Marcin Kurdziel
Wydział Informatyki
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
al. Mickiewicza 30, 30-059 Kraków
kurdziel@agh.edu.pl

Recenzja rozprawy doktorskiej dla Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Uniwersytetu Jagiellońskiego

Tytuł rozprawy: *Data-Efficient Representation Learning and Adaptation*
Autor rozprawy: mgr inż. Marcin Przewięźlikowski

Niniejszą recenzję sporządziłem w odpowiedzi na pismo Zastępcy Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Uniwersytetu Jagiellońskiego, dr hab. Barbary Strug, prof. UJ, z dnia 25 maja 2026 r., w związku z powołaniem mnie przez Radę Dyscypliny na recenzenta w postępowaniu w sprawie nadania stopnia doktora Panu mgr inż. Marcinowi Przewięźlikowskiemu.

1 Obszar badań i uzyskane wyniki naukowe

Wątkiem przewodnim recenzowanej rozprawy doktorskiej jest rozwój metod uczenia maszynowego dla problemów, w których nie dysponujemy dużymi zbiorami etykietowanych danych. Temat ten Doktorant podejmuje w ramach dwóch obszarów badań. Pierwszy z nich koncentruje się na metodach pozwalających na szybką (w sensie liczby przykładów uczących) adaptację modelu do nowych zadań, tj. na algorytmach uczenia z kilku przykładów (ang. *few-shot learning*). Drugi obszar badań Doktoranta poświęcony jest rozwojowi samonadzorowanych metod uczenia się reprezentacji (ang. *self-supervised representation learning*). Wyniki uzyskane w ramach tych prac badawczych Doktorant opublikował w sześciu artykułach naukowych:

[A1] M. Przewięźlikowski, P. Przybysz, J. Tabor, M. Zięba, P. Spurek, *HyperMAML:*

Few-shot adaptation of deep models with hypernetworks. Neurocomputing 598, 2024, 128179.

[A2] M. Sendera, M. Przewięźlikowski, K. Karanowski, M. Zięba, J. Tabor, P. Spurek, *HyperShot: Few-shot Learning by Kernel HyperNetworks.* W: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, s. 2469-2478.

[A3] M. Sendera, M. Przewięźlikowski, J. Miksa, M. Rajski, K. Karanowski, M. Zięba, J. Tabor, P. Spurek, *The general framework for Few-shot Learning by Kernel HyperNetworks.* Machine Vision and Applications 34(4), 2023, 53.

[A4] M. Przewięźlikowski, M. Pyla, B. Zieliński, B. Twardowski, J. Tabor, M. Śmieja, *Augmentation-aware Self-supervised Learning with Conditioned Projector.* Knowledge-Based Systems 305, 2024, 112572.

[A5] M. Przewięźlikowski, R. Balestriero, W. Jasiński, M. Śmieja, B. Zieliński, *Beyond [cls]: Exploring the True Potential of Masked Image Modeling Representations.* W: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025, s. 23442-23452.

[A6] M. Przewięźlikowski, M. Osial, B. Zieliński, M. Śmieja, *A deep cut into Split Federated Self-supervised Learning.* W: Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD), 2024, s. 444-459.

Powyższy cykl publikacji wraz z dołączonym autoreferatem stanowi rozprawę doktorską będącą przedmiotem niniejszej recenzji. Wyniki te omówię i skomentuję zgodnie z przyjętym przez Doktoranta podziałem na dwa główne obszary badań.

Nowe algorytmy dla zadań uczenia z kilku przykładów

W pracach [A1], [A2] i [A3] Doktorant rozwija koncepcję dostosowania modelu do nowego zadania za pomocą tzw. hipersieci (ang. *hypernetwork*). W skrócie, hipersieć to sieć neuronowa, która proponuje parametry (wagi) dla pewnej innej sieci neuronowej. W kontekście dostosowania modelu uczącego do nowego zadania, hipersieć może obserwować pewną pulę przykładów tego zadania i na ich podstawie emitować wagi dla sieci mającego klasyfikować kolejne podobne przykłady. Jednym z wczesnych przykładów zastosowania tej koncepcji jest praca L. Bertinetto i in., *Learning feed-forward one-shot learners* (NeurIPS 2016).

W pracy [A1] Doktorant wykorzystuje hipersieci do rozszerzenia jednej z klasycznych metod uczenia z kilku przykładów – algorytmu MAML (ang. *Model-Agnostic Meta-Learning.* C. Finn i in., *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*, ICML 2017). W szczególności, zaproponowany w [A1] algorytm *HyperMAML* wykorzystuje hipersieć do wyznaczania poprawki do wag klasyfikatora bazowego, adaptującej ten klasyfikator do nowego zadania. Hipersieć wyznacza poprawkę do wag, obserwując osadzenia przykładów uczących, ich prawidłowe etykiety oraz etykiety zwrócone przez klasyfikator bazowy.

