Machine Learning
Università della Svizzera italiana

Assignment 2

Morale Mariciano Jeferson

June 6, 2024

Please refer to the **ReadMe** file to get the exact all questions. This is just a template of how your report should look like. In this assignment, you are asked to:

1. Implement a fully connected feed-forward neural network to classify images from the **Cats of the Wild** dataset.
2. Implement a convolutional neural network to classify images of **Cats of the Wild** dataset.
3. Implement transfer learning.

Both requests are very similar to what we have seen during the labs. However, you are required to follow **exactly** the assignment's specifications.

1 Image Classification with Fully Connected Feed Forward Neural Networks (FFNN)

In this task, you will try and build a classifier for the provided dataset. This task, you will build a classic Feed Forward Neural Network.

1. Download and load the dataset using the following link. The dataset consist of 7 classes with a folder for each class images. The classes are 'CHEETAH' ,'OCELOT', 'SNOW LEOPARD', 'CARACAL', 'LIONS', 'PUMA', 'TIGER'. Check Cell 1 in 'example.ipynb' to find the ready and implemented function to load the dataset.
2. Preprocess the data: normalize each pixel of each channel so that the range is [0, 1].
3. One hot encode the labels (the y variable).
4. Flatten the images into 1D vectors. You can achieve that by using [torch.reshape] or by prepending a [Flatten layer] to your architecture; if you follow this approach this layer will not count for the rules at point 5.

5. Build a Feed Forward Neural Network of your choice, following these constraints:
    - Use only torch nn.Linear layers.
    - Use no more than 3 layers, considering also the output one.
    - Use ReLU activation for all layers other than the output one.
6. Draw a plot with epochs on the x-axis and with two graphs: the train accuracy and the validation accuracy (remember to add a legend to distinguish the two graphs!).
7. Assess and comment on the performances of the network on your test set, and provide an estimate of the classification accuracy that you expect on new and unseen images.
8. **Bonus** (Optional) Train your architecture of choice (you are allowed to change the input layer dimensionality!) following the same procedure as above, but, instead of the flattened images, use any feature of your choice as input. You can think of these extracted features as a conceptual equivalent of the Polynomial Features you saw in Regression problems, where the input data were 1D vectors. Remember that images are just 3D tensors (HxWxC) where the first two dimensions are the Height and Width of the image and the last dimension represents the channels (usually 3 for RGB images, one for red, one for green and one for blue). You can compute functions of these data as you would for any multi-dimensional array. A few examples of features that can be extracted from images are:
    - Mean and variance over the whole image.
    - Mean and variance for each channel.
    - Max and min values over the whole image.
    - Max and min values for each channel.
    - Ratios between statistics of different channels (e.g. Max Red / Max Blue)
    - Image Histogram (Can be compute directly by temporarely converting to numpy arrays and using np.histogram)

    But you can use anything that you think may carry useful information to classify an image.

    N.B. If you carry out point 7 also consider the obtained model and results in the discussion of point 6.

Reasoning

Every answer provided first depict the overall plain result. Then, prompts the analytical data that lead to such result. In addition, a theoretical motivation for the outcomes is given. Finally, some metaphors are provided by me for the best I could approximate such concepts to how humans behave and interact with nature, in order to ease the understanding of these concepts.

Apparatus and Configuration

The jupyter notebook was developed using the following configurations in order to produce the most deterministic outcome at every new run.
Apparatus description:

- **CPU** Ryzen 3900XT
- **GPU** RTX 3070 8GB VRAM
- **RAM** 16 GB
- **OS** Windows 11 Pro 23H2

The first scratch of the assigment was developed on Kaggle, were the only hardware specs provided were a python environment in a notebook with a NVIDIA P100 GPU dedicated to the task, CUDA enabled during training.
the configuration description for the training and assignment overall are:

- **seed** 20020309
- **learning rate** 0.001
- **batch size**
- **cuda** enabled
- **epochs** 101
- **loss function** cross entropy
- **optimizer** Adam

The reason of using CrossEntropy as **loss function** from pytorch is that the module implements both CrossEntropy and SoftMax at every layer but the last output layer, which is exactly the desired behavior.
The training loop implemented follows a similar pattern provided by PyTorch documentation.

6

In Figure (**??**), I provide the plot with epochs on the x-axis where two graphs are plotted: the train and test accuracies.

7

Overall, after assessing the performances of the network on the unseen validation set the model is realiable in estimating the classification of the images with a best accuracy slightly lower XXX than the one on the test set, and the average of the accuracies are around XXX, with a variance of XXX.
The delusional result is due to the nature of Feed Forward Neural Network: the learning is focused on pixel per pixel, while a desired and more useful approach would be the ones resembling human interaction, i.e. a broad overview of the whole image, analyzing it at chunks.
As it is difficult for us to spot an image from a puzzle piece, the FFNN has a metaphorically resembling trouble.
For us humans, if we might be able to immediately recognize a person (macro-concept) from its smile (micro-concept). A similar concept is the key to improve the results and lead to the CNN model approach in Task 2.

