dr hab. inż. Paweł Morawiecki, prof. IPI PAN Warszawa, 31.08.2026

# Recenzja rozprawy doktorskiej „Data-Efficient Representation Learning and Adaptation" autorstwa Marcina Przewięźlikowskiego

## Tematyka rozprawy

Szeroki kontekst pracy to uczenie głębokich sieci neuronowych. Tematyka rozprawy doktorskiej skupia się na zagadnieniu uczenia się użytecznych reprezentacji z danych nieetykietowanych i szybkiej adaptacji tych reprezentacji do nowych zadań przy bardzo ograniczonej liczbie etykiet. Szczególny nacisk został położony na hiper-sieci (ang. hyper-networks), uczenie samo-nadzorowane (ang. self-supervised learning) oraz praktyczną użyteczność wyuczonych reprezentacji.

Autor stawia trzy pytania badawcze, które spajają cykl publikacji:

- jak projektować modele, które szybko adaptują się do nowych zadań przy małej liczbie etykiet?
- jakie własności reprezentacji umożliwiają skuteczny transfer i adaptację?
- jak poprawić praktyczne zastosowanie metod uczenia samo-nadzorowanego?

Pierwsze pytanie eksploatowane jest w pierwszych trzech publikacjach, z kolei dwa kolejne pytania badane są w pracach IV, V oraz VI.

Pierwszy blok publikacji [I–III] dotyczy uczenia typu few-shot. Punktem wyjścia jest klasyczne podejście, które adaptuje model do nowego zadania przez kilka kroków optymalizacji gradientowej. Doktorant przedstawia alternatywę opartą na hiper-sieciach, które zamiast wykonywać taką optymalizację, bezpośrednio generują parametry albo aktualizacje parametrów modelu dla konkretnego zadania. BayesHyperShot rozszerza ten pomysł probabilistycznie, tak aby model umiał również oceniać niepewność swoich predykcji.

Drugi blok [IV–VI] dotyczy uczenia samo-nadzorowanego i pytania o to, co właściwie sprawia, że wyuczona reprezentacja jest użyteczna. W pracy analizowane są dwa ważne nurty SSL: Joint-Embedding Architectures i Masked Image Modeling. W pracach pokazano, że wymuszanie pełnej niezmienniczości względem augmentacji może usuwać użyteczną informację, dlatego warto zachować informację o zastosowanej transformacji.
Pod lupę brany jest także sposób agregacji lokalnych cech do tworzenia globalnej reprezentacji.

**Cykl publikacji**

Praca składa się z cyklu sześciu publikacji. We wszystkich pracach Doktorant jest pierwszym autorem lub współdzieli pierwszeństwo. Prace zostały zaprezentowane na bardzo dobrych konferencjach i opublikowane w dobrych i wyróżniających się czasopismach. Poniżej przedstawiam krótką charakterystykę prac i najważniejsze wyniki.

1. „HyperMAML: Few-shot adaptation of deep models with hypernetworks", Neurocomputing, (2024).

Publikacja dotyczy problemu uczenia typu few-shot czyli szybkiego dostosowania modelu do nowego zadania na podstawie bardzo niewielkiej liczby przykładów. Punktem wyjścia jest metoda MAML, w której adaptacja do nowego zadania wymaga wykonania kilku kroków optymalizacji gradientowej. Autorzy proponują HyperMAML, gdzie aktualizacja parametrów modelu jest bezpośrednio generowana przez hiper-sieć na podstawie danych opisujących nowe zadanie. Pozwala to uniknąć iteracyjnej optymalizacji podczas adaptacji i skrócić jej czas. Eksperymenty pokazują, że HyperMAML osiąga wyniki porównywalne z konkurencyjnymi metodami few-shot learning a jednocześnie szybciej tworzy model dostosowany do konkretnego zadania.

2. „HyperShot: Few-shot Learning by Kernel HyperNetworks", IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), (2023).

Praca rozwija zastosowanie hiper-sieci skupiając się na sposobie reprezentowania informacji o nowym zadaniu. W metodzie HyperShot przykłady należące do support set są przekształcane do przestrzeni reprezentacji, a relacje między nimi są następnie opisywane za pomocą funkcji kernelowej. Powstała macierz podobieństw stanowi wejście do hiper-sieci, która generuje parametry klasyfikatora specyficznego dla danego zadania. Takie podejście pozwala modelować wzajemne relacje między przykładami i jednocześnie zachować odporność na ich kolejność. Metoda okazuje się szczególnie skuteczna w zadaniach cross-domain, w których dane testowe pochodzą z innej domeny niż dane wykorzystane podczas treningu.

3. „The general framework for Few-shot Learning by kernel HyperNetworks", Machine Vision and Applications, (2023).

Publikacja rozszerza koncepcję HyperShot i przedstawia bardziej ogólną analizę few-shot learning opartego na hiper-sieciach i metodach kernelowych. Najważniejszym rozszerzeniem jest BayesHyperShot, w którym hiper-sieć nie generuje jednego ustalonego zestawu parametrów klasyfikatora, lecz parametry rozkładu probabilistycznego, z którego mogą być próbkowane różne klasyfikatory. Dzięki temu metoda może nie tylko przewidywać klasę nowego przykładu, ale również oceniać niepewność swojej predykcji. Autorzy pokazują m.in., że model daje bardziej niepewne odpowiedzi dla danych pochodzących spoza rozkładu treningowego. Praca łączy więc szybką adaptację z możliwością ilościowego opisu niepewności modelu.