W ten sposób HyperMAML zastępuje kosztowną, iteracyjną adaptację gradientową (z klasycznej metody MAML) pojedynczą predykcją hipersieci. W pracy [A1] przedstawiono szeroką eksperymentalną ewaluację HyperMAML. Choć algorytm osiąga wyniki wyraźnie lepsze od klasycznego MAML, to wśród metod referencyjnych można wskazać algorytmy o zbliżonej lub nieco lepszej dokładności. Warto jednak podkreślić, że architektura hipersieci badana w [A1] adaptuje tylko jedną warstwę klasyfikatora bazowego, potencjalnie ograniczając uzyskane wyniki. Lektura [A1] sugeruje również, że samo trenowanie hipersieci nie jest proste: Doktorant wykorzystuje tu stopniowe przejście od klasycznej adaptacji gradientowej do adaptacji hipersiecią ([A1], równanie 7). Pojawia się tu pytanie, czy mechanizm ten jest wykorzystywany tylko podczas treningu metamodelu, czy może jest niezbędny również podczas późniejszej adaptacji do nowych zadań?

W [A2] Doktorant proponuje algorytm *HyperShot* łączący hipersieci z metodami wykorzystującymi funkcje jądrowe (ang. *kernel functions*). O ile w [A1] hipersieć operowała na osadzeniach przykładów z nowego zadania, to w [A2] zbiór przykładów reprezentowany jest macierzą Grama skonstruowaną z takich osadzeń. Na podstawie tej macierzy hipersieć proponuje parametry klasyfikatora dla nowego zadania. Sam klasyfikator również operuje w przestrzeni funkcji jądrowej. Wyniki eksperymentalne zawarte w [A2] demonstrują, że zaproponowany algorytm bardzo dobrze radzi sobie w sytuacji, gdy klasyfikator bazowy adaptowany jest do zadania z innej domeny (np. adaptacja klasyfikatora ręcznie pisanych znaków alfabetu do klasyfikacji ręcznie pisanych cyfr). HyperShot uzyskuje w tym kontekście wyniki lepsze od istniejących wówczas metod referencyjnych.

W [A3] Doktorant proponuje *BayesHyperShot* – rozszerzenie algorytmu HyperShot o mechanizm szacowania niepewności predykcji. Pomysł polega tu na zbudowaniu klasyfikatora bazowego w postaci bayesowskiej sieci neuronowej. Hipersieć proponuje wówczas parametry rozkładu nad wagami tego klasyfikatora. Tak skonstruowany metamodel uczony jest następnie klasycznym podejściem wariacyjnym. W trakcie wnioskowania próbki z wariacyjnego rozkładu a posteriori dają estymator Monte Carlo dla rozkładu prawdopodobieństwa nad klasą przykładu wejściowego.

Lektura pracy [A3] nasuwa pewne pytanie odnośnie sposobu konstrukcji modelu wariacyjnego. Model przedstawiony w [A3] (sekcja 2.6) zakłada mianowicie, że wagi klasyfikatora mają a priori rozkład normalny $\mathcal{N}(0, \mathbf{I})$. Hipersieć wyznacza natomiast parametry wariacyjnego rozkładu a posteriori. W tym miejscu można zadać pytanie, dlaczego model zaproponowany w [A3] regularyzuje parametry dla każdego zadania (do którego adaptowany jest klasyfikator) do standardowego rozkładu normalnego? Na pierwszy rzut oka bardziej intuicyjnym wydawałoby się wykorzystanie hipersieci do wyznaczania parametrów warunkowego rozkładu a priori – hipersieć proponowałaby wówczas wartość oczekiwaną i wariancję w rozkładzie a priori $\mathcal{N}\left(\boldsymbol{\mu}_0(S_i), \mathrm{diag}(\boldsymbol{\sigma}_0(S_i)^2)\right)$. Czynnik regularyzujący w ELBO utrzymywałby wówczas parametry wariacyjnego rozkładu a posteriori w pobliżu wartości zaproponowanych przez hipersieć. Niezależnie od powyższej uwagi, wyniki eksperymentalne zawarte w [A3] wskazują, że parametryzacja zaproponowana w pracy spełnia postawiony cel. Estymacja niepewności umożliwia w szczególności wykrywanie przykładów

spoza rozkładu, do którego dostosowano klasyfikator.

