Szczecin, dn. 14.09.2022

Prof. dr hab. inż. Adam Krzyżak
Wydział Elektryczny
Zachodniopomorki Uniwersytet Technologiczny
ul. Sikorskiego 37
70-313 Szczecin

R E C E N Z J A

rozprawy doktorskiej mgra inż. Kamila Szyca

„Safety and Trustworthiness of Deep Learning in Computer Vision – With Application of Out-of-Distribution Detection Techniques"

1. Zakres tematyczny i realizacja rozprawy

Niniejsza recenzja została przygotowana na podstawie pisma Prof. dra hab. inż. Michała Woźniaka, Przewodniczącego Rady Naukowej Informatyki Technicznej i Telekomunikacji, Politechniki Wrocławskiej, z dnia 17.06.2022 roku.

Recenzowana praca składa się z 3 rozdziałów i podsumowania. Całość została przedstawiona na 153 stronach. Praca poświęcona jest analizie wykrywania danych nieznanych. Praca dotyczy bardzo aktualnych i istotnych wątków badawczych realizowanych przez czołowe zespoły naukowe w Polsce i zagranicą.

Rozdział pierwszy składa się z dwóch podrozdziałów: Wstępu zawierającego postawienie problemu, cel, tezę i zakres pracy i główne osiągniecia oraz Przeglądu zawierającego omówienie podstawowych architektur i metod uczenia sieci głębokich ze szczególnym uwzględnieniem sieci konwolucyjnych jak również omówienie aktualnego stanu badań i głównych wyzwań w dziedzinie sieci głękokich, a zwłaszcza zagadnień detekcji danych nieznanych w wizji komputerowej, którym poświęcona jest niniejsza rozprawa.

Zasadnicze wyniki rozprawy zaprezentowane są w rozdziale drugim. We wstępie do tego rozdziału uzmysłowiono czytelnikowi jak trudnym zagadnieniem jest problem wykrywania danych nieznanych. W dalszej części przeanalizowano wybrane metody wykrywania danych nieznanych i ich ograniczenia. Zasadnicze metody przebadane przez Autora obejmują Extreme Value Machines (EVM), metody parametryczne oparte na

wielowymiarowym rozkładzie normalnym (MVN) wykorzystujące odległość Mahalanobisa oraz metody nieparametryczne Density-based Local Outliers (LOF). Sprawdzono wpływ metod ekstrakcji cech, wymiarów wektorów cech, rozdzielczości obrazów oraz liczby znanych klas na skuteczność metody Mahalanobisa i LOF. Zbadano też wpływ technik rozszerzania danych (ang. data augmentation) na skuteczność detekcji danych nieznanych oraz na wykrywanie wrogich ataków (ang. adversarial attacks). Autor też krytycznie się ustosunkował do problemów niestabilności metod wykrywania danych nieznanych.

W rozdziale czwartym podsumowano wyniki pracy, wyciągnięto wnioski oraz zarysowano otwarte problemy do przyszłych badań.

Celem rozprawy doktorskiej jest analiza problemu wykrywania danych nieznanych (ang. Out-of-Distribution Detection) lub detekcja OoD. Tematyka rozprawy jest silnie związana z głębokimi sieciami neuronowymi typu splotowego (ang. Convolutional Neural Networks-CNN). Skuteczna detekcja takich danych zwiększa bezpieczeństwo i wiarygodność modeli sztucznej inteligencji. W pracy skupiono się zbadaniu metod wykrywania danych nieznanch przy pomocy klasyfikacji w zborach otwartych. Zaproponowane metody zostały przeanalizowane i porównane w badaniach symulacyjnych.

W ramach realizacji celu rozprawy Autor uzyskał szereg oryginalnych i wartościowych rezultatów naukowych wymienionych poniżej:

1. Przeanalizował wpływ funkcji oceny jakości, baz danych obrazowych, architektur sieci CNN oraz metod ekstrakcji cech na wykrywanie danych nieznanych.

2. Szczegółowo przebadał w badaniach symulacyjnych zalety i wady algorytmów EVM, MVN, Simple Unified Framework (SUF) i algorytmu LOF oraz porównał te algorytmy. Pokazał, iż algorytm LOF bazujący na estymacji gęstości rozkładu wykazuje lepszą efektywność w wykrywaniu danych nieznanych niż pozostałe algorytmy.

3. Wykazał, iż w problemach o dużej skali (zarówno pod względem rozdzielczości obrazów jak też liczby klas) metoda LOF charakteryzuje się na ogół większą efektywnością niż pozostałe metody.

4. Zademonstrował, że wybór metody ekstrakcji cech ma znaczny wpływ na efektywność OoD. W szczególności pokazał, że parametryzacja popularnej metody GAP (ang. Global Average Pooling) może prowadzić do lepszych rezultatów. Pokazał również, że metody OoD są wrażliwe na redukcję wymiarowości wektora cech, a najmniej wrażliwą metodą okazuje się być metoda LOF.

5. Wykazał, iż detekcja danych OoD może być użyteczna do wykrywania wrogich ataków (ang. adversarial attacks).

6. Badania eksperymentalne zademonstrowały wrażliwość metod OoD na zmiany w modelach. Wykazały również, że nie istnieje najlepsza uniwersalna technika OoD.

