

dr hab. inż. Grzegorz Bocewicz, prof. PK
Wydział Elektroniki i Informatyki
Politechnika Koszalińska
Śniadeckich 2
75-453 Koszalin

Koszalin, 26.02.2021 r.

R e c e n z j a

rozprawy doktorskiej mgr inż. Daniela Halikowskiego

**pt. „Przetwarzanie sekwencji wideo przy zastosowaniu podejścia deep learning
do automatycznego generowania instrukcji stanowiskowych”**

Dziedzina: dziedzina nauk inżynieryjno-technicznych

Dyscyplina: informatyka techniczna i telekomunikacja

1. Obszar tematyczny rozprawy

Tematyka rozprawy dotyczy szerokokorozumianej analizy i klasyfikacji obiektów w nagraniach wideo rejestrujących aktywność pracowników przedsiębiorstw produkcyjnych. W szczególności Doktorant koncentruje się na zagadnieniach automatycznego generowania instrukcji stanowiskowych oraz automatycznej oceny poprawności realizacji wybranej usługi serwisowej (w rozważanym przypadku obsługi kotła na paliwo stałe). Możliwość automatycznej oceny aktywności pracowniczej wykorzystującej nagrania wideo wiąże się z koniecznością rozwiązania szeregu problemów cząstkowych obejmujących m.in: analizę i przetwarzanie obrazu, rozpoznawanie (identyfikację) obiektów, identyfikację kolejnych etapów usługi serwisowej (określenie relacji między zidentyfikowanymi obiektami), a także wykrywanie nieprawidłowości w realizacji tej usługi. Tematyka rozpoznawania obrazów jest znana od wielu lat, ale jak się okazuje wciąż bardzo aktualna. Powszechnie stosowane rozwiązania (w szczególności z obszaru metod uczenia głębokiego - deep learning) z powodzeniem wykorzystywane są w procesach rozpoznawania obiektów na obrazach (sterowanie robotami przemysłowymi, wykrywanie niebezpiecznych sytuacji, monitoring zachowania pracowników, itp.), jednak ich wykorzystywanie w zakresie kontroli i oceny złożonych procesów produkcyjnych (w tym przypadku przebiegu realizowanej usługi) wciąż stanowi duże wyzwanie.

W ogólności problemy rozpoznawania obrazów wymagają znacznych nakładów obliczeniowych. Oznacza to, że czasowo efektywne metody rozpoznawania zwykle charakteryzują się niewielką dokładnością i na odwrót wzrost dokładności okupiony jest dużym czasem obliczeń.



Należy zaznaczyć, że stopień „trudności” tego typu problemów rośnie, gdy w procesie rozpoznawania uwzględnia się dodatkowo relacje zachodzące między obiektami jak np. **ręka sięga po klucz**. Możliwość rozpoznania zarówno obiektów, jak i relacji zachodzących w danym momencie pomiędzy nimi jest warunkiem poprawnej identyfikacji poszczególnych etapów realizowanej usługi. Ze względu na konieczność kontroli usługi w trakcie jej realizacji proces tego typu powinien odbywać się przy określonym rygorze czasowym (przyjęty czas rozpoznania nie przekracza 1/5 s). Przy tym założeniu, możliwość budowy systemu pozwalającego w trybie on-line na identyfikację (i ocenę poprawności) określonych etapów zadanej usługi stanowi motywację prowadzonych przez Doktoranta badań.

Biorąc pod uwagę, że prowadzone prace ograniczają się do oceny aktywności pracowniczek w zakresie wybranych usług, opiniowana rozprawa koncentruje się na problemie, którego istotę można zrekonstruować jak następuje: Dany jest zbiór instrukcji stanowiskowych określających postępowanie pracownika w trakcie realizacji oferowanych usług serwisowych. Dana jest sekwencja wideo, na której zarejestrowano aktywność pracownika realizującego określoną usługę (sekwencja ta stanowi podstawę do oceny poprawności realizowanej usługi). Poszukiwana jest odpowiedź na pytanie: Czy czynności wykonane przez pracownika są zgodne z wymaganiami sformułowanymi w instrukcji stanowiskowej (np. czy kolejność wykonanych czynności pokrywa się z zaplanowanymi)? Dla tak zdefiniowanego problemu prowadzone są poszukiwania modeli sieci neuronowych (typu: YOLOv3; CNN, CNN + SVM; R-CNN; itp.) zapewniających efektywne (w trybie on-line) rozpoznawanie w zadanej sekwencji obrazów wideo, obiektów oraz relacji zachodzących pomiędzy nimi (identyfikacja czynności składających się na usługę serwisową).

