GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Optimization of Structures & Calculations in Neural Networks, PG_00064490		
Field of study	Informatics		
Date of commencement of studies	February 2026	Academic year of realisation of subject	2026/2027
Education level	second-cycle studies	Subject group	Optional subject group Specialty subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	3	ECTS credits	3.0
Learning profile	general academic profile	Assessment form	exam
Conducting unit	Department of Multimedia Systems -> Faculty of Electronics Telecommunications and Informatics -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. inż. Piotr Szczuko	
	Teachers	dr hab. inż. Piotr Szczuko	

Lesson types	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	15.0	15.0	0.0	45
	E-learning hours included: 0.0						
	eNauczanie source address: https://enauczanie.pg.edu.pl/moodle/enrol/index.php?id=47727						

Learning activity and number of study hours	Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
	Number of study hours	45	6.0	24.0	75

Subject objectives	The goal is to present theory, practice and problems solving in a domain of models optimisation. Techniques for structure prunning, sparsing, architecture simplification, calculations accelerations are presented. Various approaches for effective training, robustness assurance, accuracy and precision for real-world applications, e.g. in case of limited resources or noisy data.

Learning outcomes			
	Course outcome	Subject outcome	Method of verification
	[K7_U07] can apply advanced methods of process and function support, specific to the field of study	Student created a machine learning model and optimized it with respect to the model goal, model structure. Student correctly used chosen library and programming language.	[SU4] Assessment of ability to use methods and tools [SU1] Assessment of task fulfilment
	[K7_W10] knows and understands, to an increased extent, the basic processes occurring in the life cycle of equipment, objects and technical systems, as well as methods of supporting processes and functions, specific to the field of study	The student knows and understands advanced methods for optimizing deep learning models across the AI system lifecycle, from architecture design, through training, to deployment, including compression techniques, computational acceleration, and ensuring effective performance under conditions of limited resources and noisy data.	[SW1] Assessment of factual knowledge [SW3] Assessment of knowledge contained in written work and projects
	[K7_W11] knows and understands, to an increased extent, the general principles of creation and development of forms of individual entrepreneurship and the economic, legal and other conditions of various types of activities related to the awarded qualification, including the principles of protection of industrial property and copyright law	The student knows and understands the economic, legal, and ethical conditions related to the deployment and commercialization of optimized deep learning models, including intellectual property protection principles concerning algorithms, neural network architectures, and software	[SW3] Assessment of knowledge contained in written work and projects [SW1] Assessment of factual knowledge
	[K7_U12] is able, to an increased extent, to analyze the operation of components and systems related to the field of study, as well as to measure their parameters and study their technical characteristics, and to plan and carry out experiments related to the field of study, including computer simulations, interpret the obtained results and draw conclusions	The student is able to analyze the operation and measure the parameters of optimized deep learning models, including examining the technical characteristics of neural network architectures in terms of computational efficiency and accuracy, as well as plan and conduct experiments, concerning compression techniques, computational acceleration, and training under conditions of limited resources and noisy data, and to interpret the obtained results and draw conclusions.	[SU1] Assessment of task fulfilment [SU4] Assessment of ability to use methods and tools [SU3] Assessment of ability to use knowledge gained from the subject

Subject contents

Course content – lecture
Neural model reduction, calculations accelerations.
Quantisation, sparsification, knowledge distillation.
Noisy labels training.
Network architectures search.
Self-supervised training, pre-training.
Models uncertainty estimation (calibration, test-time dropout, ensambling, Bayes networks).
Models robustness, adversarial techniques.
Hybrid models, weight-agnostic, capsule nets.

Course content – laboratory
Experimental verification of network size reduction possibilitie.
Practical introduction to training models on noisy data. Simulated and real label errors in well-known datasets.
Leveraging self-supervised learning and model pre-training to improve generalization capabilities.
Introduction to uncertainty estimation methods in neural networks and model calibration.

Course content – project
Critical analysis of the state of knowledge and selection of model and dataset.
Analysis of the state of knowledge in terms of selecting model optimization methods.
Preparation of the research environment.
Model optimization experiments, data adjustment.
Formulation of conclusions, documentation and public presentation of the project.

Prerequisites and co-requisites

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Laboratory	51.0%	35.0%
Colloquy	51.0%	35.0%
Project	51.0%	30.0%

Recommended reading	Basic literature	Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, Alexandra Peste, (2021) Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks. [2102.00554] (arxiv.org). Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. 2020. A Survey of Model Compression and Acceleration for Deep Neural Networks. (2020). arXiv:cs.LG/1710.09282. Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural Architecture Search: A Survey. (2019). arXiv:stat.ML/1808.05377. Manish Gupta and Puneet Agrawal. 2020. Compression of Deep Learning Models for Text: A Survey. (2020). arXiv:cs.CL/2008.05221. V. Sze, Y. Chen, T. Yang, and J. S. Emer. 2017. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proc. IEEE 105, 12 (2017), 22952329. https://doi.org/10.1109/JPROC.2017.2761740.
	Supplementary literature	Tensorflow model optimization (2022) https://www.tensorflow.org/ model_optimization. Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2020. Efficient transformers: A survey. (2020). arXiv:cs.LG/2009.06732.
	eResources addresses	
Example issues/ example questions/ tasks being completed	• Describe and comment on one of chosen methods for optimisation, justify its use: network pruning and sparse processing, quantisation, knowledge distillation. • Justify the need for for data sparsification and architecture sparsification, and benefits of those operations. • Describe how a training on noisy labels can be efficiently performed. • Describe how the self-supervision and pre-training work. What are the benefits of these procedures. • Describe methdos for calibration of neural networks, dropout and models ensambling. • How to estimate the model robustness? • Give an example of weight-agnostic model, and application of capsule networks.	
Practical activites within the subject	Not applicable	