Warszawa, 12.01.2023

dr hab. inż. Tomasz Trzciński, prof. PW i UJ
Instytut Informatyki Politechniki Warszawskiej
Katedra Uczenia Maszynowego Uniwersytetu Jagiellońskiego

RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY DYSCYPLINY NAUKOWEJ INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI ŚLĄSKIEJ

Tytuł rozprawy: Optimization of deep learning network architectures for hyperspectral data classification

Autor rozprawy: Kamil Książek

1 Analiza strony merytorycznej rozprawy

1.1 Obszar problemowy

Recenzowana rozprawa doktorska dotyczy zagadnień związanych z wykorzystaniem metod głębokiego uczenia, w tym w szczególności modeli sztucznych sieci neuronowych, do analizy i klasyfikacji obrazów hiperspektralnych. Główną osią rozprawy jest rozwiązanie problemów związanych z dotychczasowo stosowanymi metodami służącymi do tej analizy, takimi jak mieszanie spektralne różnych substancji w ramach jednego piksela obrazu czy niestabilność treningu sieci neuronowych i zależność uzyskiwanych rezultatów od metod inicjalizacji wag modelu. Uważam, że proponowana problematyka pracy jest aktualna, chociaż mieści się raczej w zastosowaniach metod uczenia maszynowego do konkretnego, dosyć niszowego problemu obrazów hiperspektralnych, bez wskazania na możliwą generalizację ponad ten specyficzny typ danych - prezentowane w rozdziale czwartym eksperymenty ograniczają się do dosyć niewielkiego i mało reprezentatywnego zbioru MNIST.

Postawiony wyżej problem badawczy skłania Doktoranta do postawienia hipotezy, iż *optymalizacja architektury głębokiej sieci neuronowej oraz metody reinicjalizacji wag tejże sieci polepszają wyniki dla danych hiperspektralnych*. Moim zdaniem, hipoteza dotycząca metod reinicjalizacji wag zostaje zweryfikowana, choć niestety wyłącznie częściowo, tzn. w scenariuszu dotyczącym architektury autoenkodera i w zadaniu rozkładu czy dekompozycji spektralnej (*ang.* spectral unmixing). Dla pełni weryfikacji hipotezy, Doktorant powinien również sprawdzić czy wprowadzenie nowych metod reinicjalizacji wag w zadaniu dekompozycji spektralnej rzeczywiście pozytywnie wpływa na rezultaty osiągane przez sieć klasyfikującą dane hiperspektralne - wszakże z tego punktu wychodził on w rozdziale drugim. Jeśli zaś chodzi o pierwszą część hipotezy, tj. dotyczącą optymalizacji architektury sieci neuronowej, moim zdaniem weryfikacja tej tezy jest dosyć pobieżna, tzn. polega na przeglądowym potraktowaniu istniejących i zaproponowanych w literaturze architektur sieci neuronowych do nowego typu danych, jakim są dane hiperspektralne. Nie wyczerpuje to jednak, w moim przekonaniu, definicji słowa *optymalizacja*, gdyż takowa wymaga nie tylko wyboru odpowiedniej architektury, ale również systematycznego podejścia do jej treningu, konstrukcji, wyboru hiperparametrów oraz ich optymalizacji. Powyższe przemyślenia skłaniają mnie również do konstatacji, że tytuł recenzowanej rozprawy doktorskiej rozmija się z jej treścią - analizując realną kontrybucję zaproponowaną przez Doktoranta skłaniałbym się raczej do wskazania nowych metod reinicjalizacji architektury autoenkodera w problemie dekompozycji spektralnej, jako głównego tematu rozprawy, a co za tym idzie sugerowałbym modyfikację jej tytułu, która ten temat lepiej ujmuje.

Jako cel pracy Autor stawia sobie rozwiązanie specyficznego problemu klasyfikacji substancji będących pochodną krwi przy wykorzystaniu modeli sieci neuronowych na bazie zdjęć hiperspektralnych. To określenie jasno i klarownie wskazuje obszar zainteresowań badawczych Doktoranta, jak również główną oś publikacji, wchodzących w skład rozprawy. Mieści się ono w dziedzinie informatyki technicznej,

jednak tak jak opisałem to wyżej, stanowi raczej dosyć niszowy i specyficzny aspekt jej zastosowania do konkretnego typu danych.

