Deep Learning

Project 1 – Convolutional Neural Networks (CNNs)- Plan

Adam Majczyk 313420, Szymon Matuszewski 313434

Data to be used

LINK: https://www.kaggle.com/datasets/eclaircat/cinic-eclair-test/

Train Set

The train set contains 90000 JPG pictures, size of 32x32 pixels in 8-bit colour space. The pictures belong to one of 10 categories: truck, ship, horse, frog, dog, deer, cat, bird, automobile, airplane.

Test Set

The test set contains 90000 JPG pictures, size of 32x32 pixels in 8-bit colour space. They are unlabelled.

Planned architectures of CNNs

The architectures which will be tested will involve testing networks with varying numbers of:

1. Convolution layers (1-3 layers; few since the images are only 32x32 pixels).
2. Fully connected layers (1-4 layers).
3. Neurons per hidden layer in the fully connected layer (1-20 neurons).

We also plan to test:

1. Different kernel sizes (2x2-10x10), also different combinations of sizes, when using more than one convolution layer.
2. Different optimizers in the fully connected layer.
3. Different activation functions in the fully connected layer.
4. Max, median, mode and mean pooling.

Aside from the classic CNN architecture we plan to test pretrained models, such as AlexNet, VGG, ResNet50. We will also attempt to fine tune them for our task – by the means of retraining the fully connected layers (the whole ANN vs only the last hidden layer).

Planned tests

Testing will cover all the aspects of the project including:

1. Different architectures:
    - Evaluating the classification metrics on the test dataset, including precision, recall, auc, balanced accuracy and more.
    - Checking the training time
2. Augmentations techniques:
    - Evaluating the classification metrics on test datasets containing only the augmented images from examined augmentation group so as to decide which augmentations should be applied to improve the resilience on changes.

- Augmentation group – a combination of augmentations. The python package **albumentations** will be used here.
- Planned augmentations to be examined:
  Crop - crop region from image.
  GridDistortion - an image warping technique which is driven by the mapping between equivalent families of curves, arranged in a grid structure.
  Perspective - perform a random four-point perspective transform of the input.
  PixelDropout – perform a random dropout for pixels of the image.
  RandomScale – randomly scale the image.
  CLAHE – perform the CLAHE (Contrast Limited Adaptive Histogram Equalization) transformation of the image.
  GaussianBlur - Blur the input image using a Gaussian filter with a random kernel size.
  Rotation – rotate the image.
  Mirror – mirror the image.
  Colouring – apply a black/white, red, blue, green etc. colour mask.
  These can change as an outcome of the data exploration process.

3. Comparing results with pre-trained models:
   - The classification metrics on models AlexNet, VGG, ResNet50 will be evaluated and then compared to our models with different architectures and chosen augmentations.
4. Comparing simple-architectured CNNs with a committee of CNNs based on classification metrics.