

Tytuł, stopień, imię i nazwisko
prof. dr hab. inż. Bogdan Kwolek

WYDZIAŁ AUTOMATYKI,
ROBOTYKI I I ELEKTROTECHNIKI
POLITECHNIKA POZNAŃSKA

data 12.09.2023 r.

15-09-2023

WPŁYNEŁO

PRZEWODNICZĄCY RADY DYSCYPLINY
Automatyka, Elektronika, Elektrotechnika
i Technologia Kosmiczne

prof. dr hab. inż. Wojciech Szelaąg

RECENZJA ROZPRAWY DOKTORSKIEJ

przygotowana dla Rady dyscypliny Automatyki, Elektroniki, Elektrotechniki
i Technologii Kosmicznych Politechniki Poznańskiej

**Tytuł rozprawy: Restricted Boltzmann Machine as a binary image descriptors processor
and its application in a mobile robot for scene recognition**

Autor rozprawy: mgr inż. Szymon Sobczak

Promotorzy: dr hab. inż. Dariusz Pazderski, prof. PP, dr inż. Rafał Kapela

1. Przedstawiona mi do recenzji rozprawa doktorska poświęcona jest rozpoznawaniu obrazów w systemach wizyjnych robotów opartych o sztuczne sieci neuronowe. Zasadniczym celem prac badawczo-eksperymentalnych było opracowanie i praktyczne zweryfikowanie rozwiązań pozwalających na poprawę skuteczności rozpoznawania obrazów przy zachowaniu ograniczonego zapotrzebowania na moc obliczeniową. Wspomniany cel osiągnięto poprzez zaproponowanie nowatorskiej metody z dodatkową fazą przetwarzania danych, która poprzedza klasyfikator neuronowy. Dzięki wytrenowaniu bez nadzoru ekstraktora cech na niepoetykietowanych danych, niewnoszącego przy tym znaczących narzutów obliczeniowych i opóźnień w torze przetwarzania, uzyskano możliwość użycia prostszych i mniej wymagających obliczeniowo klasyfikatorów neuronowych, a co za tym idzie konieczność etykietowania większych rozmiarowo danych trenujących. W dysertacji pokazano eksperymentalnie, że wspomniane podejście oferuje szereg cennych korzyści dla systemów wizyjnych robotów. Prace badawczo-eksperymentalne obejmowały projektowanie oraz budowę efektywnych obliczeniowo i nieustępujących zarazem skutecznością klasyfikacji sieciom neuronowym głębokim, także dotrenowywanym w podejściu znanym pod nazwą transfer-learning. Zasadnicza część prac dotyczyła udoskonaleniu znanych metod ekstrakcji cech binarnych, a następnie przetwarzania wyekstrahowanych danych binarnych tak aby wydobyć z nich najważniejsze informacje i zależności między nimi. Wynikiem wspomnianych operacji jest tensor wartości rzeczywistych, reprezentujących wektory cech dla każdego piksela obrazu, który podawany jest wejście klasyfikatora neuronowego. Przedmiotem badań i udoskonalień były w szczególności metody ekstrakcji cech oparte na LBP oraz Ograniczone Maszyny Boltzmann RBM. Autor udoskonalił lokalny deskryptor binarny LBP oraz zaproponował podejście umożliwiające poprawę efektywności rozpoznawania obrazów przez system wizyjny robota przy zachowaniu ograniczonego zapotrzebowania na moc obliczeniową. Zaproponowany deskryptor CLBP rozszerza informację zakodowaną w wektorach cech o zależności między barwami z lokalnych sąsiedztw pikseli. W proponowanym podejściu dla cech wydzielonych tym sposobem przez wstępną warstwę wyznaczone są bardziej złożone zależności w drugiej

