

dr hab. inż. Paweł Pławiak, prof. PK i IITiS PAN

Kraków, dn. 25.09.2023 r.

Katedra Informatyki,

Wydział Informatyki i Telekomunikacji,

Politechnika Krakowska,

Warszawska 24, 31-155 Kraków

### **Recenzja rozprawy doktorskiej**

**mgr inż. Pawła Staszewskiego**

***pt.: „New methods for image retrieval using deep learning.”***

***(“Nowe metody wyszukiwania obrazów z wykorzystaniem głębokiego uczenia”)***

Promotor: dr hab. inż. Maciej Jaworski, prof. PK

Praca doktorska Pana mgra Pawła Staszewskiego została wykonana na Wydziale Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej pod kierunkiem Pana dra hab. inż. Macieja Jaworskiego, prof. PK. Tematyka rozprawy dotyczy opracowania nowych metod wyszukiwania obrazów z wykorzystaniem metod uczenia głębokiego.

Dysertacja zawarta jest na 87 stronach i napisana w języku angielskim. Rozpoczyna się od wprowadzenia opisującego problem, cele pracy, wkład autora, motywację oraz nowatorstwo. W kolejnym rozdziale zostało opisane konfigurowanie struktury głębokiej sieci neuronowej. Następnie przedstawiono nowe podejście do budowania deskryptora na podstawie warstw konwolucyjnych. Następny rozdział wyjaśnia interpretowalność wiedzy zawartej w głębokiej sieci neuronowej. Na koniec Autor przedstawił podsumowanie wniosków oraz plany przyszłych badań. Bibliografia zawiera 50 pozycji, które zostały dobrane prawidłowo, ze znajomością prezentowanego zagadnienia.

Główną tematyką niniejszej rozprawy doktorskiej jest zagadnienie wyszukiwania podobnych obrazów. W literaturze metody służące do tego zadania

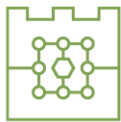


można podzielić na dwie grupy: wyszukiwanie na podstawie metadanych oraz wyszukiwanie na podstawie zawartości obrazu (ang. Content-based image retrieval, CBIR). Algorytm opracowany w ramach pracy należy do tej drugiej grupy. Istnieje wiele różnych algorytmów służących do realizacji zadania CBIR, jednak większość ostatnio proponowanych metod bazuje na sztucznych sieciach neuronowych. Doktorant zaproponował nową metodę konstruowania deskryptorów obrazów, wykorzystującą aktywację neuronów w konwolucyjnych sieciach neuronowych. W literaturze można spotkać wiele tego rodzaju deskryptorów, która dają znakomite wyniki. Skutecznie pozwalają one na wyszukiwanie obrazów podobnych semantycznie, tzn. zawierających obiekty tej samej klasy. W recenzowanej pracy Doktorant podjął się rozwiązania problemu nieco rozszerzonego. Mianowicie, celem pracy było skonstruowanie deskryptora, który umożliwi wyszukiwanie obrazów podobnych nie tylko semantycznie, ale także ze względu na inne cechy, takie jak rozkłady kolorów czy tekstury. Ponadto ostatni rozdział pracy porusza dodatkowy wątek poboczny dotyczący interpretowalności sieci neuronowych, która rozumiana jest przez Autora jako możliwość identyfikowania grupy klas, do której należy dany obraz, na podstawie jego deskryptora.

Deskryptory obrazów, zaproponowane w recenzowanej pracy, skonstruowane są z wartości aktywacji neuronów konwolucyjnej sieci neuronowej. Większość przeprowadzonych eksperymentów bazuje na gotowej sieci neuronowej VGG16, wytrenowanej na popularnym zbiorze danych IMAGENET1M. Nowością metody jest to, że deskryptory składają się z aktywności neuronów zarówno końcowych warstw gęstych (w pełni połączonych) jak i tych pochodzących z warstw konwolucyjnych. Wykorzystanie wszystkich neuronów nie byłoby dobrym podejściem, gdyż większość z nich ma znikomy wpływ na finalną decyzję podejmowaną przez sieć. Dlatego Doktorant zaproponował własną metodę wyboru aktywacji najbardziej istotnych. W ostatniej warstwie konwolucyjnej, tuż przed warstwami gęstymi, są to aktywności przekraczające pewien zadany przez użytkownika próg. W warstwach wcześniejszych wybierane są te aktywacje, które mają największy wpływ na neurony wyselekcjonowane uprzednio w warstwach po nich następujących. Ostatecznie, wybrane aktywacje są uśredniane dla każdej z map cech (ang. feature map) w danej