8 BONUS

The bonus was completed using the mean and average for each RGB channel. The results thought show a similar slightly lower accuracy value during training. The validation results highlighting XXX with variance of XXX. I would have like to put more features in order to get a better model.
The statistical comparison with the FFNN model without features yield the following results: there XXXXX statistical significantly difference with the $p - value = 0.05$. The best result is yild by model XXXXX with best accuracy XXXX and variance of XXXX.
It suffers from the same disadvantages from FFNN discussed previously, so the improvement is sadly bounded by the model nature.

2 Image Classification with Convolutional Neural Networks (CNN)

Implement a multi-class classifier (CNN model) to identify the class of the images: 'CHEETAH' ,'OCELOT', 'SNOW LEOPARD', 'CARACAL', 'LIONS', 'PUMA', 'TIGER'.

1. Follow steps 1 and 2 from T1 to prepare the data.
2. Build a CNN of your choice, following these constraints:
   - use 3 convolutional layers.
   - use 3 pooling layers.
   - use 3 dense layers (output layer included).
3. Train and validate your model. Choose the right optimizer and loss function.
4. Follow steps 5 and 6 of T1 to assess performance.
5. Qualitatively and **statistically** compare the results obtained in T1 with the ones obtained in T2. Explain what you think the motivations for the difference in performance may be.
6. **Bonus** (Optional) Tune the model hyper-parameters with a **grid search** to improve the performances (if feasible).
   - Perform a grid search on the chosen ranges based on hold-out cross-validation in the training set and identify the most promising hyper-parameter setup.
   - Compare the accuracy on the test set achieved by the most promising configuration with that of the model obtained in point 4. Are the accuracy levels **statistically** different?

3

The following train parameters were chosen for the model. The Optmizer chosen is XXXX because XXXX. The loss function used is the CrossEntropyLoss function, which fits the categorical nature of the classification problem.

4

Overall, after assessing the performances of the network on the unseen validation set the model is realiable in estimating the classification of the images with a best accuracy slightly lower XXX than the one on the test set, and the average of the accuracies are around XXX, with a variance of XXX.

5

The statistical comparison between the simple FFNN and the current CNN model yield the following results: there XXXXX statistical significantly difference with the $p - value = 0.05$. The best result is yild by model XXXXX with best accuracy XXXX and variance of XXXX.

6 BONUS

For the grid search I choose to change the following hyperparameters:

1. Learning rate: [0.001, 0.01, 0.1]
2. Batch size: [32, 64, 128]

The epochs were fixed at 100 as defined in configuration.
The results of such grid search yield me a model with peak test accuracy of 62.09% with the following best parameters:

- batch size = 64
- learning rate = 0.001
- epochs = 100

3 Transfer Learning

This task involves loading the VGG19 model from PyTorch, applying transfer learning, and experimenting with different model cuts. The VGG19 architecture have 19 layers grouped into 5 blocks, comprising 16 convolutional layers followed by 3 fully-connected layers. Its success in achieving strong performance on various image classification benchmarks makes it a well-known model.
Your task is to apply transfer learning with a pre-trained VGG19 model. A code snippet that loads the VGG19 model from PyTorch is provided. You'll be responsible for completing the remaining code sections (marked as TODO). Specifically:

1. The provided code snippet sets param.requires_grad = False for the pre-trained VGG19 model's parameters. Can you explain the purpose of this step in the context of transfer learning and fine-tuning? Will the weights of the pre-trained VGG19 model be updated during transfer learning training?

2. We want to transfer learning with a pre-trained VGG19 model for our specific classification task. The code has sections for __init__ and forward functions but needs to be completed to incorporate two different "cuts" from the VGG19 architecture. After each cut, additional linear layers are needed for classification (similar to Block 6 of VGG19). Implement the __init__ and forward functions to accommodate these two cuts:

   - This cut should take the pre-trained layers up to and including the 11th convolution layer (Block 4).
   - Cut 2: This cut should use all the convolutional layers from the pre-trained VGG19 model (up to Block 5).

     Note after each cut take the activation function and the pooling layer associated with the convolution layer on the cut

