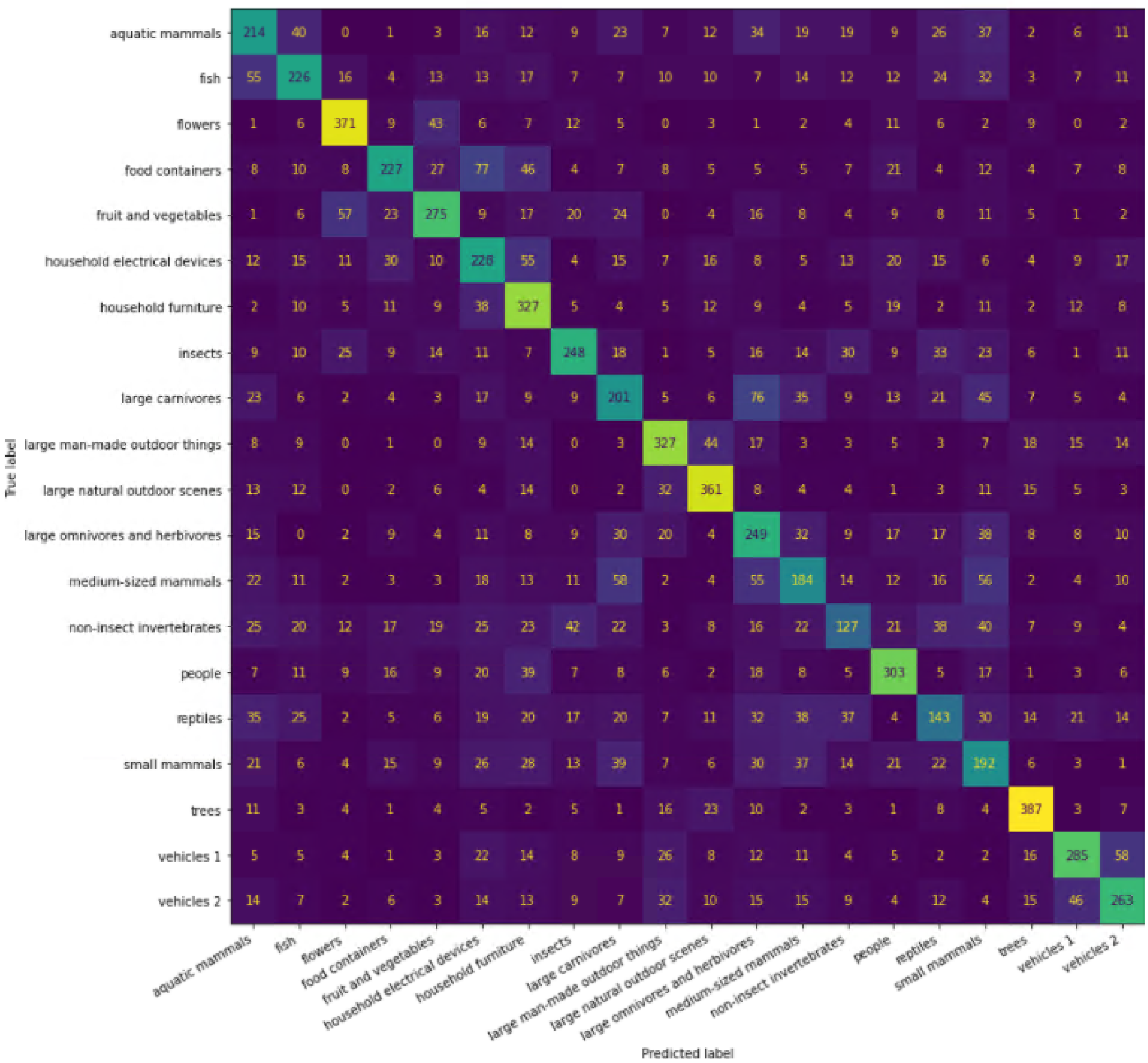
1. Mohammad, M.S., Pranitha, B., Pandula, S.G. and Sree, P.T., 2021. Object Detection with Voice Sensor and Cartoonizing the Image. *International Journal*, *10*(4).
2. Wang, X., Bai, X., Liu, W. and Latecki, L.J., 2011, June. Feature context for image classification and object detection. In *CVPR 2011* (pp. 961-968). IEEE.
3. Mjalli, F.S., Al-Asheh, S. and Alfadala, H.E., 2007. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management*, *83*(3), pp.329-338.
4. Wallach, I., Dzamba, M. and Heifets, A., 2015. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*.
5. Tammina, S., 2019. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), pp.143-150.
6. Singhania, S., Fernandez, N. and Rao, S., 2017, November. 3han: A deep neural network for fake news detection. In *International conference on neural information processing* (pp. 572-581). Springer, Cham.
7. Albawi, S., Mohammed, T.A. and Al-Zawi, S., 2017, August. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). Ieee.
8. Yu, D., Wang, H., Chen, P. and Wei, Z., 2014, October. Mixed pooling for convolutional neural networks. In *International conference on rough sets and knowledge technology* (pp. 364-375). Springer, Cham.
9. Yeh, C.W., Yeh, W.T., Hung, S.H. and Lin, C.T., 2016, October. Flattened data in convolutional neural networks: Using malware detection as case study. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems* (pp. 130-135).
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), pp.1929-1958.
11. He, J., Li, L., Xu, J. and Zheng, C., 2018. Relu deep neural networks and linear finite elements. *arXiv preprint arXiv:1807.03973*.
12. Gupta, A., Harrison, P.J., Wieslander, H., Pielawski, N., Kartasalo, K., Partel, G., Solorzano, L., Suveer, A., Klemm, A.H., Spjuth, O. and Sintorn, I.M., 2019. Deep learning in image cytometry: a review. Cytometry Part A, 95(4), pp.366-380.
13. Wong, S.C., Gatt, A., Stamatescu, V. and McDonnell, M.D., 2016, November. Understanding data augmentation for classification: when to warp?. In *2016 international conference on digital image computing: techniques and applications (DICTA)* (pp. 1-6). IEEE.