warstwie. W końcowej postaci deskryptor ma więc stałą liczbę wymiarów niezależnie od przetwarzanego obrazu, co umożliwia ich wzajemne porównywanie. Tak skonstruowany hybrydowy deskryptor, zawierający aktywacje neuronów z obu typów warstw (gęstych i konwolucyjnych) umożliwia wyszukiwanie obrazów podobnych semantycznie, ale także porównywalnych względem innych, tzw. drugorzędowych cech obrazu. Są to m.in. rozkłady kolorów, tekstury czy rodzaj tła. Podobieństwo wyszukiwanych obrazów jest już wyraźnie widoczne w rozdziale prezentującym wstępne rezultaty. Przykładowo, do obrazu przedstawiającego ptaka na tle trawy zostały znalezione obrazy innych ptaków, które także widnieją na tle zielonej roślinności. Można zauważyć różnicę względem obrazów wyszukanych z wykorzystaniem innych deskryptorów znanych z literatury – pozwoliły one także na odszukanie innych obrazów ptaków, ale niekoniecznie w tej samej scenarii. Oczywiście Doktorant nie pozostawia czytelnika z jedyną możliwością oceny swoich rezultatów w sposób wizualny i proponuje także miary ilościowe. Wyznaczane są przykładowo odległości euklidesowe między porównywanymi deskryptorami. Jeśli chodzi o ocenę podobieństw kolorów czy tekstur to należy zauważyć, że jest to zadanie niezwykle trudne, zwłaszcza że jest to problem rzadko (albo wcale) rozważany w literaturze. Jednak także z tym problemem Doktorant poradził sobie bardzo dobrze, proponując dwa rodzaje miar, które pozwalają w sposób ilościowy wyrazić tego rodzaju podobieństwo. W pierwszej mierze wykorzystane zostały estymatory funkcji gęstości prawdopodobieństwa bazujące na funkcjach jądrowych Parzena. Zostały one użyte do estymacji rozkładów kolorów. Estymatory te mają przewagę nad standardowymi histogramami kolorów, ponieważ generują gładkie funkcje. Następnie, na podstawie estymatorów Parzena dla dwóch porównywanych obrazów obliczana jest powierzchnia ich wzajemnego przekrywania się. Drugą zaproponowaną miarą jest odległość euklidesowa pomiędzy macierzami Grama, wyekstrahowanymi ze wszystkich warstw konwolucyjnych. Macierze te są stosowane w technice transferowania stylów (ang. neural style transfer), w pracy natomiast zostały użyte jako miara podobieństwa między teksturami występującymi na obrazach. Zaproponowane w pracy deskryptory zostały porównane względem wspomnianych miar ze swoimi okrojonymi wersjami, bazującymi tylko na warstwach



gęstych lub tylko na warstwach konwolucyjnych, a także z wieloma innymi deskryptorami z literatury, także tymi dostarczanymi przez autorów zbioru IMAGENET1M. Zbadano także możliwość konstruowania deskryptora dla sieci ResNet50. Sieć ta nie zawiera warstw gęstych, dlatego w deskrypcji zastąpiono odpowiadającą im część aktywacji z warstwy Global Average Pooling. Wyniki eksperymentalne są obiecujące i pokazują skuteczność zaproponowanego algorytmu w rozważanym problemie. W przedostatnim rozdziale pracy Doktorant podjął tematykę interpretowalności uzyskiwanych wyników. Sprawnie dokonując syntezy popularnych narzędzi wykorzystywanych w uczeniu maszynowym, takich jak algorytm t-SNE czy DBSCAN, wykazał, że zaproponowane w pracy deskryptory grupują się w klastry, które reprezentują szersze, bardziej ogólne grupy klas. Przykładowo, można zaobserwować klaster ptaków, zawierający klasy odpowiadające poszczególnym gatunkom. Jak napisano w podsumowaniu pracy, rezultat ten otwiera możliwość wykorzystania deskryptorów do usprawnienia procesu klasyfikacji w danej sieci neuronowej, co stwarza pewne perspektywy dalszego rozwoju opracowanej metody.

