Deep Learning

project

Jan Mycka, Maciej Żelaszczyk,
Adam Żychowski, Stanisław Kaźmierczak

Faculty of Mathematics and Information Science,
Warsaw University of Technology

2024

Overview

1. Assessment
2. Project I - Convolutional neural networks
3. Project II - Transformers
4. Project III - Diffusion

Overview

- **3 projects**
- Project I (35 points), Project II (35 points), Project III (30 points)
- groups of 2 people
- first week of delay: -5 points; second week of delay: additional -10 points (-15 points in total); exceeding 2 weeks of delay results in 0 points for the project
- grades: [0;50] $\rightarrow$ 2, [51;60] $\rightarrow$ 3, [61;70] $\rightarrow$ 3.5, [71;80] $\rightarrow$ 4, [81;90] $\rightarrow$ 4.5, [91;100] $\rightarrow$ 5,

Timetable

1.	Feb. 19/20	**Introduction to Project 1**
2.	Feb. 26/27	Project 1 plan
3.	Mar. 4/5	Tutorial hours
4.	Mar. 11/12	Tutorial hours
5.	Mar. 18/19	Initial presentation of Project 1
6.	Mar. 25/26	**Project 1 deadline, introduction to Project 2**
7.	Apr. 8/9	Project 2 plan
8.	Apr. 15/16	Tutorial hours
9.	Apr. 22/23	Initial presentation of Project 2
10.	May 6/7	**Project 2 deadline, introduction to Project 3**
11.	May 13/14	Project 3 plan
12.	May 20/21	Tutorial hours
13.	May 27/28	Tutorial hours
14.	Jun. 3/4	Initial presentation of Project 3
15.	Jun. 10/11	**Project 3 deadline**

Rules

- the mandatory classes are only those in bold font.
- initial presentation provides an opportunity to discuss progress in the project but is not obligatory.
- let me know a day before (no later than 6p.m.) if you plan to attend non-mandatory classes.
- all required artifacts (project plan, code, report, presentation) should be sent before the beginning of the respective class.
- you can utilize code from external sources (books, articles, blogs) provided that:
  - reference is cited in the report
  - some modifications to the original solution are applied
- violation of the above or any other kind of plagiarism results in a failing grade.

Project guidelines

General

- take care of reproducibility by initializing a random number generator with a constant seed.
- training models on other datasets than mentioned in the task description is not allowed.
- application of pre-trained models (Inception, EfficientNet, etc.) is permitted (and even recommended as a part of the experiments).
- full discretion in terms of libraries
- crucial elements taken into account while assessment:
  - report quality
  - achieved accuracy registered on Kaggle (if applicable to the particular task)
  - the originality of the solution
- present your experiments and results during a 10-minute slide show.

Reports

The report should include:

- description of the research problem, understandable to the person who did not see the content of the task
- instruction of the application (containing information on how to reproduce results)
- theoretical introduction
- description of the conducted experiments
- statistically processed results (presented clearly)
- conclusions, presumed reasons for successes/failures and further research proposals

Reports

Some additional remarks:

- if the experiment is not described in the report it is regarded as not conducted
- the report is an official document, so please keep it formal (table of contents, bibliography, captions under figures, tables, etc.)
- results should be commented
- to obtain statistically significant results, each experiment ought to be repeated multiple times (when possible)
- in addition to the mean, standard deviation should also be calculated (in some scenarios worse mean with low variance may be a more desirable result than a better mean with high variance)

Project I - Convolutional neural networks

Topic: **Image classification with convolutional neural networks**

Dataset: CINIC-10
`https://www.kaggle.com/datasets/mengcius/cinic10/data`

Project I - Convolutional neural networks

- Test and compare different network architecture (at least one should be a convolutional neural network)
- Investigate influence of the following hyper-parameter change on obtained results:
  - At least 2 hyper-parameters related to training process
  - At least 2 hyper-parameters related to regularization
- Investigate influence of at least X data augmentation techniques from the following groups:
  - Standard operations (where x=3)
  - More advanced data augmentation techniques like mixup, cutmix, cutout (where x=1)
- Consider application of ensemble (hard/soft voting, stacking)

Assessment: source code, report, presentation

Project I - Convolutional neural networks

Useful resources:

- Google Colab or GPU to speed up training
- https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/
- https://paperswithcode.com/sota/image-classification-on-cinic-10
- https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/
- https://machinelearningmastery.com/voting-ensembles-with-python/
- https://medium.com/kaggle-blog/profiling-top-kagglers-bestfitting-currently-1-in-the-world-58cc0e187b

Project II - Transformers

Topic: **Speech commands classification with Transformers**

Dataset: Speech Commands Dataset
`https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data`

Project II - Transformers

- test and compare different network architectures (at least one of them should be a Transformer)
- investigate influence of parameters change on the obtained results
- present confusion matrix (with appropriate discussion)
- in case of accuracy or efficiency problem a subset of classes can be selected and tested (e.g. only “yes” and “no” commands)
- please pay special attention on “silence” and “unknown” classes - test different approaches (e.g. separate network for their recognition)

Assessment: source code, report, presentation

Project II - Transformers

Useful resources:

- Speech representation and data exploration
- Attention Is All You Need
- The Annotated Transformer

Project III - Diffusion

Topic: **Image generation with diffusion models**

Dataset: 10% sample of the LSUN Bedrooms Dataset
`https://www.kaggle.com/jhoward/lsun_bedroom`

The whole dataset is $\sim$ 4.7GB in total.

Project III - Diffusion

- test and compare different network architectures (at least one of them should converge to generate satisfactory images)
- potential architectures (can use different ones): vanilla pixel diffusion, DDPM, Improved DDPM, Stable Diffusion
- calculate the Fréchet Inception Distance (FID) for your generated images and compare it to results from literature
- assess your results qualitatively
- investigate the influence of hyperparameters on obtained results

Project III - Diffusion

- discuss sets of hyperparameters which help in overcoming training collapse and mode collapse
- select two of your generated images together with their latent noise matrix; interpolate linearly between the two latent matrices to generate 8 additional latent matrices; use these 8 matrices to run the diffusion process and generate images from your model; present the 10 generated images (8 newly generated and 2 generated previously) and discuss the importance of the results
- discuss any additional findings
- no required language or libraries; suggested language: Python; suggested libraries: PyTorch or TensorFlow

Assessment: source code, report, presentation

Project III - Diffusion

Useful resources:

- Original diffusion paper
- DDPM
- Improved DDPM
- Diffusion Models Beat GANs on Image Synthesis
- Stable Diffusion

The End