Appendix

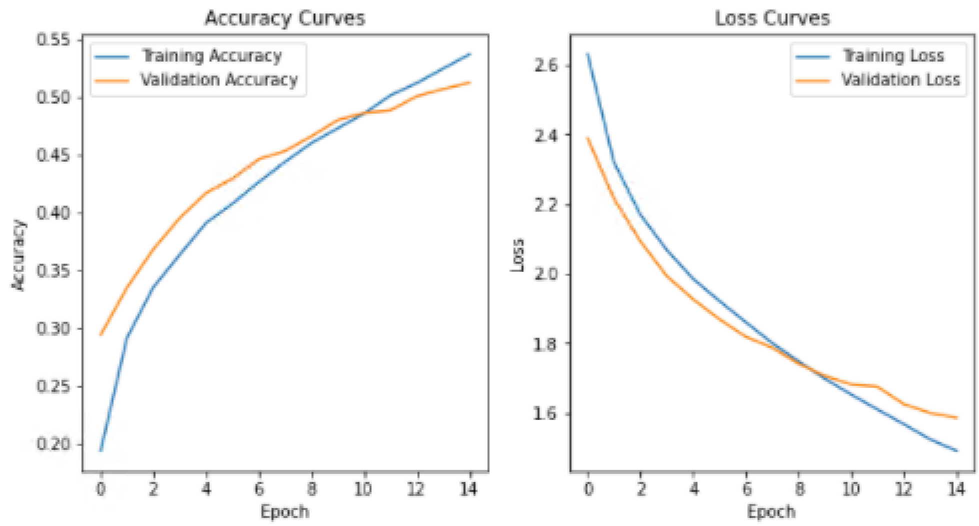
Appendix 1) Deep Neural Network Confusion Matrix Result:



Appendix 2) Convolutional Neural Network Confusion Matrix Result:



Appendix 3) Convolutional Neural Network Accuracy and Loss curves



Appendix 4) Deep Neural Network Accuracy and Loss curves