Od strony redakcyjnej praca napisana jest poprawnie. Jak na pracę doktorską ma dosyć małą objętość, ale jest napisana rzeczowo i konkretnie. Należy tutaj podkreślić precyzyjny formalizm matematyczny zastosowany przez Doktoranta, który umożliwia dokładne zrozumienie intencji Autora. Dzięki temu, odtworzenie algorytmu przez czytelnika potencjalnie zainteresowanego powtórzeniem eksperymentów numerycznych nie powinno stanowić większego problemu. Praca zawiera dość nietypową formę wstępu teoretycznego. W pierwszym rozdziale są po krótko omówione wszystkie ważniejsze zagadnienia wykorzystywane w pracy. W rozdziale drugim natomiast, zamiast tradycyjnego opisu teoretycznego metod głębokiego uczenia, Autor zamieścił coś w rodzaju praktycznego poradnika na temat dobierania odpowiednich parametrów i architektur sieci neuronowych. Opis teoretyczny działania sieci został opisany w literaturze już tak wiele razy, że zastąpienie go w pracy takim poradnikiem wprowadza do niej pewną nowość i sprawia, że wyróżnia się ona na tle pozostałych prac o tej tematyce. Warto także odnotować na plus, że recenzowana praca została napisana w języku angielskim.



Ogólnie praca napisana jest poprawnie i przedstawia nowatorski pomysł Doktoranta na podejście do zagadnienia CBIR, jednak z pewnością kilka aspektów mogłoby zostać poprawionych. Pewien niedosyt może budzić część eksperymentalna pracy.

- Wyniki zostały zaprezentowane w każdym eksperymencie dla kilku wybranych obrazów, brak jest natomiast głębszej analizy statystycznej, która mogłaby wykazać, czy przewaga zaproponowanego podejścia nad innymi jest istotna statystycznie.
- Kolejną kwestią skłaniającą do zastanowienia jest wykorzystanie różnych narzędzi do porównywania rozkładów kolorów w różnych rozdzielach. Przy wynikach wstępnych wykorzystano standardowe histogramy, natomiast w dalszej części zastąpiono estymatorami Parzena. Zmiana ta nie została w pracy skomentowana.
- Jak wspomniano wcześniej, praca generalnie napisana jest poprawnie i zwięźle, jednak miejscami zdarzają się pewne delikatne nieścisłości. Przykładowo, w rozdziale 4 na początku wymienione są w punktach kolejne etapy prowadzące do uzyskania klastrów, nie ma jednak na tej liście algorytmu t-SNE. W dalszej części pracy dowiadujemy się jednak, że algorytm ten jest bardzo istotny w całej procedurze, gdyż umożliwia wyrzutowanie deskryptorów do dwóch wymiarów celem ich zwizualizowania.
- Można mieć także drobną wątpliwość z powodu braku wyraźnego zaznaczenia w tekście celu/tezy pracy. Wprawdzie z zawartych opisów dość jasno wynika jakie Doktorant zdefiniował założenia i do czego dąży, jednak sformułowanie konkretnego celu/tezy przyjęło się jako pewien standard w pisaniu rozpraw doktorskich.

Poniższe pytania i zagadnienia mają charakter dyskusyjny i nie mają charakteru krytyki:

- Chociaż problem postawiony w pracy jest sam w sobie bardzo ciekawy i nowatorski, można zadać pytanie jakie mogłoby być potencjalne praktyczne zastosowanie zaproponowanych deskryptorów. W jakiej dziedzinie



wyszukiwanie obrazów podobnych nie tylko semantycznie, ale też pod względem tła, kolorów i tekstur, byłoby bardziej przydatne niż tylko wyszukiwanie semantyczne?

- Do porównywania podobieństw obrazów wykorzystano różne wspomniane wcześniej miary ilościowe. Czy dałoby się dokonać także podobnej oceny wyników przedstawionych w rozdziale traktującym o interpretowalności?
- W użytym zbiorze IMAGENET1M klasy obrazów ułożone są od razu w pewną hierarchię. Przykładowo, czy można by było w jakiś sposób ilościowo zmierzyć, jak bardzo utworzone klastry odzwierciedlają tę od góry zadaną hierarchię?

Na koniec warto odnotować, że metoda konstruowania deskryptorów zaprezentowana w pracy została opublikowana w bardzo prestiżowym czasopiśmie naukowym IEEE Transactions on Neural Networks and Learning Systems, któremu MEiN przypisuje 200 pkt. W opublikowanym artykule Doktorant jest pierwszym współautorem. Pozostałe treści, w tym zagadnienie interpretowalności, także były publikowane w różnych materiałach konferencyjnych.

