DEEP LEARNING CLASSIFICATION MODELS FOR IMAGE RECOGNITION BASED ON CIFAR-10 DATA

Apoorv Saraogee
Masters of Data Science Course 458
Artificial Intelligence and Deep Learning Section 57
November 21, 2022

Abstract

The growth of the mobile ecosystem has led to an unprecedented increase in the amount of digital imaging data. Alongside the considerable increase in computing power, deep learning neural networks are becoming an attractive option for computer vision applications in image classification. This study explores different network topologies and hyperparameters for traditional and convolutional neural networks using the CIFAR-10 dataset (Canadian Institute of Advanced Research) of 60,000 images and 10 categories. The best performing model had a testing accuracy score of 77% and was with 3 hidden convolutional layers in a stacked topology with a fully connected layer and dropout regularization. Overall, convolutional neural network models performed better than traditional neural networks suggesting the suitability of convolution for computer vision applications. However, a key drawback is the high processing time in training models with convolutional layers.

Keywords: convolution, regularization, imaging, classification, deep, network, layers, cifar-10

DEEP LEARNING CLASSIFICATION MODELS FOR IMAGE RECOGNITION BASED ON CIFAR-10 DATA

The growth of the mobile ecosystem has led to an enormous amount of digital imaging data that can be used for applications in computer vision such as autonomous vehicles, neurobiology and facial recognition. Deep learning classification models are an attractive option to be used in these applications while leveraging existing imaging data as the many layers in deep networks allow for hierarchal processing of the data (Glassner, 2021). Unlike single layer networks, deep learning networks model combinations of pixels. In deep learning networks, all of the model weights in the input layer or the properties of each pixel affect the model weights of the next layer or properties of groups of pixels. Multi-layer convolutional neural networks have become state-of-the-art in computer vision applications (Doon et al., 2019). This study explores deep learning networks with a dataset of images with 10 classifications from the Canadian Institute of Advanced Research, also known as CIFAR-10 (Krizhevsky, 2009). The study compares various network topologies and hyperparameter settings with a focus on deep neural networks and convolutional neural networks, number of layers and with/without regularization. Other research questions will include adding layers to convolutional neural networks.

Literature Review

In the early days of deep learning model development for computer vision, there were difficulties due to limited computer processing power (Chai et al., 2021). However, recently there has been rapid growth in deep learning models for image classification as computer processing limitations diminish with convolutional neural networks being the most suitable for computer vision (Chai et al., 2021). Convolutional neural networks and traditional deep learning networks have both been employed with the CIFAR-10 dataset with varying results (Calik & Demirci, 2019;

Doon et al., 2019; Jimmy Ba & Caruana, 2014; Urban et al., 2017). Although most studies showed that convolutional neural networks with deeper networks performed the best with >90% testing accuracy (Doon et al., 2019; Urban et al., 2017), deep networks have also been shown to be efficient classifiers. This study uses similar parameters as a study that showed traditional deep neural networks (stacked layers) with 2 layers that have 2000 neurons and dropout regularization performed better than convolutional neural networks (Jimmy Ba & Caruana, 2014). This study also similarly tests adding a fully connected final classification layer to convolutional neural networks which can aid in dimensionality reduction and result in higher testing accuracy (Calik & Demirci, 2019). In contrast to some other studies with more layers, this study is limited to experiments with 2 or 3 neural layers, both of which have been shown to have comparable testing accuracy to models with 4 layers (Urban et al., 2017).

Methods

This study analyzes a dataset of 60,000 colored images (32x32 pixels) curated by the Canadian Institute of Advanced Research for 10 categories (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck) also known as CIFAR-10 (Krizhevsky, 2009). Analysis is done in a Jupyter notebook using kernels for Python3 run locally with models in the sklearn, tensorflow and keras packages. The data was split with 45,000 images used for training, 5000 images used for validation and a holdout of 10,000 images for testing. There were 10 experiments run with varying network topologies, hyperparameter and regularization settings detailed in Table 1.

Table 1: Experiment details with traditional deep and convolutional neural networks

Experiment number	Description
1	2 layer deep neural network with 2000 neurons
2	3 layer deep neural network with 2000 neurons
3	2 layer deep convolutional neural network with 128, 256 neurons
4	3 layer deep convolutional neural network with 128, 256, 512 neurons
5	2 layer deep neural network with 2000 neurons and 0.3 dropout regularization
6	3 layer deep neural network with 2000 neurons and 0.3 dropout regularization
7	2 layer deep convolutional neural network with 128, 256 neurons and 0.3 dropout regularization
8	3 layer deep convolutional neural network with 128, 256, 512 neurons and 0.3 dropout regularization
9	3 layer deep convolutional neural network with 128, 256, 512 neurons and 0.6 dropout regularization
10	3 layer deep convolutional neural network with 128, 256, 512 neurons , 0.3 dropout regularization and a fully connected classification layer with 100 neurons

All experiments were trained using stochastic gradient descent with a constant number of 200 epochs, a batch size of 64 and sparse categorical cross-entropy for the loss function. All of the convolutional neural networks used a stride of 1 and kernel size of 3 for the convolutional layers and a stride of 2 and kernel size of 2 for the polling layers. All traditional neural layers used the softmax activation function while convolutional layers used the ReLU activation function.