Uważam, że opiniowana rozprawa podejmuje ważny i aktualny problem. Elementy nowości przejawiają się w oryginalnym sformułowaniu problemu rozpoznawania obiektów i łączących je relacji, propozycji modeli sieci neuronowych i wspierających je algorytmów oceny poprawności realizowanych usług serwisowych. Podjęcie przedstawionej problematyki jest uzasadnione zarówno ze względów poznawczych, jak i możliwości wielu praktycznych zastosowań związanych m.in. z systemami wsparcia działów utrzymania ruchu przedsiębiorstw produkcyjnych.

2. Kompozycja i treść rozprawy

Opiniowana rozprawa liczy 118 stron i składa się z 7 rozdziałów, wykazu cytowanej bibliografii, spisów stosowanych skrótów, rysunków i tabel. Załączona lista źródeł bibliograficz-

nych obejmuje ważniejsze pozycje literaturowe z zakresu przedmiotu pracy. W bibliografii występują dwie współautorskie publikacje Doktoranta. Jedna z nich została opublikowana w czasopiśmie Sensors (IF = 3.275) indeksowanym w bazie Scopus oraz Journal Citation Reports.

W rozprawie wyodrębnić można dwie zasadnicze części. W części pierwszej (obejmującej rozdziały 1-3) sformułowano problem badawczy, tezy, cel główny rozprawy oraz cele szczegółowe. Przedstawiono również podstawowe pojęcia oraz stan wiedzy z zakresu rozpoznawania, analizy obrazów i ich klasyfikacji. W szczególności przedstawiono sformułowanie problemu automatycznego generowania zbioru instrukcji stanowiskowych i weryfikacji aktywności pracownika na stanowisku roboczym (ocena poprawności realizacji usługi serwisowej). Przyjęta formuła prezentacji jest w większości przypadków klarowna, choć miejscami tekst jest zbyt ogólny. Przykładem może tu być brak formalnej definicji instrukcji stanowiskowej determinującej przebieg realizowanej przez pracownika usługi. Strukturę usługi czytelnik poznaje pośrednio dopiero w rozdziale 4. Tezy rozprawy zostały sformułowane poprawnie, choć pewien niedosyt odczuwalny jest w zakresie ich uzasadnienia. Poza krótką dyskusją zawartą we „Wprowadzeniu” brakuje szerszego uzasadnienia, z jakiego powodu Doktorant zdecydował się na takie, a nie inne rozwiązanie. Szczególnie ważne w tym kontekście wydaje się przedstawienie pogłębionej dyskusji tego na ile przyjęte technologie spełniają wymagania (wynikające z warunków w jakich wykonywana jest usługa) końcowego użytkownika (dział utrzymania ruchu).

Warto zwrócić uwagę na bardzo obszerną charakterystykę dostępnych rozwiązań AI (rozdziały 2 i 3), w której szczegółowo przedstawiono współczesne modele (sieci neuronowych typu: CNN; R-CNN; YOLO) oraz metody analizy i rozpoznawania obrazów. Pomimo wysokich walorów użytkowych, do niedostatków tej części rozprawy należy zaliczyć brak dyskusji i krytycznej oceny Doktoranta w zakresie możliwości wykorzystania omawianych rozwiązań w procesie oceny aktywności ludzkiej (identyfikacji obiektów i zachodzących między nimi relacji). Lekturę tej części pracy utrudnia również brak jednolitego systemu oznaczeń dla wszystkich omawianych metod (np. do określania wag neuronów wykorzystano aż cztery warianty oznaczeń: w_i , W_i , $W_{i,j}$, w_{hk}), co niejednokrotnie prowadzi do niejednoznaczności.