Rozwój metod samonadzorowanego uczenia się reprezentacji

W pracach [A4], [A5] i [A6] Doktorant badał i rozwijał metody samonadzorowanego uczenia się reprezentacji. Prace te skupiały się na dwóch rodzinach algorytmów: algorytmach budujących osadzenia na podstawie dwóch zmodyfikowanych wersji tego samego obrazu (ang. *Joint-embedding Architectures*, JEA) oraz algorytmach budujących osadzenia poprzez uczenie się uzupełniania zamaskowanych fragmentów obrazu (ang. *Masked Image Modeling*, MIM).

W [A4] Doktorant zaproponował metodę *Conditional Augmentation-aware Self-supervised Learning* (CASSLE). Klasyczne algorytmy z rodziny JEA uczą się budować osadzenia, które są niezmiennicze ze względu na pewne modyfikacje obrazu, np. pewne zmiany kolorów. Choć osadzenia te są skuteczne w wielu zadaniach docelowych, ich niezmienniczość może być źródłem problemów. Przykładowo, niezmienniczość względem zmian kolorystyki może utrudniać klasyfikację obiektów rozróżnialnych głównie kolorem. Algorytm CASSLE rozszerza więc klasyczną architekturę JEA o moduł kodujący informację o charakterze wykonanej modyfikacji obrazu. Doktorant argumentuje, że moduł ten wymusza na sieci osadzającej zachowanie informacji o wykonanej modyfikacji. I faktycznie, wyniki eksperymentalne zawarte w [A4] wskazują, że osadzenia budowane przez CASSLE są bardziej wrażliwe na modyfikacje obrazu niż osadzenia z metod referencyjnych. Co więcej, ewaluacja klasyfikatorami liniowymi pokazuje, że osadzenia CASSLE rozróżniają klasy obrazów z większą dokładnością niż osadzania budowane przez metody referencyjne.

W pracy [A5] Doktorant bada i rozwija architektury MIM. Punktem wyjścia do tej pracy jest bardzo ciekawa obserwacja empiryczna. Doktorant bada mianowicie osadzenia budowane przez transformery wizyjne (ang. *Vision Transformer*, ViT). W modelach tych informacja determinująca klasę wejściowej obserwacji agregowana jest w specjalnym tokenie `[cls]`. Doktorant porównał wzorce atencji dla tego tokenu, gdy ViT wykorzystany jest jako rdzeń modelu JEA i MIM. Okazuje się, że w architekturach JEA token `[cls]` koncentruje wagi atencji na wybranych, istotnych elementach obrazu. W architekturze MIM wagi atencji dla tokenu `[cls]` mają względnie jednostajny rozkład na całej powierzchni obrazu. Innymi słowy, osadzenia tokenu `[cls]` budowane w architekturze MIM nie wykazują specyficzności do istotnych (z perspektywy klasyfikacji) cech obrazu. Doktorant argumentuje, iż różnica ta może wyjaśniać słabsze wyniki uzyskiwane przez algorytmy z rodziny MIM.

Kierując się powyższą obserwacją, Doktorant proponuje rozszerzenie klasycznej architektury MIM o dodatkowy mechanizm selektywnej agregacji osadzeń tokenów. Wykorzystuje w tym celu znaną z literatury metodę *Attention-based Multiple Instance Learning Pooling*. Uzyskane wyniki eksperymentalne pokazują, że zaproponowana selektywna agregacja znacząco poprawia jakość osadzeń budowanych za pomocą architektury MIM. Rezultat ten jest potwierdzony dla kilku różnych architektur z rodziny MIM, zarówno dla małych, jak i

dużych sieci ViT. Co więcej, ta sama modyfikacja wprowadzona w architekturę JEA praktycznie nie zmienia jakości osadzeń, co pośrednio potwierdza hipotezę stojącą za motywacją pracy [A5]. Uważam, że jest to najciekawszy wynik naukowy w zaprezentowanym cyklu. Został on opublikowany na wiodącej konferencji naukowej z obszaru widzenia komputerowego.

W [A6] Doktorant bada metody uczenia samonadzorowanego w kontekście uczenia federacyjnego, tj. w sytuacji, gdy model trenowany jest w sposób rozproszony. Przedmiotem badań Doktoranta jest w szczególności klasyczna metoda MocoSFL (J. Li i in., *MocoSFL: enabling cross-client collaborative self-supervised learning*, ICLR 2023). MocoSFL to metoda uczenia się reprezentacji z rodziny JEA. W algorytmie tym grupa węzłów obliczeniowych wspólnie trenuje model osadzający. Równocześnie każdy z węzłów utrzymuje wykładniczą średnią kroczącą wag lokalnej kopii tego modelu - jest ona wykorzystywana do budowy osadzenia zmodyfikowanej wersji obrazu. Doktorant dostrzegł pewną niekonsekwencję w tym schemacie rozproszenia: podczas gdy model trenowany kolaboratywnie jest cyklicznie agregowany pomiędzy węzłami, lokalna wykładnicza średnia krocząca jego wag nie jest synchronizowana. Wprowadza to niestabilność w procesie uczenia. Doktorant proponuje więc zmodyfikowany wariant algorytmu MocoSFL, w którym cyklicznie agregowany jest zarówno model osadzający, jak i wykładnicze średnie kroczące jego wag. Modyfikacja ta znacząco poprawia stabilność uczenia, dzięki czemu model może zostać podzielony (pomiędzy serwer i rozproszone węzły obliczeniowe) dowolnie głęboko w sekwencji warstw. Pozwala to zmniejszyć koszt komunikacji i poprawić prywatność danych uczących.

