Optimization of deep learning network architectures for hyperspectral data classification

Kamil Książek

Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
44-100 Gliwice, Poland

Department of Data Sciences and Engineering
Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology
44-100 Gliwice, Poland

kksiazek@iitis.pl, kamil.ksiazek@polsl.pl

Supervisor: dr hab. inż. Przemysław Głomb
Co-supervisor: dr Krisztián Búza

Abstract

In recent years, deep learning has achieved undisputed success in many domains, including the classification of hyperspectral images. These images are acquired by hyperspectral sensors that capture data in hundreds or thousands of narrow channels per pixel from various ranges of the electromagnetic spectrum. They measure energy from the region of visible light, the near-infrared, the short-wave infrared or even the long-wave infrared. Hyperspectral images have various applications in non-invasive substance classification, including geology, precision agriculture, environmental monitoring, hydrology and military science.

In this work, we focus on the problem of optimization of deep neural networks applied to the classification of hyperspectral data. Due to the variety and complexity of neural architectures, the network configuration for a particular problem poses a challenge. We chose the classification of blood and blood-like substances on a dataset of hyperspectral images as an

experimental setting for the neural network optimization task. We applied several state-of-the-art architectures, i.e. one-, two- and three-dimensional convolutional neural networks, a recurrent network based on GRU units and a multilayer perceptron. We designed two different scenarios of experiments, i.e. a commonly used *Hyperspectral Transductive Classification* (HTC) scenario in which samples from both a training set and a test set come from one image and a more realistic *Hyperspectral Inductive Classification* (HIC) scenario in which a test set comes from a different source than a training set. In the HIC scenario, we also evaluated the influence of diverse background substances and varying days of image capture on network performance. We observed that for the HTC scenario, the classification accuracy of all methods tested exceeded 90%, while for the HIC scenario, the performance was much lower and varied between 57.2 and 99.5%. Furthermore, in some cases, more sophisticated methods such as convolutional neural networks were less efficient than the simplest feedforward architecture. We also identified a problem with the network stability for one of the architectures used in the experiments. The problem was more investigated in the following chapters and led to the study about vanishing gradients and, finally, to network reinitialization methods.

One of the main limitations of network performance in the HIC scenario is the presence of pixels containing a mixture of spectra of several substances. Therefore, for further studies, we selected a branch of neural network architectures, i.e. linear autoencoders. They are trained without labels, and, despite the simplicity, they can achieve competitive results. We conducted a series of hyperspectral unmixing experiments in which original spectra of mixed pixels are being reconstructed. We observed that, in some cases, trained networks with a given set of hyperparameters achieve different levels of performance. We prepared an extensive statistical study that confirmed the important impact of weight initialization on the final network reconstruction error. Furthermore, we identified a problem of dead activations in networks using the ReLU activation function, related to the large number of inactive neurons. This phenomenon makes in some situations training of such networks difficult or even impossible. We proposed three network reinitialization methods that alleviate the negative consequences of dead activations. Based on the threshold for ratios of dead activations, we reinitialize a subset or all network weights, depending on the selected reinitialization method. We confirm that for experiments with hyperspectral data, the proposed methods increase network performance. Finally, we evaluated the robustness of these approaches using the MNIST dataset. The observed results lead to the conclusion that network reinitialization methods are an effective solution and reduce network error, especially for suboptimal sets of hyperparameters.

The following dissertation contributes to the problem of optimization of deep learning networks for hyperspectral data classification through an extensive study of different architectures. Furthermore, the problem of network stability was discussed and three network reinitialization methods were proposed to mitigate the identified dead activation phenomenon.