Gliwice, 4.10.2023

Prof. dr hab. inż. Marcin Woźniak
Wydział Matematyki Stosowanej
Politechnika Śląska

Sz. P.
Prof. dr hab. inż. Robert Nowicki
Przewodniczący
Rady Dyscypliny Naukowej
Informatyka Techniczna i Telekomunikacja
Politechnika Częstochowska

Recenzja rozprawy doktorskiej mgr inż. Pawła Staszewskiego pt.: „New methods for image retrieval using deep learning" opracowanej pod kierunkiem dr hab. inż. Macieja Jaworskiego, Prof. PK.

W odpowiedzi na pismo R-WIMiI-510-9/18 z dnia 14.07.2023 przedkładam niniejszą recenzję.

Przedstawiona do recenzji praca została napisana w języku angielskim i składa się z 5 głównych rozdziałów obejmujących w sumie 87 stron. W skład bibliografii wchodzi 49 pozycji. Rozprawa zawiera 41 rysunków oraz 11 tabel.

We wstępie pracy Doktorant przedstawił podstawowe zagadnienie Image Retrieval oraz Content-based Image Retrieval (CBIR). Doktorant zaprezentował wybrane zastosowania CBIR, np. w medycynie. Omówiony został zbiór danych Tiny Imagenet

oraz IMAGENET1M wykorzystywane w części eksperymentalnej pracy. Ostatni akapit przedstawia motywacje i główny cel pracy jakim jest konstrukcja deskryptora pozwalającego na wyszukiwanie obrazów podobnych względem wybranych cech obrazu, kolorystyki tła czy tekstury. Wypunktowane zostały także najważniejsze osiągnięcia i nowości zaprezentowane w pracy.

Rozdział drugi jest przedstawiony w formie poradnika praktycznego dot. wybierania odpowiednich architektur konwolucyjnych sieci neuronowych. Porównanych zostało sześć różnych modeli, różniących się typami warstw poolingu, parametrem dylatacji, rozmiarami filtrów konwolucyjnych lub wartością parametru regularyzacji. Skuteczność modeli została przetestowana na zbiorze Tiny Imagenet. Przedstawione zostały wartości różnych miar dla każdego modelu, takie jak dokładność i wartość funkcji straty na zbiorach treningowych i walidacyjnych, f1-score oraz średni czas uczenia.

Rozdział trzeci stanowi centralną część pracy i prezentuje autorski algorytm konstruowania deskryptorów obrazów. Jako sieć bazową Doktorant zastosował VGG16. Do skonstruowania deskryptora wykorzystywano aktywacje z warstw w pełni połączonych oraz konwolucyjnych. Doktorant prezentuje algorytm wyboru najistotniejszych aktywacji w poszczególnych warstwach, które mają największy wpływ na finalną decyzję sieci. Algorytm konstrukcji deskryptora został przedstawiony zarówno za pomocą formalizmu matematycznego jak i pseudokodu. Działanie deskryptorów zostało przetestowane na zbiorze IMAGENET1M i porównane z innymi deskryptorami znanymi z literatury. W eksperymentach numerycznych przetestowano także analogiczny deskryptor bazujący na sieci ResNet50. W celu zmierzenia podobieństw wyszukiwanych obrazów ze względu na drugorzędowe cechy obrazu zaproponowane zostały dwie miary. Pierwszą z nich jest stopień podobieństwa estymatorów rozkładu kolorów, bazujący na funkcjach jądrowych Parzena. Drugą miarą jest odległość euklidesowa między macierzami Gramma, która wyznacza podobieństwo tekstur występujących na obrazach.

Rozdział czwarty przedstawia interpretowalność sieci neuronowych. Doktorant opisuje interpretowalność jako możliwość wykorzystania deskryptorów do pogrupowania obrazów
w klastry odpowiadające klasom nadrzędnym. W tym celu deskryptory dla obrazów ze zbioru IMAGENET1M zostały uśrednione dla każdej z klas. Następnie uśrednione deskryptory zostały rzutowane do przestrzeni dwuwymiarowej z wykorzystaniem

algorytmu t-SNE. W kolejnym kroku zastosowano algorytm DBSCAN do pogrupowania deskryptorów. Uzyskano 19 klastrów odpowiadających grupom klas, np. psy różnych ras, budynki czy pojazdy.

Rozdział piąty, oprócz podsumowania uzyskanych wyników, nakreśla możliwe kierunki dalszych badań. Doktorant przedstawił zastosowanie architektur sieci neuronowych innych niż VGG16 czy ResNet50, a także użycie innych metod redukcji wymiarów i grupowania.

Ocena merytoryczna przedstawionej rozprawy doktorskiej:

Przedłożona rozprawa doktorska mgr inż. Pawła Staszewskiego dotyczy metod uczenia maszynowego dedykowanych analizie obrazów i detekcji cech charakterystycznych. Badania pokazują, że najbardziej skutecznymi w tej dziedzinie są metody oparte o głębokie sieci neuronowe. Wartości aktywacji neuronów sieci dla danego obrazu tworzą wektor deskryptora, który charakteryzuje obraz i umożliwia wyszukiwanie podobnych obrazów w bazie danych. W pracy doktorskiej rozważane jest rozszerzone zadanie, czyli wyszukiwanie obrazów podobnych nie tylko ze względu na klasę abstrakcji, ale także na tzw. drugorzędowe cechy. Dzięki temu wyszukiwane obrazy są do siebie podobne w znacznie szerszym stopniu, nie tylko semantycznie, ale także kolorystycznie.