4. „Augmentation-aware Self-supervised Learning with Conditioned Projector", Knowledge-Based Systems, (2024).

Publikacja dotyczy uczenia samo-nadzorowanego i problemu niezmienniczości reprezentacji względem augmentacji danych. Typowe metody joint-embedding uczą model, aby różne przekształcone wersje tego samego obrazu otrzymywały podobne reprezentacje, jednak może to prowadzić do utraty informacji istotnej dla późniejszych zadań, np. informacji o kolorze. Autorzy proponują CASSLE, w którym informacja o zastosowanej augmentacji jest przekazywana do projektora wykorzystywanego podczas treningu. Dzięki temu encoder nie musi całkowicie usuwać z reprezentacji cech zmienionych przez augmentację. Eksperymenty pokazują, że takie reprezentacje zachowują więcej informacji o właściwościach obrazu i mogą lepiej sprawdzać się w późniejszych zadaniach.

5. „Beyond [cls]: Exploring the True Potential of Masked Image Modeling Representations", IEEE/CVF International Conference on Computer Vision (ICCV), (2025).

Praca analizuje, dlaczego reprezentacje uzyskiwane przez metody Masked Image Modeling, takie jak MAE, często gorzej sprawdzają się bez dodatkowego dostrajania niż reprezentacje modeli joint-embedding. Autorzy pokazują, że problem nie musi wynikać z braku użytecznej informacji w reprezentacjach poszczególnych fragmentów obrazu, lecz ze sposobu, w jaki informacja ta jest agregowana do jednej globalnej reprezentacji. W modelach MIM token [cls] rozkłada atencję stosunkowo równomiernie na wiele fragmentów obrazu, podczas gdy modele joint-embedding bardziej selektywnie skupiają się na istotnych obszarach. Na tej podstawie zaproponowano Selective Aggregation, czyli prosty mechanizm uczący się, które tokeny obrazu powinny mieć największy udział w reprezentacji globalnej. Metoda znacząco poprawia jakość zamrożonych reprezentacji wielu modeli MIM bez konieczności ponownego trenowania samego encodera.

6. „A deep cut into Split Federated Self-supervised Learning", European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), (2024).

Publikacja dotyczy trenowania modeli w środowisku rozproszonym, w którym dane pozostają na urządzeniach klientów, a model jest dzielony pomiędzy klientów i centralny serwer. Autorzy analizują MocoSFL i zauważają, że przesunięcie punktu podziału modelu do głębszych warstw — korzystne z punktu widzenia komunikacji i prywatności — prowadzi do dużego spadku jakości reprezentacji. Analiza wskazuje, że przyczyną jest niespójność pomiędzy aktualizowanym modelem online a jego momentum/teacher odpowiednikiem po synchronizacji klientów. Zaproponowana metoda MonAcoSFL synchronizuje również modele momentum, dzięki czemu zachowana zostaje zgodność obu ścieżek uczenia. Pozwala to trenować model z głębszym podziałem sieci, ograniczyć koszty komunikacji i jednocześnie zachować znacznie lepszą jakość reprezentacji.

**Redakcja Tekstu**

Autoreferat i publikacje są napisane poprawnym angielskim. Wprowadzenie w tematykę badań i podkreślenie kluczowych problemów i wyzwań jest pomocne.

W Autoreferacie znalazłem kilka drobnych błędów językowych:

- „Both approaches benefits" (powinno być „…benefit")

- Suma we wzorze nr 9 (strona 26) indeksowana od 0, a tuż poniżej w zdaniu indeksacja od 1?

- w polskim Streszczeniu:
  wykrywać -> wykrywają
  metodę CASSLE, które poprawia -> metodę CASSLE, którA poprawia

- Hypernetwork Approach to Bayesian MAML. Hypernetwork Approach to Bayesian MAML (strona 30) - zdublowany tytuł

**Uwagi i pytania**

Podczas lektury rozprawy nasunęły mi się następujące pytania:

1. W opisie metody MoCo (strona 24) Autor rozróżnia dwa typy enkoderów: student (online) oraz teacher/momentum. Na końcu opisu pojawiają się jednak parametry nazwane „offline". Czy to jest tożsame z teacher/momentum? Czy jeszcze coś innego?

2. W Tabeli 4 (strona 22), Doktorant klasyfikuje metodę I-JEPA jako MIM (a nie jako JEA)? Dlaczego? W JEPA, jak wskazuje akronim, też mamy joint embedding architecture (wzbogacony o predykcję).

3. W metodzie CASSLE informacja o zastosowanej augmentacji jest wykorzystywana przez conditioned projector podczas pretrainingu, ale nie podczas ewaluacji. Czy Autor badał możliwość wykorzystania tej informacji również podczas inferencji, np. poprzez wykonanie kilku predykcji dla różnych augmentacji obrazu wraz z odpowiadającymi im wektorami augmentacji i agregację uzyskanych wyników? Czy taki mechanizm mógłby dodatkowo poprawić jakość klasyfikacji downstream?

**Konkluzja**

Przedłożony cykl prac oceniam wysoko, prace dotyczą ważnego problemu z dużym potencjałem naukowo-aplikacyjnym, a uzyskane wyniki są przekonywujące. Uważam, że złożona rozprawa mgr Marcina Przewięźlikowskiego spełnia wymagania ustawowe i zwyczajowe stawiane pracom doktorskim i może stanowić podstawę nadania stopnia doktora.

Biorąc pod uwagę uzyskane wyniki i miejsca publikacji, wnioskuję o wyróżnienie rozprawy.