7. Autor również zademonstrował wpływ technik rozszerzania danych (ang. data augmentation) na skuteczność detekcji danych OoD.

8. Podsumowując osiągniecia Autora należy docenić Jego ogromny wysiłek włożony w realizację wielu algorytmów i technik głębokiego uczenia i przetwarzania obrazów.

9. Na wyróżnienie zasługuje również Rozdział 2, w którym Autor krytycznie zaprezentował wyczerpujący przegląd technik i algorytmów uczenia głębokiego ze szczególnym uwzględnieniem sieci konwolucyjnych oraz omówił aktualne i najważniejsze pozycje literaturowe. O wnikliwości Autora świadczy fakt, że omówiona literatura obejmuje prawie 300 pozycji.

Najważniejsze wyniki rozprawy zostały opublikowane w materiałach konferencji międzynarodowych: 2018 *IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, *International Conference on Dependability and Complex Systems*, Springer 1919, 2021, 2022.

Rozprawa jest zredagowana bardzo starannie. Autor sprawnie posługuje się zaawansowanymi narzędziami oprogramowania i wszystkie wywody są klarowne i poprawne. Poniższe uwagi mają charakter dyskusyjny i nie wpływają na wysoką ocenę pracy doktorskiej:

1. Najlepsze wyniki w tabelach w Rozdziale 3 powinny być wyróżnione tłustym drukiem. Uwaga ta dotyczy zdecydowanej większości tabeli w tym rozdziale, np. Tab. 3.5, 3.9-3.16. Przykładem poprawnej tabeli jest Tab. 3.8.

2. W Tabeli 3.7 porównano metodę z pracy [132] z metodą LOF zbadaną przez Autora i stwierdzono, że metoda LOF przewyższa tę z pracy [132]. Porównując wartość AUROC 99.14 z [132] z wartościami AUC w Tab. 3.7 (rząd 3-6, kolumna AUC) zauważamy nieznaczną różnicę. Czy ta nieznaczna różnica usprawiedliwia wniosek Autora?

3. Śledząc dokładność DTACC przebadanej metody w Tab. 3.14 w zależności od n można zaobserwować wartości maksymalne dla podejścia LOF dla n=64. Dlaczego Autor w ostatnim zdaniu podsumowania Sekcji 3.4.2 na str. 98 sugeruje wybór n=64? W tym miejscu nasuwa się ogólne pytanie czy można wyciągać wnioski o optymalnej wartości parametrów na podstawie eksperymentu na jednej bazie danych?

4. Jakiego modelu dotyczy pierwsza część Tabeli 3.17 na str. 103? Z opisu wynika, że chodzi o model ResNet-101 i bazę obrazów CIFAR-10, ale w samej tabeli ta informacja została pominięta.

5. Obrazki na Rys. 3.9 są zbyt małe, aby rzetelnie ocenić efekty rozszerzeń.

6. Przedstawione w rozprawie wnioski są oparte na wynikach eksperymentów komputerowych. Czy istnieje lub jest rozwijana teoria potwierdzająca wyniki zaprezentowane w niniejszej rozprawie? Komentarz w tej kwestii byłby cennym uzupełnieniem Rozdziału 2.

7. Lista akronimów na str. XI jest niekompletna. Np. brakuje na niej UF, LOF, SCDFA i wielu innych.

8. Jakkolwiek rozprawa jest napisana bardzo starannie zawiera ona sporo uchybień językowych w wersji angielskiej i kilka drobnych uchybień natury redakcyjnej. Na przykład

   a) str. 10, l. 16: The era of deep learning has begun [129].

   b) str. 10: równanie (2.1) powinno być umieszczone na końcu Sekcji 2.1.1.1.

   c) str. 11, l. 13: "consecrations" ?

   d) str. 13, l. 22: .. how to generate high-resolution images...

e) str. 13, l. -3: We can distinguish two directions:

f) str. 18, l. -8: ... this technique can even be harmful.

g) str. 34, środek strony: ... developed a new net called FixResNet...

h) str. 77, 2-ga linia za Rys. 3.2: ... are the mean and standard deviation...

i) str. 101, ostatni paragraph: For Downscale no augmentation achieved...

W podsumowaniu stwierdzam, co następuje:

a) W rozprawie doktorskiej Autor rozwiązał ważny problem dotyczący bezpieczeństwa i wiarygodności modeli sztucznej inteligencji poprzez przebadanie wybranych technik wykrywania danych nieznanych. Między innymi zademonstrował w badaniach symulacyjnych, że metoda nieparametryczna LOF sprawdza się lepiej niż inne popularne techniki.
b) Rozprawa zawiera szereg oryginalnych i wartościowych rezultatów naukowych wymienionych powyżej i opublikowanych przez autora w materiałach międzynarodowych konferencji.
c) Rozprawa jest zredagowana bardzo starannie, a poszczególne wątki przedstawione są w sposób zrozumiały i kompetentny.

W konkluzji stwierdzam, że rozprawa doktorska „Safety and Trustworthiness of Deep Learning in Computer Vision – With Application of Out-of-Distribution Detection Techniques", której autorem jest mgr. inż. Kamil Szyc, spełnia wymagania stosownej ustawy o stopniach naukowych i tytule naukowym. Wnoszę o jej przyjęcie i dopuszczenie do publicznej obrony.

Adam Krzyżak