



---

**POZNAN UNIVERSITY OF TECHNOLOGY**

---

FACULTY OF COMPUTING AND TELECOMMUNICATION

**BEYOND TRADITIONAL CURRICULUM LEARNING: SHAPING  
THE OPTIMIZATION PATH TO FIND BETTER MINIMA, FASTER**

Izabela Krysińska

Supervisor

dr. hab. inż. Mikołaj Morzy, prof. PP

POZNAŃ 2025

# Abstract

Training of modern deep neural networks traditionally relies on presenting data in a random, unstructured order. This thesis challenges that paradigm by systematically exploring curriculum learning (CL), a training methodology inspired by the human learning principle of starting with easy concepts before moving to more complex ones. Despite its intuitive appeal and early promise, CL has failed to become a standard tool in the machine learning practitioner’s toolkit. This thesis addresses the central paradox of CL: the wide gap between its theoretical potential and its inconsistent, often disappointing, empirical results in practice. The core motivation of this work is to move beyond the collection of ad-hoc methods that characterize much of the CL literature. It aims to establish a more principled, analytical foundation for the field by systematically investigating not just *if* a curriculum works, but *why* and *how* it influences the fundamental dynamics of neural network training.

This work introduces a formal framework for curriculum learning and systematically investigates its core components — scoring function that defines data or task difficulty and pacing function that schedules the curriculum — through a series of novel methods and comprehensive empirical studies. We explore the scoring function across two primary dimensions: in the data space and in the task space. First, we propose a typicality-based static scoring, hypothesizing that prioritizing samples that are good representatives of their classes will help the model reach better minima and improve robustness. Second, we introduce VoG-Guided Learning (VGL), a data-driven curriculum that employs an adaptive scoring function based on the Variance of Gradients (VoG) to sample examples based on the model’s uncertainty in real-time. Third we propose Transitional Objective Learning (TOL), a task-space curriculum that leverages a static, domain-specific hierarchy of objectives in speech recognition. We then conduct an in-depth investigation of the pacing function through the lens of batch composition, treating dynamic batch size scheduling as a primary curriculum mechanism and analyzing its effect in combination with different scoring functions.

Across a diverse range of tasks, from tabular data analysis, image classification, text classification, to speech recognition, our findings demonstrate that principled curriculum design significantly improves training outcomes. VGL consistently outperformed baselines by dynamically focusing on the most informative samples. TOL was shown to accelerate convergence by leveraging the natural hierarchy of phonemes. Additionally, our study of batch orchestration reveals that dynamic, non-monotonic pacing functions can act as a powerful implicit regularizer, effectively mitigating the generalization gap in large-batch training and outperforming simple warm-starting strategies. In summary, this thesis provides a systematic perspective on curriculum learning, moving it from a collection of disparate heuristics towards a principled discipline. By formally analyzing and empirically validating the roles of scoring and pacing, this work demonstrates that the deliberate orchestration of the training process is a powerful tool for more efficient development of higher-performing deep learning models.

# Streszczenie

Trenowanie nowoczesnych, głębokich sieci neuronowych tradycyjnie opiera się na prezentowaniu danych w losowej kolejności. Niniejsza rozprawa doktorska podważa ten paradygmat, systematycznie badając uczenie programowe (curriculum learning, CL) — metodologię trenowania inspirowaną procesem uczenia u ludzi, polegającą na rozpoczynaniu od koncepcji łatwych przed przejściem do bardziej złożonych. Pomimo swojej intuicyjnej atrakcyjności i obiecujących wyników początkowych, CL nie stało się standardowym narzędziem w arsenale praktyków uczenia maszynowego. Rozprawa ta traktuje o głównym paradoksie związanym z CL: o niezgodności pomiędzy jego teoretycznym potencjałem a często rozczarowującymi, wynikami empirycznymi. Główną motywacją tej pracy jest wykroczenie poza zbiór niesystematycznych metod charakteryzujący znaczną część literatury na temat CL. Celem jest ustanowienie bardziej pryncypialnych, analitycznych podstaw dla tej dziedziny poprzez zbadanie nie tylko *czy* CL działa, ale przede wszystkim *dlaczego* i *jak* wpływa ono na fundamentalną dynamikę optymalizacji uczenia sieci neuronowych.

W pracy tej wprowadzono formalne ramy dla uczenia programowego i systematycznie zbadano jego kluczowe komponenty — funkcję określającą trudność (scoring function), która definiuje złożoność danych bądź zadań, oraz funkcję sterującą tempem (pacing function), która kontroluje czego w danym momencie model się uczy — poprzez serię nowatorskich metod i kompleksowych badań empirycznych. Funkcje określające trudność badamy w dwóch głównych wymiarach: w przestrzeni danych i w przestrzeni zadań. Po pierwsze, proponujemy statyczną ocenę trudności opartą na typowości próbek, zakładając, że priorytetyzacja próbek dobrze reprezentujących swoją klasę pomoże znaleźć lepsze minimum i wzmocni odporność na szum danych. Po drugie, wprowadzamy uczenie kierowane wariacją gradientów (VGL), program nauczania zdefiniowany w przestrzeni danych, który wykorzystuje adaptacyjną funkcję określającą trudność opartą na wariacji gradientów (VoG) do próbkowania przykładów na podstawie niepewności modelu w czasie rzeczywistym. Po trzecie, proponujemy uczenie z progresywną funkcją celu (TOL), program nauczania w przestrzeni zadań, który wykorzystuje statyczną, specyficzną dla domeny hierarchię celów w rozpoznawaniu mowy. Następnie przeprowadzamy dogłębne badanie funkcji sterującej tempem przez pryzmat kompozycji wsadów (batch composition), traktując dynamiczne harmonogramowanie rozmiaru wsadu jako główny mechanizm programu nauczania i analizując jego działanie w połączeniu z różnymi funkcjami określającymi trudność.

Nasze wyniki, uzyskane na zróżnicowanym zbiorze zadań, od analizy danych tabelarycznych, przez klasyfikację obrazów i tekstu, po rozpoznawanie mowy, pokazują, że analityczne projektowanie programu nauczania znacząco usprawnia proces trenowania. Modele trenowane z użyciem VGL konsekwentnie przewyższały dokładnością modele trenowane standardowymi metodami dzięki dynamicznemu skupianiu się na najbardziej informatywnych próbkach. Wykazano, że TOL przyspiesza zbieżność do minimum lokalnego, dzieląc złożony problem na składowe poprzez wykorzystanie naturalnej hierarchii fonemów. Dodatkowo, nasze badanie orkiestracji wsadów (batch orchestration) pokazuje, że dynamiczne, niemonotoniczne funkcje sterujące tempem nauczania

mogą działać jako czynnik regularyzacji, skutecznie poprawiając generalizację modelu w trenowaniu z użyciem dużych wsadów i przewyższając proste strategie typu warm-starting. Podsumowując, niniejsza rozprawa dostarcza formalne ramy dla uczenia programowego, przekształcając je w kierunku zdefiniowanej dyscypliny. Poprzez analizę i empiryczną walidację funkcji określających trudność i sterujących tempem, praca ta dowodzi, że świadoma orkiestracja procesu trenowania jest skutecznym narzędziem do wydajniejszego trenowania dokładniejszych głębokich sztucznych sieci neuronowych.