Pierwszy rozdział rozprawy to wstęp, który opisuje motywację podjęcia pracy badawczej w określonym wyżej kierunku, krótko omawia aktualnie istniejące rozwiązania oraz ich ograniczenia. W rozdziale tym Autor opisuje również wkład w rozwój nauki opisany w dalszej części rozprawy, jak również stawianą hipotezę badawczą. Rozdział pierwszy kończy się krótkim streszczeniem kolejnych rozdziałów pracy, w tym tych będących rozszerzeniem publikacji tworzących główny wkład merytoryczny rozprawy.

W rozdziale drugim Autor przedstawia przegląd metod uczenia maszynowego w ich zastosowaniu do problemu klasyfikacji plam po krwi na bazie zdjęć hiperspektralnych. W tym celu Autor wybiera różnego rodzaju architektury sieci głębokich, w tym sieci splotowe oraz rekurencyjne, jak również metody klasyczne takie jak maszyny wektorów nośnych. Rozdział ten zawiera interesującą, lecz dosyć odbiegającą od głównego trzonu zainteresowań informatyki technicznej sekcję dotyczącą chemicznej analizy roztworów krwi oraz ich rozkładu, natomiast rozumiem jej potrzebę wynikającą z wpływu tego procesu na uzyskiwane obrazy hiperspektralne. Następnie Doktorant przedstawia wybrane architektury sieci głębokich, a także opisuje scenariusz ich ewaluacji, jak również wyniki tejże ewaluacji. Rozdział kończy konkluzja dotyczącą wyboru rozwiązania dającego najlepsze rezultaty oraz wskazanie, iż bardziej skomplikowane architektury niekoniecznie dają najlepsze rezultaty. Autor wskazuje również w konkluzjach pojawienie się problemu stabilności treningu sieci neuronowych a także mieszania spektralnego, które wg Autora wpływa na jakość uzyskiwanych rezultatów.

W związku z powyższym w rozdziale trzecim Doktorant zajmuje się architekturą autoenkodera trenowanego w celu dekompozycji spektralnej i niwelowania problemu zachodzenia na siebie widm sygnałów w obrazach hiperspektralnych. W szczególności przeprowadza on testy statystyczne mające na celu wskazanie zależności pomiędzy sposobami inicjalizacji wag a stabilnością uzyskiwanych rezultatów. Konkluzje płynące z tego rozdziału wskazują, że sposób inicjalizacji wag ma znaczący wpływ na osiągane przez sieć rezultaty dekompozycji spektralnej, co, w domyśle, pośrednio może się też przenosić na rezultaty klasyfikacji uzyskiwane przez architektury omawiane w rozdziale drugim, choć takiej analizy nie przeprowadzono.

W rozdziale czwartym, Doktorant proponuje oryginalny pomysł reinicjalizacji wag sieci w zależności od zidentyfikowanych *martwych* neuronów, zdefiniowanych przez Autora jako takie które dają na wyjściu wyniki zerowe. Zaproponowana metoda jest zewaluowana na zbiorach danych hiperspektralnych jak również na pojedynczym zbiorze niezawierającym tych danych, tj. MNIST. Należy zauważyć, że w tym przypadku ewaluacja podobnie jak w rozdziale trzecim dotyczy wyłącznie architektury autoenkodera i wyłącznie zadania dekompozycji spektralnej, a nie postawionego na początku rozprawy celu klasyfikacji obrazów.

Rozprawę kończy rozdział piąty, w którym Autor przedstawia wnioski końcowe, podsumowuje przedstawione w poprzednich rozdziałach wyniki, jak również wskazuje możliwe dalsze prace nad proponowanymi metodami.

Rozprawa zawiera reprezentatywną dla poruszanej problematyki bibliografię, która w większości przypadków właściwie ilustruje omawiane zagadnienia i dokumentuje wkład własny Autora.