Results for each of the experiments were analyzed with the accuracy and loss scores for the training, validation and testing sets. Visualizations included a confusion matrix and plots of accuracy/loss scores during each training epoch. Feature maps were visualized for all convolutional neural networks to check for edge recognition using the max pooling layer filters. Finally, the results from the third experiment were visualized using t-SNE chart for each category and activation values for the last classification layer method. Suitability of the method was evaluated by presence of clusters for each category.

Results

Results for each of the experiments are compared with accuracy/loss scores and process time for the training, test and validation sets in Table 2 below. It can be seen that experiment 10 (stacked 3 layer CNN, fully connected dense layer, 0.3 dropout normalization) has the highest testing accuracy score of 0.768 or ~77%. The confusion matrices and plots of accuracy/loss scores during each epoch are shown in the attached supplementary files.

Table 2: Accuracy/loss scores and process times for training/test/validation sets

Experiment number	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss	Process Time (s)
1	0.244	1.975	0.234	1.991	0.242	1.981	918
2	0.098	2.303	0.097	2.303	0.100	2.303	664
3	0.891	0.309	0.708	1.076	0.721	0.868	1635
4	0.911	0.259	0.721	0.993	0.720	0.841	3400
5	0.793	0.595	0.736	0.755	0.743	0.769	1872
6	0.100	2.303	0.097	2.303	0.239	1.989	750
7	0.768	0.670	0.724	0.794	0.741	0.750	1452
8	0.841	0.450	0.784	0.642	0.643	1.038	4746
9	0.653	1.001	0.700	0.869	0.695	0.896	6320
10	0.786	0.614	0.775	0.658	0.768	0.674	9828

Overall, the results with deep neural networks had the highest testing accuracy of 74% in experiment 5 with a shallower network and 2 hidden layers. This is in line with results in other studies using traditional neural networks which saw lower accuracy scores with more than 2 layers (Urban et al., 2017). In contrast to traditional neural networks, convolutional neural networks performed comparably with 2 and 3 layers in experiments 3 and 4 with testing accuracy scores of 72%. Notably, process time was much significantly higher in convolutional neural networks with ~10000 seconds in experiment 10 compared to ~1000 in experiment 1. Regularization improved both traditional neural networks (improvement of 50% between experiment 1 and 5) and convolutional neural networks (improvement of 2% between experiments 3 and 7). Adjusting the dropout hyperparameter in experiment 9 from 0.3 to 0.6 also resulted in a 5% increase in testing

accuracy compared to experiment 8. This suggests overfitting of the model and is evidenced by the high training accuracy scores of ~90% in experiments 3 and 4. Figure 2 below also shows negligible changes in the validation testing accuracy during training and a flat slope for validation accuracy for Experiment 3 compared to Experiment 7.