Badania Autora przedstawione w rozprawie zostały opisane w następujących 10 publikacjach:

1. **Staszewski, P.**, Jaworski, M., Cao, J., Rutkowski, L. (2021). A new approach to descriptors generation for image retrieval by analyzing activations of deep neural network layers. IEEE Transactions on Neural Networks and Learning Systems, MEiN = 200 pkt., IF = 10.4.
2. Jaworski, M., Rutkowski, L., **Staszewski, P.**, Najgebauer, P. (2021, June). Monitoring of Changes in Data Stream Distribution Using Convolutional Restricted Boltzmann Machines. In International Conference on Artificial Intelligence and Soft Computing (pp. 338-346). Springer, Cham.
3. **Staszewski, P.**, Jaworski, M., Rutkowski, L., Tao, D. (2020, October). Explainable Cluster-Based Rules Generation for Image Retrieval and





- Classification. In International Conference on Artificial Intelligence and Soft Computing (pp. 85-94). Springer, Cham.
4. Woldan, P., **Staszewski, P.**, Rutkowski, L., Grzaneck, K. (2019, June). On Proper Designing of Deep Structures for Image Classification. In International Conference on Artificial Intelligence and Soft Computing (pp. 223-235). Springer, Cham.
  5. Rutkowski, T., Romanowski, J., Woldan, P., **Staszewski, P.**, Nielek, R., Rutkowski, L. (2018, July). A content-based recommendation system using neuro-fuzzy approach. In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.
  6. Rutkowski, T., Romanowski, J., Woldan, P., **Staszewski, P.**, Nielek, R. (2018, June). Towards interpretability of the movie recommender based on a neuro-fuzzy approach. In International Conference on Artificial Intelligence and Soft Computing (pp. 752-762). Springer, Cham.
  7. Korytkowski, M., **Staszewski, P.**, Woldan, P., Scherer, R. (2016, October). Fast computing framework for convolutional neural networks. In 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom) (pp. 118-123). IEEE.
  8. **Staszewski, P.**, Woldan, P., Korytkowski, M., Scherer, R., Wang, L. (2016, June). Query-by-example image retrieval in microsoft SQL server. In International Conference on Artificial Intelligence and Soft Computing (pp. 746-754). Springer, Cham.
  9. Korytkowski, M., Scherer, R., **Staszewski, P.**, Woldan, P. (2015, June). Bagof features image indexing and classification in Microsoft SQL server relational database. In 2015 IEEE 2nd International Conference on Cybernetics (CYBCONF) (pp. 478-482). IEEE.



10. **Staszewski, P.**, Woldan, P., Ferdowski, S. (2015, June). Mobile fuzzy system for detecting loss of consciousness and epileptic seizure. In International Conference on Artificial Intelligence and Soft Computing (pp. 142-150). Springer, Cham.

Niestety nie został podany wkład własny Autora rozprawy, tym bardziej że jest pierwszym autorem tylko w 4 z 10 publikacji. Z 10 publikacji jedna została opublikowana w bardzo prestiżowym czasopiśmie IEEE Transactions on Neural Networks and Learning Systems za 200 pkt. z IF = 10.4, a reszta to referaty konferencyjne co może budzić lekki niedosyt.

W bazie Scopus widnieje 10 dokumentów Autora (1 artykuł i 9 konferencji), sumaryczna liczba cytowań wynosi 73 oraz indeks Hirscha równy 4. Podobnie w bazie Web of Science znajduje się 14 dokumentów, cytowanych sumarycznie 91 razy, a indeks Hirscha wynosi 4. Uważam, że są to dobre wyniki jak na aktualny etap kariery naukowej Doktoranta, potwierdzające istotność prowadzonych badań.

#### **Wniosek Końcowy:**

W końcowej ocenie z przekonaniem stwierdzam, iż praca doktorska **mgr inż. Pawła Staszewskiego pt.: „New methods for image retrieval using deep learning.”** (**“Nowe metody wyszukiwania obrazów z wykorzystaniem głębokiego uczenia”**) odpowiada warunkom określonym w art. 13 Ustawy z dnia 14 marca 2003 o tytule naukowym i stopniach naukowych (Dz.U. z 2017 r. poz. 1789 z późn. zmianami) i wnioskuję o dopuszczenie rozprawy doktorskiej **mgr inż. Pawła Staszewskiego** do publicznej obrony i dalszych etapów przewodu doktorskiego w dyscyplinie Informatyka.

dr hab. inż. Paweł Pławiak, prof. PK i IITiS PAN