1.2 Ocena wyników oraz stopnia ich oryginalności

Recenzowana rozprawa dotyczy niszowej problematyki analizy obrazów hiperspektralnych, które ze względu na swoją specyfikę wymagają odpowiedniego dostosowania proponowanych w literaturze metod, w tym metod uczenia głębokiego. Oryginalność osiągniętych wyników ma więc w największym zakresie charakter aplikacyjny, tj. dotyczy adaptacji metod zaczerpniętych z literatury do innego typu danych. Poniżej przedstawiam opis przedstawionych w rozprawie wyników, które moim zdaniem można uznać za najistotniejszej wraz z oceną ich oryginalności:

- **ewaluacja istniejących modeli sieci neuronowych w zadaniu klasyfikacji obrazów spektralnych**: Przedstawiona w rozdziale drugim ewaluacja jest *de facto* pracą przeglądową i zawiera minimalny wkład jeśli chodzi o oryginalność opisywanych eksperymentów, polegający głównie na dostosowaniu implementacji do nowego typu danych oraz analizie wyników.
- **ewaluacja wpływu inicjalizacji wag na rezultaty osiągane przez autoenkoder w zadaniu dekompozycji spektralnej**: Przedstawiona w rozdziale trzecim ewaluacja przedstawia porównanie wyników osiąganych przez autoenkoder w zadaniu dekompozycji spektralnej na bazie testów statystycznych, zawiera również konkluzje dotyczące zaobserwowanego zjawiska tzw.

znikającego gradientu. Oryginalność tej części pracy jest wyższa niż rozdziału drugiego, jednak w przeważającej mierze zawiera ona opis i definicje istniejących i proponowanych wcześniej modeli oraz metryk ewaluacyjnych.

- **wprowadzenie metody reinicjalizacji wag na podstawie liczby tzw. martwych neuronów**: Przedstawiona w rozdziale czwartym metod identyfikacji tzw. martwych neuronów oraz bazujący na niej algorytm reinicjalizacji wag jest w moim przekonaniu jedynym istotnie oryginalnym osiągnięciem Doktoranta opisanym w recenzowanej rozprawie doktorskiej. Do pełni obrazu wpływu tejże metody na jakość rezultatów osiąganych w ramach postawionego celu brakuje jednak porównaniu wyników metod prezentowanych w rozdziale pierwszym po zastosowaniu tejże metody reinicjalizacji wag proponowanej w rozdziale czwartym.

Należy podkreślić, że część rozprawy pochodzi z publikacji wieloautorskich, w których główną i dominującą rolę pełnił Doktorant. Jego wkład jest jasno opisany i znajduje odzwierciedlenie w pozycji Doktoranta na liście autorów. prac, o czym świadczy również jego pierwsza pozycja na liście autorów.

Zauważone niedoskonałości dysertacji dotyczą kwestii merytorycznych oraz logicznego powiązania elementów rozprawy. W poniższym zestawieniu zwracam uwagę na te dostrzeżone niedostatki i proszę Doktoranta o ustosunkowanie się do nich.

- str. 11, ostatni wiersz: "If one of the innermost hidden layers has two or three neurons then the autoencoder can be used for visualization of the dataset, often with data separated into clusters according to the different classes." - Założenie dotyczące 2 lub 3 neuronów nie jest konieczne, wystarczy przecież zastosowanie znanych algorytmów wizualizacji przestrzeni wielowymiarowych takich jak t-SNE (L. van der Maaten, and G. Hinton, 2008) lub UMAP (McInnes et al., 2018).

- str. 56, przedostatni wiersz: "We noticed a similar phenomenon of undertrained models in the case of autoencoders for hyperspectral unmixing." - Brakuje rezultatów przedstawionych w pracy, które wskazują na ten fenomen, a jest on kluczowy w ciągu logicznym prowadzącym czytelnika przez całą dysertację.