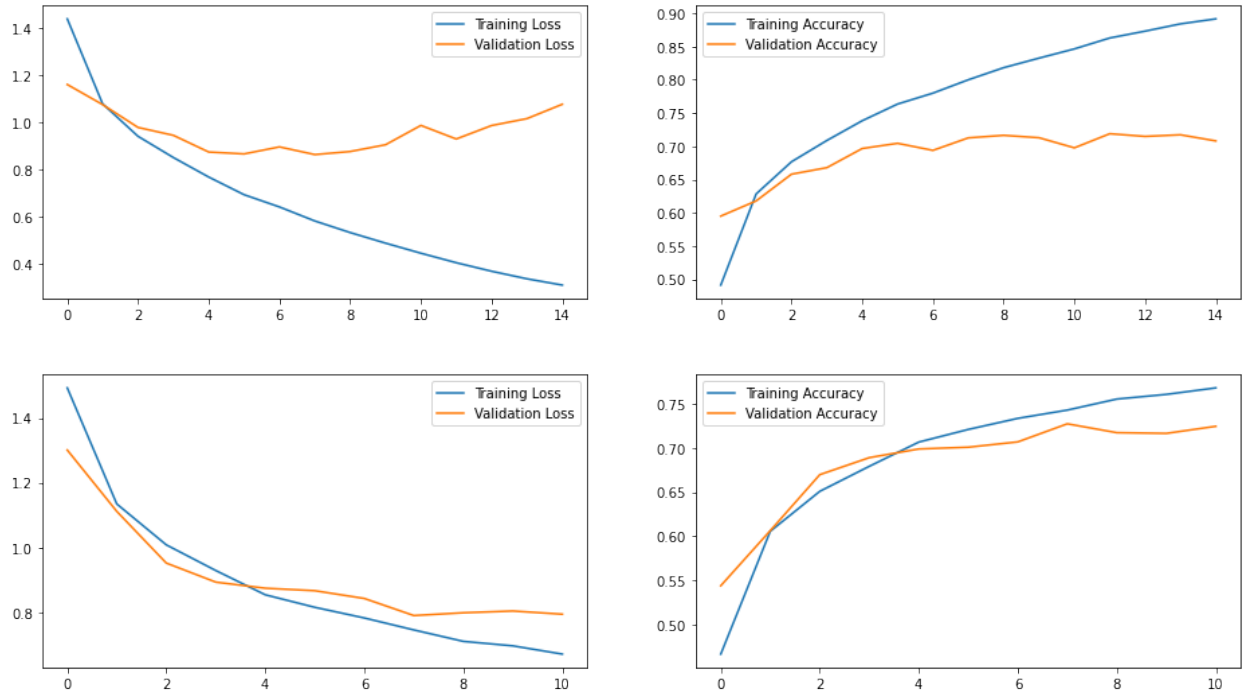

Figure 1. (top left and right) Training and validation loss/accuracy scores for Experiment 3 without regularization. (bottom left and right) Training and validation loss/accuracy scores for Experiment 7 with regularization.

Since the max pooling layers are of the same size as the input, we can show all of the pixels that light up for a particular photo in a category. The number of features is the same as the number of neurons in the layer. Visualizations with different filters or features for Experiment 3 with the activation values for each pixel are shown in Figure 2 for a bird photo which can be clearly deciphered in the pixels that 'light' up. The lit up pixels suggests that the model is learning correct features that make up a bird's face with a beak and eyes in different perspectives.