Głównym rezultatem przedstawionej pracy doktorskiej jest algorytm konstruowania deskryptorów obrazów na podstawie aktywacji głębokiej sieci neuronowej. W odróżnieniu od innych tego typu deskryptorów, rozwiązanie zawiera aktywacje neuronów zarówno z warstw typu fully-connected jak i warstw konwolucyjnych. W przypadku tych ostatnich, zaproponowana została metoda wyboru tylko tych neuronów, które mają największy wpływ na finalną decyzję sieci. Finalnie, aktywacje dla każdej tzw. „mapy cech” w każdej warstwie konwolucyjnej są uśredniane, dzięki temu końcowa postać deskryptora ma zawsze ten sam wymiar.

W ramach pracy przeprowadzono wiele eksperymentów porównawczych na zbiorze treningowym IMAGENET1M. Przebiegają one według schematu: losowo wybieranych jest 10 obrazów ze zbioru i do każdego z nich wyszukiwanych jest po 5 obrazów z najbliższymi deskryptorami. Rezultaty można porównywać zarówno wizualnie na

podstawie zestawionych obrazów, a także przy pomocy trzech miar: odległości L1 między deskryptorami, powierzchni różnic rozkładów kolorów oraz odległości L1 między macierzami Gramma. Rozkłady kolorów zostały wyznaczone przy pomocy metody estymacji bazującej na funkcjach jądrowych Parzena. Macierze Gramma natomiast są swojego rodzaju identyfikatorem stylu danego obrazu i pozwalają na porównywanie podobieństwa tekstur dwóch obrazów.

Porównanie działania deskryptorów zostało przeprowadzone na różnych płaszczyznach. Po pierwsze, porównano działanie zaproponowanego deskryptora, bazującego na aktywacjach obu typów warstw (FC+Conv), z dwoma jego warstwami okrojonymi: pierwszą bazującą na samych aktywacjach z warstw konwolucyjnych (Conv) i drugą bazującą na samych aktywacjach typu fully-connected (FC). Już w tym porównaniu widać, że opracowany deskryptor spełnia stawiane założenia: pozwala wyszukać obrazy podobne zarówno pod względem klasy jak i drugorzędowych cech obrazu, takich jak np. tło. Dla porównania, obrazy wyszukane z użyciem deskryptorów typu FC wyszukują obrazy podobne semantycznie, natomiast tło i tekstury przeważnie zupełnie się różnią. Użycie deskryptorów typu Conv powoduje, że wyszukane obrazy są podobne kolorystycznie, nie zawsze jednak przedstawione na obrazach klasy są ze sobą zgodne.

W pracy porównano zaproponowane deskryptory z tymi, które można znaleźć w literaturze. Jednym z nich jest deskryptor powstający po zastosowaniu operacji „average pooling” na ostatniej warstwie konwolucyjnej. Innym rozwiązaniem jest zastosowanie operacji „pyramid pooling”. Ponadto, w porównaniu zastosowano też deskryptory dostarczane przez autorów zbioru IMAGENET1M. Autorski deskryptor osiągnął najlepsze wyniki biorąc pod uwagę wszystkie wspomniane powyżej miary, aczkolwiek przy porównaniu odległości między macierzami Gramma deskryptor bazujący na „average pooling” sprawdzał się porównywalnie dobrze.

Wspomniane deskryptory bazowały na popularnej głębokiej sieci neuronowej VGG16. W celu zbadania uniwersalności zaproponowanej metody skonstruowano także podobny deskryptor dla sieci ResNet50. Sieć ta nie posiada warstw typu „fully connected”, dlatego ta część deskryptora musiała zostać zastąpiona przez aktywacje warstwy „global average pooling”. Deskryptory oparte o sieć ResNet50 nie dawały tak dobrych rezultatów jak te zbudowane dla sieci VGG16, niemniej jednak dalej wypadły bardzo dobrze na tle wspomnianych wcześniej deskryptorów z literatury.

W ostatnim rozdziale pracy podjęto dodatkowy wątek badawczy dotyczący interpretowalności sieci neuronowych. Stosując redukcję wymiaru deskryptorów metodą t-SNE a następnie grupując je algorytmem DBSCAN wykazano, że deskryptory umożliwiają grupowanie obrazów w większe podzbiory, składające się z obrazów należących do podobnych klas. Uzyskane wyniki nie zostały zmierzone w sposób ilościowy, należy je rozpatrywać w kategorii rezultatów wstępnych.

