Deep Learning 2023 - Project I plan

Bartłomiej Sobieski, Władysław Olejnik

February 2023

1 Project plan

Project I of the *Deep Learning* course is concerned with the application of convolutional neural networks (CNNs) to the image classification task on the *CIFAR-10*[3] dataset. In recent years, this type of neural networks has achieved state-of-the-art results in many complex tasks ranging from general computer vision problems (beginning with the famous *AlexNet* [4]) through achieving superhuman performance in chess and its variants [5], proving to this day it's general applicability and high performance in comparison to e.g. classical dense neural networks. Recently, a shift in the domain of computer vision based on deep neural networks has been observed due to the introduction of a *Transformer*-based [7] architecture (initially introduced in the context of natural language processing) called *Vision Transformer* (*ViT*) [2]. This family of neural networks outperforms convolutional neural networks under conditions of very large number of parameters and data. This transition, from dense neural networks to CNNs to visual transformers, is the motivation behind the experiments that we want to conduct.

1.1 Experiment I

In this experiment, we will limit ourselves to using only dense networks on flattened images from the dataset. Our comparison will include at least 3 networks with different sizes. As mentioned in the requirements for this project, we will test the performance of this network using different hyperparameters related to the training process, regularization and data augmentation techniques. This experiment will serve as a baseline for more sophisticated architectures from other experiments.

1.2 Experiment II

In this experiment, our main focus will be to test the performance of convolutional neural networks with manually designed architecture and varying size. This experiment will also include testing different hyperparameters related to the training process, regularization and data augmentation techniques. The aim of this experiment is to maximize the performance of convolutional neural networks trained from scratch and check how it scales with an increase in the number of parameters used.

1.3 Experiment III

In this experiment, we will manually implement a smaller version of *ViT*. This experiment, just like the previous ones, will include testing different hyperparameters related to the training process, regularization and data augmentation techniques. The goal of this experiment is to see how *ViT* performs in a scenario of small number of parameters and images in the dataset.

1.4 Experiment IV

In this experiment, we will focus on fine-tuning famous pretrained architectures such as *ResNet* [1] or *EfficientNet* [6]. Because these models achieve high out-of-the-box performance and have been trained on more powerful hardware than we have at our disposal, they will serve as an upper limit to which we can refer.

2 Remarks

This is our initial idea of the experiments we want to carry out, so it is difficult to include some specific details, such as the architecture that we will use in experiment II.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[2] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. An image is worth 16x16 words: Transformers for image recognition at scale. 2021.

[3] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[5] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. Mastering chess and shogi by self-play with a general reinforcement learning algorithm, 2017.

[6] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. 2019.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.