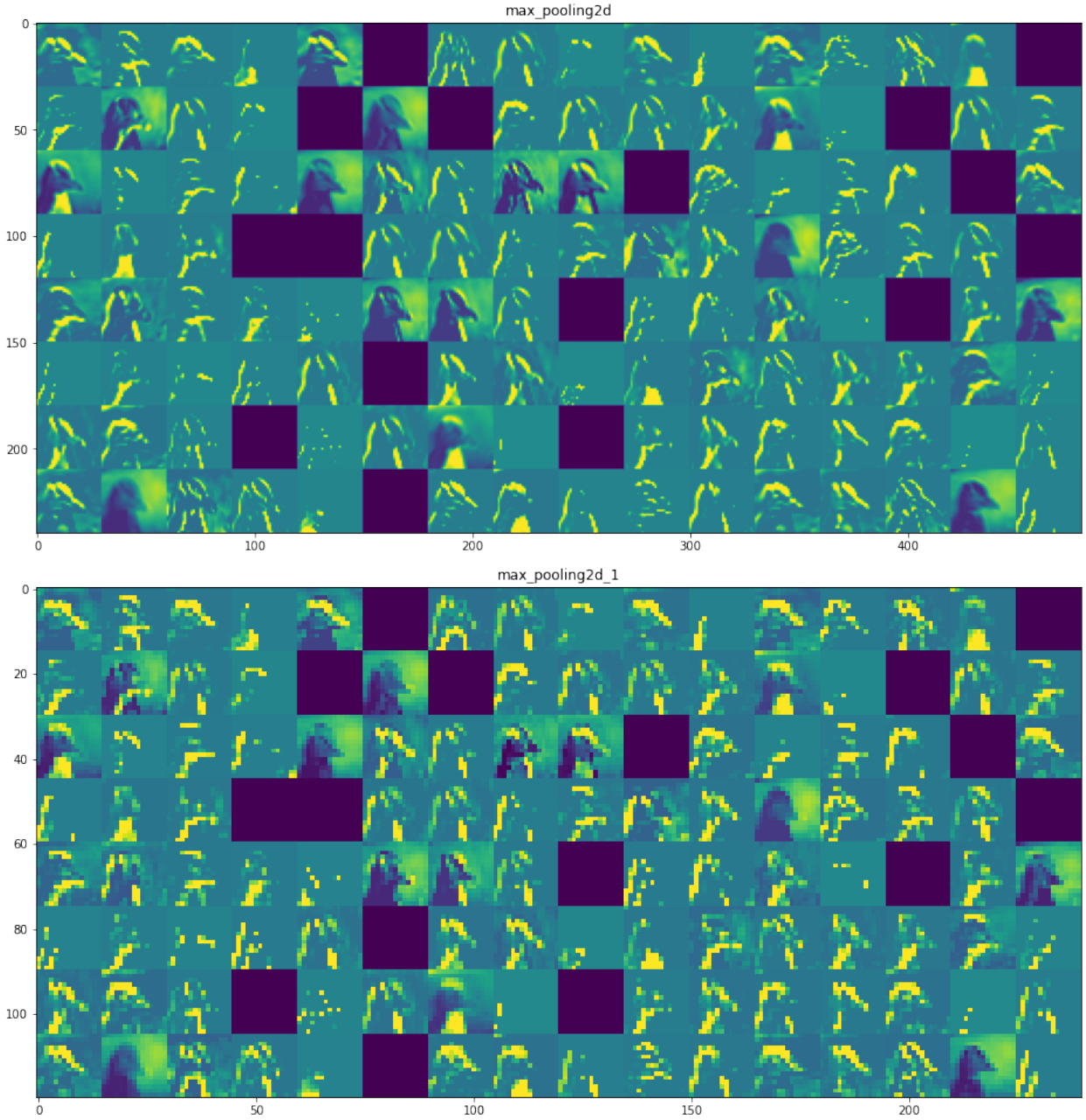

Figure 2. Visualizing feature maps for Experiment 3 max pooling layers 1 (top) and 2 (bottom) for a bird category test image.

Finally, the study also visualized the final activation values for the last dense layer using a t-SNE chart in Figure 3 and shows strong clustering of images in each category except cat/dog. Presence of clusters show suitability of the model. It is interesting to see that the cluster for cats and dogs overlap indicating high similarity in features between them. Indeed, the predictions for a test image of a dog had 6% chance of being a cat with the remaining 94% chance of being a dog.

Figure 3. Clustering map using t-SNE on activation values in the final classification layer of experiment 3 showing strong clustering in all categories except cat/dog.

Conclusions

Various deep learning networks were evaluated in this study with traditional and convolutional neural network models for the classification of images. Network topologies with 2 or 3 layers and hyperparameters with dropout regularization were varied between 10 experiments. The best performing model had 77% testing accuracy in Experiment 10 with a hyperparameter value of 0.3 for dropout regularization and a stacked network topology having a 3-layer convolutional neural network and fully connected layer. Regularization improved model performance by indicative of overfitting as a common problem in neural networks. Overall, convolutional neural networks were more robust to parameter changes compared to traditional neural networks for image classification in line with a report in the literature (Urban et al., 2017). A key drawback is the high processing time with convolutional neural networks and demand for processing power reduces scalability. However, this report also postulates that more layers are

better which was not seen in our study. More layers might indeed not make a difference depending on your network topology and hyperparameters and is reported as such in a dissenting report (Jimmy Ba & Caruana, 2014). Future studies would focus on varying more hyperparameters in Experiment 10 to achieve >90% testing accuracy for state-of-the-art performance. The model in Experiment 10 using convolutional networks could also suitable for use as a facial recognition software on mobile devices if it had higher accuracy. As there are many more categories, more than one layer may be required than as with the CIFAR-10 dataset as more complex groupings can occur with faces compared to feature differences between a car and a cat. The model can be trained in the same way as the CIFAR-10 dataset where thousands of images in a dataset are used to 'learn' the features in the pooling hidden layers and then validated and tested with unseen data.

Supplementary files

- exp1.html
- exp2.html
- exp3.html
- exp4.html
- exp5.html
- exp6.html
- exp7.html
- exp8.html
- exp9.html
- exp10.html

References

Calik, R. C., & Demirci, M. F. (2019). *Cifar-10 Image Classification with Convolutional Neural Networks For Embedded Systems*. https://doi.org/10.1109/AICCSA.2018.8612873

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*, 100134. https://doi.org/10.1016/J.MLWA.2021.100134

Doon, R., Rawat, T. K., & Gautam, S. (2019). *Cifar-10 Classification using Deep Convolutional Neural Network*. IEEE Xplore. https://doi.org/10.1109/PUNECON.2018.8745428

Glassner, A. (2021). *Deep Learning: A visual approach*. No Starch Press.

Jimmy Ba, L., & Caruana, R. (2014). Do Deep Nets Really Need to be Deep? *NIPS*.

Krizhevsky, A. (2009). *Learning Multiple Layers of Features from Tiny Images*.

Urban, G., Geras, K. J., Kahou, S. E., Aslan, O., Wang, S., Mohamed, A., Philipose, M., Richardson, M., & Caruana, R. (2017). DO DEEP CONVOLUTIONAL NETS REALLY NEED TO BE DEEP AND CONVOLUTIONAL? *ICLR*.