- str. 89, w. 13: "We identified that for some models the gradient has vanished during the first iterations of the training session. This was due to the dead activations' and dead neurons' phenomena which are related to the ReLU activation function [97, 131]." - To zdanie wymaga wcześniejszego zdefiniowania martwych aktywacji i neuronów, którego brakuje. Co więcej, nie jest do końca jasne jak wskazana aktywacja typu ReLU, która jest nieliniowa, ma się do założenia ze str. 45, w. 10 dot. wyboru liniowych modeli.

1.3 Zagadnienia dyskusyjne

Poniższe uwagi dotyczą ogólniejszych kwestii poruszonych w rozprawie i nie odnoszą się bezpośrednio do treści pracy. Niemniej jednak liczę na analizę tych zagadnień i odniesienie się do nich przez Doktoranta.

- str. 6, w. 4: "CNNs (...) are designed to process multidimensional data like 2D images or 3D hyperspectral images" - jeśli CNNy zostały zaprojektowane do pracy z hiperspektralnymi obrazami 3D to czy nie powinny uzyskiwać lepszych rezultatów niż metody prostsze? Moim zdaniem sieci splotowe zostały zaprojektowane przede wszystkim w celu wykorzystania lokalnych korelacji pomiędzy sąsiadującymi w wektorach wartościami. Czy takie lokalne korelacje zachodzą w przypadku danych hiperspektralnych?

- str. 11, w. 17: "the autoencoder with at least one hidden layer and with a sufficiently large number of units can reconstruct original data with any non-zero error value" - jak zdefiniowana jest "wystarczająco duża liczba jednostek"?

- str. 17, w. 12. "The problem of hyperparameter optimization or methods like transfer learning, while important for overall network effectiveness, we do not treat it as a part of architecture optimization. We consider architecture optimization as a process of the choice of the neural network type (e.g. linear, convolutional, recurrent ones) and its individual elements, i.e. the number, type and order of layers, etc." - Taka definicja optymalizacji jest dosyć uboga, co więcej wskazuje raczej na dyskretny jej charakter, podczas gdy współczesne uczenie maszynowe dąży raczej do

rozwiązań typu AutoML, które samodzielnie przy wykorzystaniu technik meta-learningu dobierają zarówno moduły architektury, jak i ich parametry. Dlaczego Doktorant zdecydował się na tak znaczące uproszczenie definicji?

- str. 18, w. 28: "the training and the test set come from the same image. It is assumed that the training samples are a good representation of the data distribution." - Procedura ewaluacyjna zakładająca pochodzenie próbek trenujących i testujących z tego samego obrazka, a więc poniekąd z tej samej albo podobnej dystrybucji, w naturalny sposób budzi zastrzeżenia co do potencjalnego przecieku informacji. Rozumiem, że jest to standard ewaluacji danych hiperspektralnych, jednak czy Doktorant może przedstawić swoją opinię na temat potencjalnych ryzyk takiego sposobu ewaluacji? W szczególności Doktorant pisze dalej, że scenariusz ewaluacji HIC łamie *podstawową* zasadę stosowaną w uczeniu dotyczącą niezależności próbek (i.i.d.), podczas gdy np. metody uczenia ciągłego (*ang. continual learning*) takiego założenia nie przyjmują. Czy mogą one wg Doktoranta znaleźć zastosowanie w przypadku danych hiperspektralnych?

- str. 38, w. 25: "For frame images, all architectures achieved high efficiency, i.e. between 96.8% and 99.9%" - tak wysokie wyniki wskazują na nasycenie metod ewaluowanych na zbiorze testującym - czy Doktorant rozważył możliwość utrudnienia klasyfikacji w takim przypadku, np. poprzez dodanie szumu czy innych sposobów zanieczyszczenia próbek?

- str. 40, w. 29: "It proves that individual training sessions with this architecture and the selected set of hyperparameters can diverge and lead to degenerate results. Although it was not the only situation in which such a phenomenon occurred, most runs lead to high-performing results" - sformułowanie "proves" bez podania teoretycznego dowodu jest dosyć ryzykowne. Co więcej, w kolejnym rozdziale Doktorant wnioskuje o braku stabilności architektur autoenkodera na podstawie *de facto* innej architektury i innego zadania (prezentowanego w rozdziale drugim). Czy Autor może się odnieść do tego powiązania dwóch odrębnych architektur?