warstwie przez sieć RBM trenowaną na niepoetykietowanych danych. Wspomniane rozwiązania są ogólne i mogą być użyteczne dla szerszej klasy problemów wymagających szybkich i efektywnych modułów rozpoznających obrazy. Na dorobek Autora składa się także oparta na wspomnianym podejściu metoda rozpoznawania na zaszumianych obrazach, a także Autorska metryka, która umożliwia nie tylko porównywanie obrazów, ale także szacowanie podobieństwa zestawów obrazów i zadecydowanie, czy zasadne byłoby użycie danej sieci neuronowej, która parametryzowana jest przez pewien zestaw parametrów, do klasyfikacji innych obrazów (np. ujęciach tej samej sceny) bez dodatkowego dotrenowania, przy jednoczesnym utrzymaniu wymaganej skuteczności. Wyniki prac eksperymentalnych uzyskane w kilku zadaniach typowych dla systemów wizyjnych robotów mobilnych potwierdziły znaczący potencjał proponowanych rozwiązań. Autor zaproponował oryginalne rozwiązania, przygotował kompletne implementacje dla systemów wbudowanych z akceleracją obliczeń na GPU, a także zweryfikował je w dwudziestu eksperymentach z różnymi scenariuszami i jasno postawionymi celami, które każdorazowo weryfikowano doświadczalnie. Tezy wyrażają się w przekonaniu, iż (i) można zwiększyć zdolność generalizacji klasyfikatora neuronowego za pomocą proponowanej metody przetwarzania wstępnego, opartej na deskryptorach binarnych i rekurencyjnej sieci neuronowej, (ii) proponowane wstępne przetwarzanie umożliwia zmniejszenie złożoności klasyfikatora neuronowego bez istotnego spadku dokładności klasyfikacji, (iii) proponowane przetwarzanie wstępnego może skutkować mniejszą wrażliwością rezultatów klasyfikacji przez sieć neuronową na zniekształcenia obrazu, (iv) wstępne przetwarzanie umożliwia wyznaczenie wartości metryk na potrzeby wyznaczenia podobieństwa i odległości pomiędzy dwoma zbiorami obrazów. Tezy pracy są konsekwentnie rozwijane, a następnie weryfikowane empirycznie. Tematyka podjęta w rozprawie jest w pełni uzasadniona pod względem poznawczym, a na płynące z niej wnioski istnieje zapotrzebowanie w obszarze robotyki i sztucznej inteligencji. Wyniki badań eksperymentalnych są rzetelne i przekonujące.

2. Rozprawa napisana jest w języku angielskim. Praca składa się z siedmiu zasadniczych części w których wprowadzono w problematykę pracy, omówiono techniki inżynierii cech obrazów, zaprezentowano sieci RBM, omówiono przetwarzanie i klasyfikację obrazów w oparciu o binarne deskryptory obrazów i RBM, zaprezentowano wyniki badań eksperymentalnych, omówiono wyniki uzyskane w badaniach eksperymentalnych z użyciem robota oraz podsumowano rozprawę. Rozprawa liczy łącznie 189 stron i jej treść podzielono na sześć rozdziałów, podsumowanie oraz trzy dodatki. Bibliografia liczy łącznie 199 pozycji. Streszczenia pracy przedstawiono zarówno w języku angielskim jak i polskim. Na wstępie zawarto wprowadzenie do problematyki pracy, omówiono zagadnienia badawcze i proponowane rozwiązania, a także zawartość rozprawy. W rozdziale drugim omówiono techniki inżynierii cech obrazów, ekstrakcję cech obrazu w oparciu o metody nienadzorowane, w tym w oparciu sieci neuronowe Hopfielda, techniki oparte o LBP, a następnie zaprezentowano autorski CLBP. W rozdziale trzecim zaprezentowano kolejno sieci RBM, uczenie sieci RBM, a następnie sieci DBM (Deep Boltzman Machines). W kolejnym rozdziale szczegółowo omówiono autorski algorytm wstępnego przetwarzania danych w oparciu o CLBP i RBM, w tym także własną implementację, a następnie potencjał sieci RBM w charakterze procesora i agregatora cech binarnych, wprowadzając kolejno trzy architektury: (1) CD-KNN, opartą na metryce CD (Contrastive Divergence), wyszukiwaniu selektywnym oraz algorytmie KNN do klasyfikacji, (2) CD-CNN, opartą na gęstych cechach i przestrzeni cech na metryce CD,