2 Ocena uzyskanych wyników

Wyniki naukowe przedstawione w recenzowanej rozprawie doktorskiej oceniam bardzo dobrze. Doktorant podjął ciekawy i aktualny problem uczenia maszynowego, jakim jest samonadzorowane uczenie się reprezentacji obrazów. Zaproponował nowe architektury metamodeli oraz nowe algorytmy budowania osadzeń. Każda z omawianych prac prawidłowo i rzetelnie ocenia proponowane rozwiązania. Doktorant każdorazowo odnosi uzyskane wyniki do uznanych metod referencyjnych. Obszernie dyskutuje również stan wiedzy w podejmowanej tematyce badawczej.

W mojej ocenie najciekawszym wynikiem przedstawionym w rozprawie jest ulepszony algorytm uczenia się reprezentacji poprzez rekonstrukcję zamaskowanych obszarów obrazu (praca [A5]). Doktorant przedstawił tu bardzo ciekawą obserwację empiryczną: klasyczne metody MIM budowane na bazie architektury ViT nie agregują w odpowiedni sposób cech istotnych z punktu widzenia klasyfikacji obrazu. Następnie zaproponował prosty, ale skuteczny mechanizm zapobiegający temu zjawisku. Wyniki eksperymentalne pokazują trafność postawionej hipotezy badawczej i praktyczną poprawę istotnej klasy metod uczenia się reprezentacji. Wynik ten został opublikowany na konferencji ICCV (ranga CORE A*), jednej z wiodących konferencji z obszaru widzenia komputerowego. Interesujące są również

propozycje wykorzystana metauczenia w modelach uczenia z kilku przykładów oraz ich rozszerzenia o techniki z obszaru funkcji jądrowych i wnioskowania bayesowskiego. Moje uwagi zawarte w poprzedniej sekcji należy traktować jedynie jako zachętę do dalszej dyskusji uzyskanych rezultatów.

Wszystkie artykuły, na podstawie których Doktorant przygotował rozprawę są wieloautorskie. Autoreferat precyzyjnie opisuje wkład Doktoranta w każdą z tych prac. W moje ocenie jest to każdorazowo wkład znaczący lub kluczowy dla uzyskanych rezultatów.

3 Pozostała aktywność Doktoranta

Poza wkładem do prac [A1]-[A6], Doktorant jest również współautorem w ośmiu innych publikacjach naukowych. Uzyskał finansowanie w ramach grantu NCN PRELUDIUM oraz był wykonawcą w trzech innych grantach NCN. W 2025 r. otrzymał stypendium Ministra Nauki i Szkolnictwa Wyższego dla wybitnych młodych naukowców. Doktorant aktywnie współpracuje również z innymi ośrodkami naukowymi. Odbył między innymi wizytę badawczą w grupie prof. Randalla Balestriero na Uniwersytecie Browna. Uczestniczył w pracach komitetach programowych wielu konferencji z obszaru uczenia maszynowego i widzenia komputerowego.

W mojej ocenie jest to ponadprzeciętna aktywność na tym etapie kariery naukowej.

4 Konkluzja

W mojej ocenie rozprawa doktorska Pana mgr inż. Marcina Przewięźlikowskiego stanowi oryginalne rozwiązanie problemu naukowego i spełnia wymogi sformułowane w art. 187 Ustawy *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r., z późniejszymi zmianami. Wnioskuję o dopuszczenie Pana Marcina Przewięźlikowskiego do dalszych etapów postępowania w sprawie nadania stopnia doktora w dyscyplinie *Informatyka Techniczna i Telekomunikacja.*

Ponadto, biorąc pod uwagę bardzo ciekawy wynik uzyskany w pracy [A5], fakt opublikowania tej pracy na wiodącej konferencji z obszaru widzenia komputerowego oraz uzyskanie finansowania w konkursie NCN PRELUDIUM, wnoszę o wyróżnienie rozprawy doktorskiej.

Z poważaniem,
Marcin Kurdziel

