Exercise Sheet 6

Neural Computation
University of Birmingham

Week 6 2025

Exercise 1. You are designing a convolutional autoencoder for images and trying different architectures (similar to a UNet), using an encoder with multiple layers of convolutions and pooling, and a decoder using upsampling and convolutions.

1. You find that a deeper network (more pooling and upsampling operations) paradoxically increases the loss on the training set and validation set. The best loss is achieved by an architecture that uses no pooling and upsampling at all, but only convolutions. Why could this be the case?

2. In order to improve the reconstruction loss for deeper networks, you decide to introduce skip-connections between the encoder and decoder (like in a UNet). What should you consider?

Solution:

1. The autoencoder tries to learn the identity; it is therefore not always desirable to find an architecture with the lowest reconstruction loss, as this may indicate that the bottleneck is too large and the autoencoder does not produce meaningful codes. If there are no pooling operations and only convolutions, the network will very likely be able to just implement the identity function.

2. Introducing skip-connections will lead to perfect reconstruction loss but at the same time prevent the network from learning a meaningful representation of the data. Instead, the information for all pixels will be directly passed to the output.

Exercise 2. You want to train and test an image classifier that can diagnose a type of liver disease from an ultrasound scan. You can find three public datasets, which you want to use to train and test your neural network. The data from hospital A comes from routine liver scans and only about 5% of the patients have the disease. The data from hospital B comes from a study focussing on the liver disease and 95% of patients in the dataset suffer from the disease. The data from hospital C comes from another targeted study and includes roughly 50% of both; positive and negative patients. All three hospitals use different scanning equipment from different manufacturers. To avoid overfitting and use as much training data as possible, you combine all three datasets together and shuffle them, hoping that this will enable the network to generalise over the different scanner types.

Part A: Is this a good approach? What do you have to consider to make sure that the classifier will work reliably with data from all three scanners?

Part B: Consider the case where the task is not image classification, but semantic segmentation with the goal of segmenting the liver, which can change its shape when affected by the disease. Do we have to consider the same issues as for classification?

Solution:

Part A:

We have to consider the Clever Hans effect in this situation.

Because the datasets come from different scanners, they may contain different types of imaging artefacts that the network could learn to recognise. For example, the network might learn

that images from hospital A are generally more likely to be negative, and images from hospital B are more likely to be positive. This way, the network could achieve a good training and validation loss and accuracy without actually learning to recognise the disease. Instead, it would simply learn to recognise the type of scanner or hospital source.

When such a system is deployed, it could produce distorted results based on which scanner was used, rather than detect the actual presence of disease.

To address this issue, one approach would be to either use only dataset C, which has a balanced distribution of positive and negative cases, or to combine the data in such a way that each hospital contributes an equal number of positive and negative examples. This could be done by randomly disregarding negative examples from dataset A and positive examples from dataset B. Alternatively, oversampling (i.e. using multiple copies of) the minority class examples in datasets A and B could help balance the training data.

Other options to consider are:

- Using a balanced validation set, ideally from hospital C, to ensure that the model is evaluated on data that is not biased by scanner type or class imbalance.
- It could be possible to preprocess the images to minimise scanner-specific artefacts. This may include normalising image resolution, contrast, and intensity across datasets, and removing metadata that could leak hospital identity.

Part B:

In the case of semantic segmentation, the goal is to accurately delineate the liver in each image. The same problem of the clever Hans effect could still potentially apply. Based on the specific imaging artefacts in dataset A or B, the network could learn to prefer certain liver shapes over others in difficult images. The same mitigation strategies would be applicable.