z użyciem CNN-MLP do klasyfikacji, (3) HS-CNN, opartą na gęstych cechach i ukrytej przestrzeni z wielu warstw RBM, z użyciem CNN-MLP do klasyfikacji, a w ostateczności potencjał RBM do wyznaczania podobieństw między obrazami z dwóch zbiorów. W kolejnym piątym rozdziale omówiono wyniki badań eksperymentalnych, które uzyskano na kilku ogólnie dostępnych zbiorach danych. W szóstym rozdziale zaprezentowano wyniki uzyskane w badaniach eksperymentalnych na obrazach pozyskiwanych przez kamerę pokładową robota mobilnego. Badania eksperymentalne realizowano celem pokazania potencjału proponowanych rozwiązań oraz zweryfikowania prawdziwości postawionych hipotez. Prezentacja wielu zagadnień, w tym ekstrakcji i opisu cech, a w szczególności sieci neuronowych i ich uczenia odbywa się z użyciem dość rozbudowanego aparatu matematycznego. Warta podkreślenia jest konsekwentna i spójna notacja, a także bogaty materiał ilustracyjny. Wartość pracy znacząco podnoszą precyzyjne zapisy algorytmów wraz ze stosownymi wyjaśnieniami w tekście pracy. Wyniki badań eksperymentalnych zaprezentowano w piętnastu tabelach oraz na kilkudziesięciu wykresach. Wyniki zaprezentowane na kilkudziesięciu obrazach oraz poglądowych ilustracjach bardzo dobrze uzupełniają wspomniane zestawienia tabelaryczne uzyskanych wyników. Na uwagę zasługują liczne analizy i wnioski podsumowujące każdy ze zrealizowanych eksperymentów. W analizach porównawczych porównywano i analizowano m.in. dokładności klasyfikacji, krzywe uczenia sieci neuronowych, czasy odpowiedzi w zależności od parametrów oraz wizualne reprezentacje cech. W rozdziale poświęconemu prezentacji wyników na obrazach pozyskiwanych przez robot mobilny nie ograniczono się jedynie do uzyskanych rezultatów, ale także omówiono kinematykę robota, identyfikację parametrów robota oraz architekturę systemu wizyjnego. W dodatkach uszczegółowiono równoległą implementację algorytmu CLBP oraz sieci RBM, przytoczono skrypty implementujące algorytm CLBP oraz warstwy RBM, podano schemat trenowania sieci neuronowych, a także szczegóły związane z wyznaczaniem orientacji robota w oparciu o obrazy pozyskiwane z pokładowego systemu wizyjnego. Od strony formalnej rozprawa zredagowana jest poprawnie, z właściwą systematyką podjętych problemów naukowych. Następstwo rozdziałów jest właściwe, tytuły rozdziałów i podrozdziałów dają syntetyczny pogląd na zawartą w nich treść. Prezentacja podjętej problematyki badawczej jest wyczerpująca. Autor precyzyjnie i poprawnie formuje wnioski wynikające ze zrealizowanych prac eksperymentalnych. Wnioski podsumowujące rozprawę mają oparcie w proponowanych rozwiązaniach oraz uzyskanych wynikach empirycznych. Nowoczesny i bogaty warsztat badawczy oraz kreatywność Autora zaowocowały wartościowym opracowaniem naukowym. Zagadnienia podjęte w recenzowanej rozprawie są ważne nie tylko w aspekcie naukowym, ale także w aspekcie praktycznym.

3. We współczesnych systemach wizyjnych robotów do klasyfikacji/rozpoznawania obrazów głównie wykorzystywane są sieci neuronowe głębokie, zwykle trenowane end-to-end. We wspomnianym nurcie mieszczą się także systemy wizyjne oparte o sieci głębokie, które trenowano na dużych repozytoriach obrazów, a następnie dotrenowywano na mniejszych zbiorach dla własnych zadań. Stosunkowo nieliczne są rozwiązania, w których sieci trenowane są bez nadzoru lub nawet wręcz z częściowym nadzorem. O ile przed pojawieniem się pierwszej sieci głębokiej AlexNet zaproponowano rozwiązania ze wstępną ekstrakcją cech dla sieci konwolucyjnej, to wspomniane rozwiązania były motywowane chęcią zwiększenia liczby warstw klasyfikatora, tzn. głębokości klasyfikatora, a w mniejszym stopniu potrzebą zwiększenia zdolności generalizacyjnych sieci, wytrenowania jej na mniejszym rozmiarowo zbiorze trenującym, czy też skrócenia czasu potrzebnego na wyznaczenie odpowiedzi sieci przy

