Deep Learning

project

Karolina Bogacka, Stanisław Kaźmierczak,
Zinelabidine Leghelimi, Adam Żychowski

Faculty of Mathematics and Information Science,
Warsaw University of Technology

2025

Agenda

1. Rules
2. Project I - Convolutional neural networks
3. Project II - Transformers
4. Project III - Generative models

Timetable

1.	Feb. 25/26	**Introduction to Project 1**
2.	Mar. 4/5	Project 1 plan
3.	Mar. 11/12	Tutorial hours
4.	Mar. 18/19	Tutorial hours
5.	Mar. 25/26	Initial presentation of Project 1
6.	Apr. 1/2	**Project 1 deadline, introduction to Project 2**
7.	Apr. 8/9	Project 2 plan
8.	Apr. 15/16	Initial presentation of Project 2
–/9.	Apr. 22/23	No classes / Tutorial hours
9./10.	Apr. 29/30	**Project 2 deadline, introduction to Project 3**
10./11.	May 6/7	Project 3 plan
11./–	May 13/14	Tutorial hours / No classes
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 (1)

- Project I (35 points), Project II (35 points), Project III (30 points).
- 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.
- Groups of 2 people
- All required artifacts—report, source code, and presentation slides—should be uploaded on the LeON platform before the beginning of the class when the project is presented.
- 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.
- Presentation of experiments and results in a 10-minute slideshow.
- Presence of both members of the group is required during the presentation.

Rules (2)

- First week of presentation delay: -2 points; second week of delay: an additional -4 points (a total of -6 points); exceeding 2 weeks of delay results in 0 points for the project.
- The project plan should be uploaded to the LeON platform before the start of the relevant class as per the timetable.
- The mandatory classes are only those in bold font.
- Other classes, the initial presentation in particular, provide an opportunity to discuss progress on the project, but are not mandatory.
- Let me know a day in advance (no later than 6 p.m.) if you plan to attend non-mandatory classes.
- 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

- 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:
  - Research process
  - Report quality
  - The novelty/originality of the solution
  - Model performance (last and least)

Reports

The report should include:

- Description of the research problem, understandable to the person who did not see the content of the task.
- Theoretical introduction and literature review.
- Description of the conducted experiments.
- Statistically processed results (presented clearly).
- Conclusions, presumed reasons for successes/failures and further research proposals.
- Instruction of the application (containing information on how to reproduce results).

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; avoid list of tables and charts with minimal commentary.
- 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).

Working with limited computing resources

- We are aware that access to computing power is limited.
- From the assessment perspective, the research process is more important than the quantitative results achieved.
- If you encounter a problem due to insufficient computing power:
  1. Document it in the report.
  2. Consider what can be changed to make the task feasible with fewer resources, e.g., training on a smaller number of observations, completely removing observations of some classes, reducing image resolution, etc.

Project I - Convolutional neural networks

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

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

Project I - Convolutional neural networks

- Test and compare different network architectures (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 ($x = 3$).
    - More advanced data augmentation techniques like cutmix, cutout, or AutoAugment ($x = 1$).
- Implement one method dedicated to few-shot learning.
  - Reduce the size of the training set and compare the obtained results with those trained on the entire dataset (from the previous point).
- Consider application of ensemble (hard/soft voting, stacking).

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).
- Hint: In case of accuracy or efficiency problems, start with a subset of classes (e.g., only the 'yes' and 'no' commands).
- Pay special attention to the "silence" and "unknown" classes—test different approaches (e.g., a separate network for their recognition, under/oversampling, etc.).

Project II - Transformers

Useful resources:

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

Project III - Generative models

Topic: **Image generation**

Dataset: Cats
`https://www.kaggle.com/datasets/borhanitrash/cat-dataset`

Project III - Generative models

- Use the given dataset to generate images of dogs and cats.
- It is up to you which approach to apply.
- Potential methods include, but are not limited to, generative adversarial networks, diffusion models, and variational autoencoders, as well as their extensions (e.g., VQ-VAE, VQ-VAE-2).
- Investigate the influence of hyperparameters on obtained results.
- Compare the results of the methods quantitatively, e.g., using the Fréchet Inception Distance (FID). Assess them qualitatively as well.
- Address the mode collapse problem if it occurs.
- 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 generate images from your model; present the 10 generated images (8 newly generated and 2 generated previously) and discuss the importance of the results.

Project III - Generative models

- Discuss any additional findings.
- Remark 1: The images have different resolutions. You should decide how to address this issue.
- Remark 2: Training generative models may be time- and resource-consuming. An essential part of this task is determining how to approach this issue; please refer to "Working with limited computing resources" slide.
- Additional task: Train a model on a dataset containing cats and dogs (`https://www.kaggle.com/competitions/dogs-vs-cats/`). Compare the results with those obtained from the dataset containing only cats. Do the generated observations resemble cats and dogs, or rather a combination of both?

Project III - Generative models

Useful resources:

- Original diffusion paper
- DDPM
- Improved DDPM
- Diffusion Models Beat GANs on Image Synthesis
- Stable Diffusion
- GAN tutorial
- An Introduction to Variational Autoencoders

The End