Deep Learning - Homework 2

Aitore Issadykova
Data Science
Nazarbayev University
Nur-Sultan, Kazakhstan
aitore.issadykova@nu.edu.kz

Assem Kussainova
Data Science
Nazarbayev University
Nur-Sultan, Kazakhstan
assem.kussainova@nu.edu.kz

***Abstract*—Convolutional Neural Networks is the widely known tool for machine learning tasks, especially for classification. The modern versions of Convolutional Neural Networks such as MobileNet and ImageNet achieve the highest performance in the Computer Vision field. This research aims to built a simple convolutional neural network and evaluate on a combined dataset of training and testing data with 4819 images from 5 classes of CIFAR-10 dataset and 5 classes of fruits dataset. Firstly, simple architecture of CNN for classification purposes is implemented and its performance is evaluated. Further, the parameters of the base network are tuned for increasing the accuracy of class prediction. As observed from the experiments, tuning the features of the convolutional network enhances the performance relatively to its basic parameters, however, most of the accuracy results are not sufficient due to the amount of the dataset provided.**

***Index Terms*—Convolutional Neural Networks, CIFAR-10 Dataset, Fruits Dataset, Deep Learning**

I. Introduction

Computer vision technologies are highly popular and widely used in the modern world. They are used for face, pedestrian, object recognition, medical analysis, autonomous car navigation and many other areas. In order to replicate human vision, a lot of researches have been done to perform image recognition with the help of computers. The latest state-of-the-art methods in image classification involve the active utilization of various deep neural network architectures. Among them, Convolutional Neural Networks (CNN) have been used in an overwhelming number of works since obtaining successful results on the ImageNet Large-Scale Visual Recognition Challenge [9]. Starting from that moment, a lot of different architectures have been developed. Most of them use deeper network structures for better performance, but CNN is still in high demand as it can efficiently solve problems as well as more serious neural networks.

In this assignment the CNN architectures performance have been evaluated on a given data from CIFAR-10 and fruits datasets. The rest of the paper is divided into the following five sections. In the Related Works section, a brief explanation of the latest works related to image classification with deep CNN is provided. Following the Dataset section contains a description of used datasets size and structure and after, in the Methods section, we propose our methodology for classifying a combination of CIFAR-10 and Fruits datasets. In the Results section, we present our experimental results for classifying the dataset and in the Discussion section, those results are analyzed and explained. In the last section of the paper, a conclusion of the work done and future suggestions in image classification of CIFAR-10 and Fruit datasets are provided.

II. Related Works

Since the publication of CIFAR-10, many articles have been published trying to achieve maximum accuracy on this dataset. On average, different convolutional neural networks with different settings and additional data preprocessing show better results. One of the early successful CNN architectures, which later began to be frequently used in subsequent studies on this dataset, was presented by He et al. [8] as a deep architecture Residual Network (ResNet) with 152 layers. They were able to gain higher accuracy by going deeper without becoming more complex by using residual blocks. In one of such studies based on the CIFAR-10 dataset, HasanPour et al. [10] proposed a new architecture with a paradigm of making the network simpler with a fewer number of parameters and faster learning. The architecture called SimpleNet, which uses 13 convolution layers with 3x3 and 1x1 kernels and 2x2 pooling kernels, beats every existing state-of-the-art in CIFAR-10 and MNIST. They used batch normalization to prevent overfitting and vanishing gradient problem. This year, the classification results of this dataset were also tried to be improved in the work of Devnath et al. [7], which deals with 8 state-of-the-art complex networks with different datasets that get optimized to reach low power at the cost of less than 0.3 percent accuracy drop, whose aim was successfully achieved. In addition, authors proposed a mathematical explanation in terms of quantization error as to why mantissa size depends mainly on the number of layers of the network.

The fruits dataset used is also included in the list of frequently used basic data in various studies since 2017 after its creation. In the past year, many successful works were released that managed to obtain high prediction accuracy using data from this dataset. Among those researches, Dang Thi Phuong Chung and Dinh Van Tai [4] provided a concise explanation of convolution neural networks (CNNs) and the EfficientNet architecture to recognize fruit using fruits dataset. From the whole dataset were selected 17624 pictures from 25 different categories. The network is built based on the EfficientNet architecture and trained for 35 epochs with a batch size of 20. Authors compared their model with the present state-of-the-art models and results were exceptional. The accuracy of the proposed model was 95.67%. Another study used a different model from the one described above, which is the work of Pande et al. [2]. The Inception V3 Model utilized in this project gives 90% accuracy on fruits dataset, which is lower than in previous approach.