zachowaniu skuteczności rozpoznawania. Niniejsza praca wychodzi poza utarte ścieżki z uwagi na wprowadzenie możliwości efektywnego wykorzystania niepoetykietowanych danych do wytrenowania niewprowadzającej dużych narzutów sieci neuronowej RBM w oparciu o wydzielone w pierwszej kolejności cechy binarne, która to odpowiedzialna jest za wydobycie najważniejszych informacji i zależności pomiędzy nimi, a przez to stworzenie możliwości użycia prostszych i mniej wymagających obliczeniowo klasyfikatorów neuronowych, a co za tym idzie konieczności etykietowania większych rozmiarowo danych trenujących. Obecnie znaczącą część uwagi poświęca się udoskonaleniu technik uczenia sieci neuronowych głębokich w kierunku lepszego wykorzystania danych niepoetykietowanych. Podejście zaprezentowane w niniejszej pracy jest dość oryginalne z uwagi na wstępne wyekstrahowanie informacji reprezentowanej przez cechy binarne przez sieć neuronową wytrenowaną bez nadzoru, a następnie klasyfikacji obrazu przez sieć neuronową trenowaną na mniejszym rozmiarowo zbiorze danych treningowych. Wyróżnikiem konkurencyjnych rozwiązań end-to-end jest to, że modele, przykładowo sieci konwolucyjne optymalizowane są jako całość, tzn. zarówno wagi ekstraktora cech jak i klasyfikatora optymalizowane są w jednym zadaniu.

Na wartościowy i oryginalny element pracy składa się nowa architektura systemu wizyjnego w której sieć neuronowa RBM wydobywa najważniejsze informacje i zależności pomiędzy wydzielonymi w pierwszej kolejności cechami binarnymi dla klasyfikatora neuronowego wytrenowanego na poetykietowanych danych. Na oryginalny dorobek Doktoranta składa się także binarny ekstraktor cech CLBP, który w odróżnieniu od pierwowzoru LBP rozszerza informację zakodowaną w wektorach cech o zależności między barwami z lokalnych sąsiedztw pikseli. Proponowane rozwiązania są oryginalne. Warto podkreślić jest to, że z uwagi na ograniczenia dostępnych implementacji sieci RBM Autor zaimplementował wspomniane sieci, w tym także skrypty do trenowania sieci RBM, co z uwagi szereg trudności praktycznych związanych uczeniem tego typu sieci nie jest zadaniem całkowicie trywialnym. Proponowany klasyfikator (moduł rozpoznający) oraz sieci neuronowe zaimplementowano, a następnie przebadano eksperymentalnie w wielu konfiguracjach w eksperymentach z jasno określonymi celami. Zaimplementowano także system wizyjny ze wsparciem GPU na układzie wbudowanym, który przebadano eksperymentalnie w zadaniach z robotem mobilnym, m.in. w zadaniu estymacji rotacji robota w oparciu o moduł CLBP-RBM, wyznaczania dokładności obrotu kamery w oparciu o histogramy w porównaniu do dokładności na podstawie pomiarów sensorycznych z enkoderów robota, klasyfikacji obiektu w oparciu o HS-CNN, pozycjonowania robota w oparciu o sprzężenie wizyjne, generowania mapy położenia obiektów w sąsiedztwie robota oraz porównywania zbiorów obrazów przez system wizyjny. Treści zawarte w pracy wskazują, że Autor ma bardzo dobre wyczucie przedmiotowego problemu naukowego.