Atutem pracy jest udostępnienie przez Doktoranta kodów źródłowych. Pozwala to na samodzielne powtórzenie eksperymentów i porównanie uzyskanych rezultatów. Praca napisana jest poprawnie i zwięźle. We wstępie bardzo wyraźnie wypunktowano wszystkie nowatorskie aspekty pracy. Proponowane algorytmy opisane są zwięźle i zrozumiale za pomocą formalizmu matematycznego. Algorytm konstruowania deskryptora jest także wyrażony w formie schematu blokowego. Praca została napisana w języku angielskim. Dzięki temu uniknięto konieczności tłumaczenia nazw i terminów z dziedziny uczenia maszynowego, które zdecydowanie zdominowane jest przez literaturę angielskojęzyczną.

Po przeczytaniu przedstawionej rozprawy doktorskiej nasuwają się następujące pytania
i sugestie:

i. Wyniki badań niestety nie zostały należycie przedstawione za pomocą analizy statystycznej. Zatem ciężko jest zweryfikować przewagę opracowanej metody nad innymi. W rezultacie może się pojawiać pytanie o wpływ np. przypadkowego doboru obrazów testowych.
ii. Do przedstawionych miar oceny warto dodać odpowiedni komentarz. Czy użyte w pracy miary (stopień przekrywania estymatorów rozkładu kolorów i odległość L1 między macierzami Gramma) były już gdzieś wcześniej stosowane, czy są autorskimi propozycjami Autora pracy? Czy miarę opartą o macierze Gramma można zmodyfikować tak, żeby w różnym stopniu uwzględniała różne warstwy konwolucyjne (w obecnej formie wszystkie warstwy uwzględniane są z tą samą wagą). Czy taka zmiana spowodowałaby, że miara byłaby bardziej adekwatna?
iii. Doktorant nie wyjaśnił w jakim sensie zastosował termin „interpretowalność”. Zwykle rozumie się go jako możliwość zrozumienia przez użytkownika jak model uczenia maszynowego podejmuje decyzję. Natomiast z pracy wynika, że Doktorant rozumie interpretowalność jako możliwość przewidzenia do jakiej

nadrzędnej klas abstrakcji należy dany obraz. Czy te rozumowania są ze sobą powiązane?

iv. W rozdziale 4 deskryptory konstruowane są inaczej, niż we wcześniejszych rozdziałach. Zawierają aktywacje tylko z ostatniego bloku konwolucyjnego, natomiast we wcześniejszych wersjach pod uwagę brane były wszystkie warstwy. Dlaczego zdecydowano się na taką modyfikację? Czy zastosowanie pełnej wersji deskryptora polepszyłoby wyniki uzyskane w rozdziale 4?

Ocena formy, taksonomii, języka i edycji pracy:

Przestawiona rozprawa doktorska jest napisana poprawnym językiem. Edycja i forma dokumentu są poprawne. Rozprawę napisano w języku angielskim, a na jej końcu dołączono streszczenie w języku polskim. Przedstawione rysunki, tabele i wzory zostały ponumerowane stosownie do przynależności do rozdziałów opracowania.

Ocena przedstawionego dorobku naukowego:

W przedstawionej pracy Doktorant zawarł referencje do trzech prac naukowych, których jest współautorem. Są to artykuły opublikowane w wysoko punktowanych czasopismach tematycznych dla przedstawionego zagadnienia oraz wolumenach konferencyjnych.

Warto zauważyć, że najważniejsze wyniki pracy, w tym algorytm konstruowania deskryptorów, został opublikowany w pracy w czasopiśmie IEEE Transactions on Neural Networks and Learning Systems, w której mgr inż. Paweł Staszewski jest pierwszym autorem.

Przedstawiony dorobek naukowy Doktoranta pokazuje zaangażowanie w prace naukowe, które zostały uznane w międzynarodowym środowisku naukowym. Moim zdaniem zaprezentowany dorobek pozwala stwierdzić obeznanie Doktoranta z dyscypliną prowadzenia badań naukowych, analizą wyników i podejmowaniem wniosków z prowadzonych badań naukowych.

Podsumowanie:

Przedstawiona do recenzji rozprawa doktorska porusza temat zastosowania modeli uczenia maszynowego do analizy obrazów. Uczenie maszynowe za pomocą modeli głębokich sieci neuronowych jest tematem wielu aktualnych opracowań naukowych. Jest to temat ważny z punktu widzenia rozwoju informatyki. Doktorant rozwiązał problem naukowy dotyczący opisu danych wejściowych przy pomocy autorskiego modelu i wykonał testy dla podstawowych architektur głębokich sieci neuronowych. Wyniki pokazały, że zastosowanie proponowanego modelu przetwarzania pozwala na zwiększenie wydajności.

Na podstawie przedstawionej powyżej oceny stwierdzam, iż rozprawa doktorska mgr. inż. Pawła Staszewskiego pt.: „New methods for image retrieval using deep learning" spełnia warunki określone w art. 13 ust 1 ustawy z dnia 14 marca 2003r. o stopniach naukowych i tytule naukowym (Dz. U. z 2017 r. poz.1789) stosowane w postępowaniach o nadanie stopnia doktora. **Wnoszę o przyjęcie w/w rozprawy doktorskiej i dopuszczenie jej do publicznej obrony w dyscyplinie informatyka.**