III. Dataset

In this assignment, the performance of implemented models has been evaluated on a combination of two datasets: 5 classes from CIFAR-10 and 5 classes from fruits.

The first set of images have been created by Krizhevsky in 2009 [1]. It contains 60,000 color pictures of 32x32 pixels, categorized into one of ten classes: airplanes, cars, cats, deer, dogs, frogs, horses, ships and trucks. The dataset contains 6,000 pictures of each class. This dataset is considered one of the basic in machine learning: various machine learning methods are tested on it before they can be scaled. For our task, only subset of images of airplanes, birds, dogs, frogs, and horses classes were used.

The second dataset was developed by Murean and Oltean [6]. The latest version of the dataset contains images of 120 fruits and vegetables. Each image is in the size of 100x100 pixels, has a white background and stored in the database in the '.jpg' format. This dataset is relatively new and has been commonly used for image classification purposes. In this work were provided only subsets of images of apples, grapes, kiwis, lemons, and strawberries.

In total, there are 4547 images of train dataset and 272 of test images. This dataset have a relatively small size, which is not enough for more complex architectures.

IV. Methods

A. Preprocessing

1) Manual: Both datasets were provided in the separate folders and were already splitting to the train and test sets. In order to conduct an experiment with two datasets and train both data within one model, both datasets were joined under one folder. The train data folders of Fruits dataset were moved to train folder of the subset of CIFAR-10 dataset.However for the test data of Fruits dataset, it was manually created folders, which stores name of the class, and images were separated manually to these folders, then all folders were moved under the test folder of subset of CIFAR-10 dataset.

2) Preprocessing with PyTorch: As it was mentioned in Section III, the data of two datasets have different size dimensions: for CIFAR-10 subset the image size is 32x32 pixels and for Fruits dataset the image size is 100x100 pixels. For loading datasets purpose, it was created ImageFolder dataset from torchvision library by using resizing transform to resize input to size 32x32 in order to obtain identical input size. Additional transforms such as creating tensors and normalization was performed. For normalization transform, it was inserted parameters recommended by best practice methods for most of the datasets and mostly used for CIFAR-10 dataset whic equals to array [0.5, 0.5, 0.5].

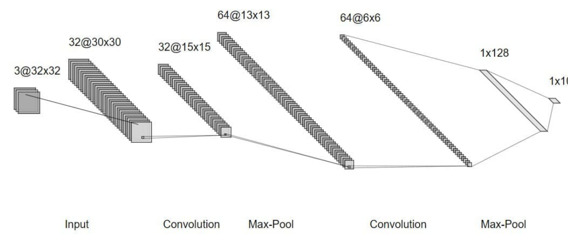

Fig. 1. Basis of CNN architecture

In order to validate the model in the future training process, it was used 80:20 split data to train and validation data with "random split" feature. The data was loaded three times as train, validation and full train which includes both train and validation data. Moreover, it was divided on the 64 batches and shuffled as it used in best practices for CNN parameters.

B. Models

1) Architecture: The CNN architecture that was used as a baseline for the experiments are shown in the Fig. 1. The basis of CNN structure consist of two Convolutional2D Layers, each of them followed by Maxpooling Layers and ends with Flattening Layer followed with 2 linear transformation functions. The initial model was user Relu activation functions after Convolution models and declared kernel size for Convolution layers as 3 and for Maxpooling the size of stride and kernel are 2. The output obtained after applying Log-Softmax activation layer.

The proposed final model based on the basis of CNN architecture. In order to diminish probability of the over-fitting model, the Dropout layer with rate 0.5 was added after second Maxpooling layer. Moreover, the final model have also some additional activation RELU function after Flattening Layer.