4. Poniższe uwagi odnoszą się do bardziej szczegółowych kwestii związanych z realizacją celu rozprawy i postawionej tezy, nie są zarzutami i mają na ogół polemiczny, a nie krytyczny charakter, nie ujmujący jednak wartości pracy. Słabych stron rozprawy jest niewiele i są one mniej istotne. Pracy z pewnością nie zaszkodziłoby szersze odniesienie się do technik kompresji sieci neuronowej (w tym do quantization, knowledge distillation), rozwiązań typu student-teacher, nowoczesnych podejść do konstrukcji części sieci CNN odpowiedzialnej za klasyfikację, a wręcz nawet konstrukcji efektywnych sieci CNN z użyciem separowanych konwolucji. Z pewnością wartość pracy i proponowanych rozwiązań wzrosłaby po przeanalizowaniu pracy pt. Resource Efficient Arithmetic Effects on RBM Neural Network Solution Quality Using MNIST, IEEE, 2011. Dysertacji nie zaszkodziłoby także

przeanalizowanie pracy pt. Enhanced gradient for training restricted Boltzmann machines, 2013, a mając na względzie systemy wizyjne robotów przytoczenie pracy przeglądowej pt. A review on generative Boltzmann networks applied to dynamic systems, 2020. Od strony poznawczej, w tym także w kontekście ogólności prezentowanych rozwiązań warto byłoby odnieść się do możliwości zastosowania proponowanego podejścia w systemie stereowizyjnym robota, lub w ostateczności w systemach wizyjnych operujących na sekwencjach obrazów. Interesujące byłoby szersze porównanie LBP i CLBP w warunkach zmiennego oświetlenia. Przy starannym przejrzaniu rozprawy dostrzegłem szereg błędów o charakterze edycyjnym – niemniej liczba potknięć nie odbiega od normy (streszczenie pracy w języku polskim wymagałoby przeredagowania i staranniejszej korekty językowej).

Powyższe uwagi mają przede wszystkim charakter dyskusyjny i nie umniejszają w najmniejszym stopniu wartości pracy, która stanowi oryginalny wkład w dziedzinę robotyki i inteligencji maszynowej.

5. Przeważająca mi do recenzji rozprawa doktorska mgr. inż. Szymona Sobczaka proponuje skuteczne i efektywne rozwiązania dla systemów wizyjnych robotów opartych o sztuczne sieci neuronowe. Rozprawa potwierdza znaczącą wiedzę Doktoranta w obszarze robotyki oraz szeroko rozumianej inteligencji maszynowej, umiejętność samodzielnego prowadzenia badań naukowych, a także opanowaniem warsztatu badawczego, zarówno od strony teoretycznej, jak i eksperymentalnej. Pan mgr. inż. Szymon Sobczak wykazał się znajomością matematycznych podstaw projektowania i trenowania złożonych sieci neuronowych, jak również ich efektywnej implementacji dla systemów wbudowanych ze wsparciem GPU. Rozprawa jest ważnym przyczynkiem w zakresie wiedzy dotyczącej projektowania i realizacji systemów rozpoznających obrazy z wymaganą skutecznością, z użyciem modeli trenowanych w oparciu o małe repozytoria danych trenujących, a przy tym niewymagających znaczących zasobów obliczeniowych. W podsumowaniu stwierdzam, że recenzowana rozprawa mgr. inż. Szymona Sobczaka spełnia wszystkie wymagania ustawowe zwyczajowe stawiane rozprawom doktorskim. Stwierdzam niniejszym, że przedstawiona mi do recenzji rozprawa doktorska spełnia kryteria określone w art. 13 ust. 1 Ustawy. Wnoszę o dopuszczenie Pana mgr. inż. Szymona Sobczaka do kolejnych etapów postępowania o nadanie stopnia doktora nauk inżynieryjno-technicznych w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne. Ponadto, wysoki poziom merytoryczny pracy, potencjał praktyczny zaprojektowanych i zrealizowanych rozwiązań, a w szczególności nowatorskie rozwiązania zaproponowane w pracy są przesłanką do wnioskowania do Rady Dyscypliny Automatyki, Elektroniki, Elektrotechniki i Technologii Kosmicznych Politechniki Poznańskiej o wyróżnienie rozprawy doktorskiej Pana mgr. inż. Szymona Sobczaka.

Boydun Kwokly