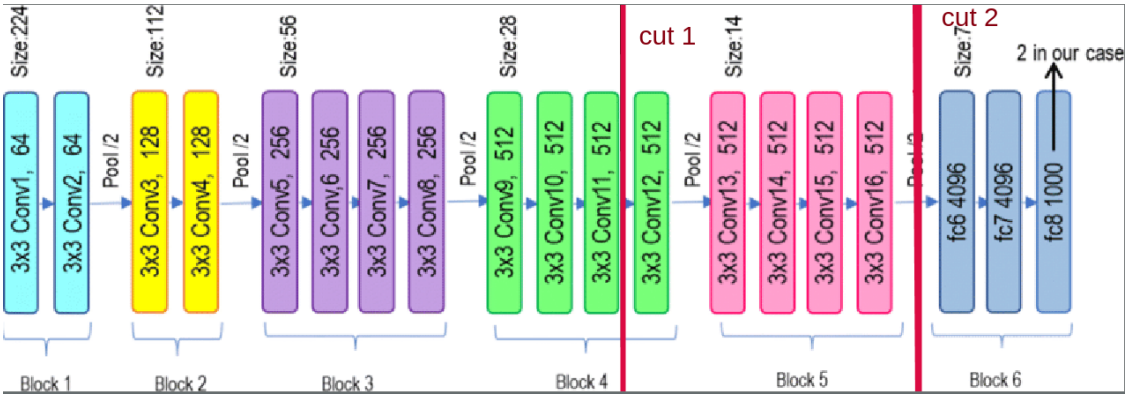

Figure 1: Cuts in VGG19

3. In both cases, after the cut, add a sequence of layers (of your choice) with appropriate activation functions, leading to a final output layer with the desired number of neurons for your classification task. Train the two models (one with Cut 1 and another with Cut 2) on your chosen dataset. Once training is complete, compare their performance statistically.

4. Based on the performance comparison, discuss any observed differences between the two models. What could be the potential reasons behind these results?

5. BONUS (optional): Try different cuts in each block of VGG19, and plot one single figure with all the train-validation-test accuracies. Explain in detail the reasons behind the variation of results you get.

1

Within the context of transfer learning, the model VGG19 from PyTorch is loaded with the architecture and weights of a pre-trained model. The purpose of setting `param.requires_grad =False` is to **freeze the weights** of the model, so that the model does not update the weights during the training of the new model. In fact, it would not make sense to update the weights of the pre-trained model in the context of transfer learning. Particularly, transfer learning is a tecnique where a model for one task, in this case the VGG19 for image classification, is reused as the starting point for a model on a second task, where the fine tuning to train the model over our feline dataset happens. Hence, during transfer learning traning, the weights of the pre-trained VGG19 will XXX be updated.

2

To select the first cut, we need to from the firt feature children of the VGG19 architecture to index 25, in order to get the 11th convolution layer and its activation function and pooling layers associated. The correctness can be easily checked thanks to the already templated `print(self.features)` after the cut, so you will see all the requirements satisfied.
For the second cut, we trivially load all the features from the model.

3

For fine tuning we put also a sequence of 3 linear layers with relu activation function complementing with a dropout strategy to try to avoid overfitting. The number of neuron were 128 in both input and hidden, ending with an output correspoding to our cat types in the dataset, i.e. 6.
The statistical comparison between the cut1 and the cut2 pre-trained and fine-tuned models yield the following results: there XXXXX statistical significantly difference with the $p - value = 0.05$. The best result is yield by model XXXXX with best accuracy XXXX and variance of XXXX.

4

Based on the performance comparison, the reason why the cut 2 performs significantly better than cut 1 lies within the transfer learning of the VGG19 itself: its architecture is meant to be used XXXX.
Hence, the potential reason of the lower accuracy for cut 1 woud be indeed in the difference of XXX.

5 BONUS

For the bonus part, I arbitrarly chose the following different cuts among the blocks of the VGG19 model:

- **Cut A** XXX

- **Cut B** XXX
- **Cut C** XXX

In Figure (2), the result plot of the train-test-validation of accuracies is showcased.

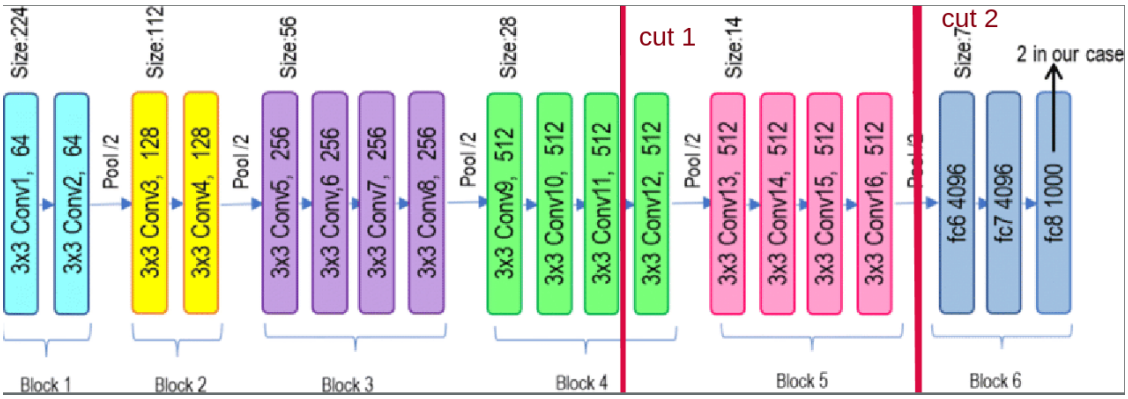

Figure 2: train-test-validation accuracies plot

The reason behind such variation of results leads to a stronger support of the previously mentioned difference between cut1 and cut2: XXXXXX.