2) Parameters: There will be conducted several running experiments. First it will be changed the Dropout rate as well as the performance of the model will be evaluated by changing size of kernels and strides. Moreover, behaviour of results will be examined based on addition of several Relu activation functions before linear transformation as well as changing the Optimizers and Loss Functions

V. Results

The results of the several different models with Adam and SGD optimizer and different learning rates are represented in Table 1. The other setup parameters that not showed in Table 1 was already optimized by several experiments for particular optimizer and loss functions. All experiments was conducted with 20 epochs. The optimal learning rate for SGD optimizer

Model Parameters	Accuracy on Validation	Accuracy on Training
Basic Model	73%	75%
SGD Optimizer and Cross-Entropy Loss	79%	82%
Adam Optimizer with l=0.01 and Cross-Entropy Loss	75%	75%
Adam Optimizer with l=0.001 and Cross-Entropy Loss	81%	81%
Adam Optimizer with l=0.005 and Cross-Entropy Loss	79%	82%
SGD Optimizer and NLL Loss	80%	83%
Adam Optimizer with l=0.004 and NLL Loss	82%	84%

TABLE I
SOME RESULTS OF THE MODEL TESTING

was found as 0.05 and for momentum is 0.05. From different parametric setups for SGD optimizer it was found that model with same kernel sizes and strides from basis architecture as well as 0.5 Dropout proportion and single Relu activation before Linear transformation outperformed other setups.

For Adam Optimizer it was found that best learning rate is 0.04 and equal parameters for SGD Optimizer outperform other setups for Adam Optimizer. However, the main difference that for achieving highest accuracy in the testing set it is necessary to include second Relu activation function before linear transformation.

Both Adam and SGD optimizers achieve the highest accuracy on the test dataset with NLL loss function. The running time for whole training and prediction labels of testing dataset was around 262 seconds for SGD optimizer and 261 seconds for Adam Optimizer.

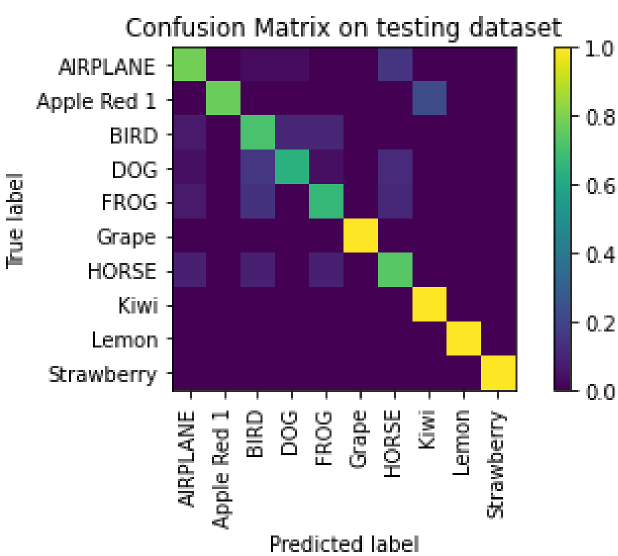

Fig. 2. Confusion Matrix of the best model

The Adam Optimizer with learning rate 0.004 and NLL loss function provides best pairs of accuracy for the validation set and testing set. Fig.2 and Fig. 3 represented the confusion matrix of the best model with Adam Optimizer and NLL function. As it can be seen from the confusion matrix, classification of fruits dataset is better than for CIFAR-10. Especially in CIFAR-10 Dataset, three classes - bird, dog and

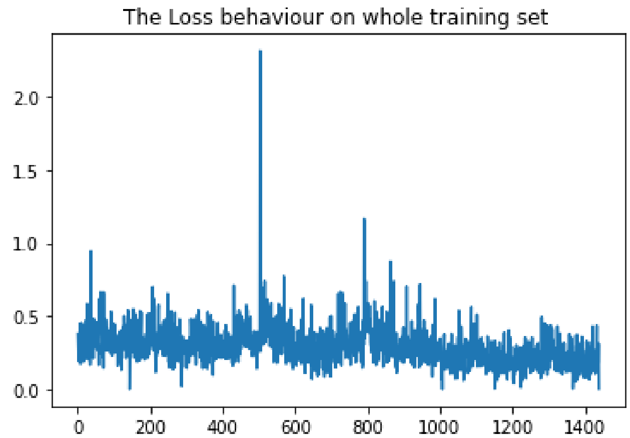

Fig. 3. Loss changes on whole dataset for best model

frog - can be misclassified between each other. The losses graphs provide the value of loss in each batch for 20 epochs. It can be observed that there are some peaks in losses for dataset and it is possibly due to overfitting issue on the datasets.