- str. 45, w. 10: "we decided to focus on simple linear architectures because, based on the results from this chapter, we can conclude that they may achieve high performance. Their analysis should be simpler than in the case of architectures with many linear or convolutional layers." - wybór prostszej architektury oczywiście ułatwił pracę Doktorantowi, ale czy ma to również uzasadnienie praktyczne? W dzisiejszych realiach stosuje się głównie modele wykorzystujące nieliniowość.

2 Analiza strony formalnej rozprawy

2.1 Ocena układu pracy i redakcji manuskryptu

Recenzowana rozprawa doktorska napisana jest w języku angielskim i obejmuje w kolejności: streszczenie w języku angielskim, streszczenie w języku polskim, podziękowania, listę użytych skrótów oraz symboli, pięć rozdziałów zasadnicze (w tym "Introduction" oraz "Conclusions"), a także bibliografię. Praca liczy 170 stron.

Obfita bibliografia liczy 177 uporządkowane alfabetycznie pozycje. W rozprawie cytowane są dwie prace własnych Autora (wszystkie współautorskie). Prace te, będące jednocześnie zawartością rozdziałów drugiego i trzeciego, pochodzą z czasopisma *MDPI Sensors* o IF 3.56 i 100 pkt. MEiN, oraz materiałów konferencyjnych lokalnej włoskiej konferencji ICIAP (70 pkt. MEiN). Spośród znajdujących się w bibliografii prac większość ukazała się w przeciągu ostatnich 5 lat co świadczy o dobrym umiejscowieniu tematyki rozprawy w aktualnym w skali światowej nurcie badań. Bibliografia rozprawy, po włączeniu w nią utworów cytowanych w ramach załączonych artykułów, nie budzi zastrzeżeń od strony merytorycznej, a jej redakcja jest staranna i nie dostrzegłem w niej żadnych błędów.

Układ rozprawy jest prawidłowy, jest ona także starannie opracowana pod względem edytorskim. Jedyne zastrzeżenie budzi umiejscowienie wykresów oraz tabel na końcu rozdziałów co znacząco utrudnia zapoznawanie się z nimi w trakcie czytania, na przykład Tabela 2.8 w rozdziale drugim jest referowana na stronie 41, podczas gdy znajduje się ona na stronie 54. Dodatkowo tabele często wykraczają poza stosowane na innych stronach marginesy, co może utrudnić czytelność pracy w przypadku druku. Użyta terminologia jest właściwa dla obszaru problemowego rozprawy w zakresie informatyki technicznej.

2.2 Uwagi szczególne

Tekst rozprawy jest w przeważającej większości poprawny pod względem językowym i stylistycznym. Podczas lektury zauważyłem nieliczne błędy redakcyjne i pomyłki wymagające korekty, np:

- str. 2, opis rysunku 1.1: "Jasper Ridge dataset" - brak referencji.
- str. 4, w. 1: "endmembers and fractional abundances" - terminy użyte bez wprowadzenia.
- str. 21, w. 16: "Authors" - niepotrzebna wielka litera.
- str. 45, w. 10: "underperfomed" - underperforming.

3 Konkluzja

Pomimo wskazanych powyżej uwag i zastrzeżeń, uważam, że recenzowana dysertacja Pana mgr. Kamila Książka **spełnia w sposób dostateczny** stawiane rozprawom doktorskim przez *Ustawę z 20 lipca 2018 roku (Dz. U. 2018 poz. 1668) Prawo o szkolnictwie wyższym i nauce*, ponieważ zawiera oryginalną koncepcję i rozwiązuje istotny problem aplikacyjny w dziedzinie informatyki technicznej. Uzyskane wyniki eksperymentalne w wystarczającym stopniu dokumentują poprawność proponowanej koncepcji oraz skuteczność działania jej implementacji. Na tej podstawie **wnioskuję o dopuszczenie mgr. Kamila Książka do dalszych kroków procedury uzyskania stopnia doktora nauk technicznych.**

Tomasz Trzciński