VI. DISCUSSION

The highest accuracy achieved by proposed model is 84% and this accuracy score was also obtained on the CIFAR-10 Dataset by using CFNN [3] for classification. The accuracy on CIFAR-10 part of the dataset may be improved by performing additional transformations for the input images in order to extract significant features. The running time of the proposed model is near to 7 minutes and the model can be speed up by using GPU or TPU accelerator and cuda architecture from PyTorch. In this study, cuda was not available in the Google Colab, as well as GPU was lowering speed of inserting data.

For the future studies it is recommended to use more data from both Datasets in order to improve accuracy of the model. Moreover, the analysis of the size of each classes and if the data is imbalanced, provide some weights correction or new sampling techniques. In addition, the performance of the well-known CNNs such as ResNet, AlexNet and ImageNet can be compared in both accuracy and running time with the proposed model.

VII. CONCLUSION

The classification of images is the one of the popular topics in Computer Vision and Deep Learning. There are several well-known Convolutional Neural Networks that is widely used in different spheres. This assignment provided some overview on the existing studies of classification of the CIFAR-10 Dataset and Fruits 360 Dataset. The current stusy used the basic Convolutional Neural Network and tuned parameters to achieve the highest performance. The best model results are comparable with results of CFNN neural network which was used for classification of the CIFAR-10 dataset [3]. As a limitation of this study , it should be mentioned the technical issues with CUDA and accelerators of the Colab environment and lack of time to provide some data analysis for the distribution of each data class. It is recommended to conduct data distribution analysis for the future experiments and provide some performance, i.e accuracy and runtime comparison of the proposed model with widely known ResNet, AlexNet and ImageNet neural networks.

REFERENCES

[1] A. Krizhevsky, “Learning multiple layers of features from tiny images,” Tech. Rep., 2009.

[2] A. Pande, M. Munot, R. Sreeemathy and R. V. Bakare, ”An Efficient Approach to Fruit Classification and Grading using Deep Convolutional Neural Network,” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-7, doi: 10.1109/I2CT45611.2019.9033957.

[3] Borisov, V. Korshunova, K.. (2019). Multiclass Classification Based on the Convolutional Fuzzy Neural Networks. 10.1007/978-3-030-30763-9-19.

[4] Chung, Dang Tai, Dinh. (2019). A fruits recognition system based on a modern deep learning technique. Journal of Physics: Conference Series. 1327. 012050. 10.1088/1742-6596/1327/1/012050.

[5] Devnath, J. K., Surana, N., Mekie, J. “A Mathematical Approach Towards Quantization of Floating Point Weights in Low Power Neural Networks”, In 2020 33rd International Conference on VLSI Design and 2020 19^{th} International Conference on Embedded Systems (VLSID) (pp. 177-182). IEEE, Jan 2020

[6] H. Murean and M. Oltean, “Fruit recognition from images using deep learning,” vol. 10, pp. 26–42, 06 2018.

[7] J. K. Devnath, N. Surana and J. Mekie, ”A Mathematical Approach Towards Quantization of Floating Point Weights in Low Power Neural Networks,” 2020 33rd International Conference on VLSI Design and 2020 19th International Conference on Embedded Systems (VLSID), Bangalore, India, 2020, pp. 177-182, doi: 10.1109/VLSID49098.2020.00048.

[8] K. He, X. Zhang, S. Ren and J. Sun, ”Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, Dec 2015. [Online]. Available: https://doi.org/10.1007/s11263-015-0816-y

[10] S. H. HasanPour, M. Rouhani, M. Fayyaz, and M. Sabokrou, “Lets keep it simple, using simple architectures to outperform deeper and more complex architectures,” CoRR, vol. abs/1608.06037, 2016. [Online]. Available: http://arxiv.org/abs/1608.06037