W części drugiej (rozdziały 4-7) zawierającej główne (poznawcze) wyniki rozprawy przedstawiono autorską metodę generowania instrukcji stanowiskowych. Metoda ta bazuje na sześciu algorytmach umożliwiających w szczególności: wyodrębnienie klatek referencyjnych kolejnych etapów procedury serwisowej, identyfikację obiektów na klatkach referencyjnych, identyfikację wykonywanej czynności, wykrywanie nieprawidłowości realizowanej procedury

serwisowej. W przedstawionym kontekście najważniejszym elementem, tej części rozprawy, jest propozycja dwuetapowej analizy obrazu obejmująca:

- Etap 1: identyfikację obiektów na testowanym obrazie.
- Etap 2: ocenę podobieństwa testowanego obrazu z obrazem referencyjnym (klatką referencyjną odpowiadającą określonej czynności procedury serwisowej).

W odróżnieniu od rozwiązań spotykanych w literaturze tego typu podejście umożliwia identyfikowanie na obrazie zarówno obiektów, jak i relacji występujących między nimi.

Efektywność metody bazującej na zaproponowanym podejściu została zweryfikowana w serii eksperymentów badawczych (rozdział 5). Badania pokazują, że w rzeczywistych warunkach produkcyjnych (usługa serwisowa kotła na paliwo stałe) dokładność identyfikacji czynności pracownika nie jest mniejsza niż 70%. Uzyskane wyniki stanowiły przesłankę do budowy systemu wspomaganie szkoleń opisanego w rozdziale 6.

Wśród niedostatków tej części rozprawy w pierwszej kolejności należy zaliczyć brak analizy złożoności obliczeniowej proponowanych algorytmów oraz brak oceny skali rozmiarów instancji, dla których możliwe jest efektywne rozpoznawanie czynności pracownika tzn.:

- określenia maksymalnej liczby obiektów, jaką metoda pozwala efektywnie (z zadaną dokładnością) identyfikować na obrazie,
- oceny rodzaju i „złożoności” wykonywanych czynności w ramach wykonywanej usługi.

Z tego względu trudno jest ocenić możliwość wykorzystania proponowanego podejścia w bardziej złożonych (niż te analizowane w rozprawie) procedurach.

Szkoda również, że Doktorant nie przeprowadził eksperymentów porównawczych w innym środowisku np. przy zmiennym oświetleniu, innym pomieszczeniu, innym zestawem wykorzystywanych narzędzi, itp. Tego typu badania pozwoliłyby określić obszar zastosowań opracowanej metody (niekoniecznie ograniczający się do przedsiębiorstw produkcyjnych). Analogiczna uwaga dotyczy procesu trenowania zaprojektowanego systemu, dla którego Doktorant założył stały zbiór rozpoznawanych obiektów. W pracy zabrakło badań pozwalających ocenić skuteczność metody dla innej klasy obiektów (np. dla większej ich liczby lub zmiany ich kształtu, koloru, itp.).

W pracy zabrakło również matematycznej postaci opracowanego modelu umożliwiającego generowanie instrukcji stanowiskowych. Zamiast tego podany jest mało czytelny rysunek ze str. 65. Część elementów składających się na model wyjaśniono dopiero przy opisie algorytmów (algorytmy 1-6), a część nie zostało zdefiniowanych w ogóle (np. FA, OFA, SA, cecha klatki, str. 66).

Do mankamentów rozprawy należy również zaliczyć sposób opisu przeprowadzonych eksperymentów. W przyjętej strukturze, na którą składają się: założenia, cele szczegółowe, przebieg, rezultaty, wnioski, brakuje celu głównego. O motywacji doktoranta i celu prowadzonych badań można przekonać się dopiero po lekturze przebiegu poszczególnych eksperymentów.

Pomimo wskazanych niedostatków należy podkreślić, że przyjęty sposób narracji tej części rozprawy jest na ogół poprawny i świadczy o dobrym przygotowaniu oraz wysokich kompetencjach Doktoranta.

3. Oryginalne osiągnięcia

Do najważniejszych osiągnięć rozprawy, wyróżniających je spośród dostępnych w literaturze przedmiotu, można zaliczyć:

1. Opracowanie autorskiego podejścia dwuetapowej analizy obrazu umożliwiającej identyfikowanie na obrazie obiektów zadanej klasy i relacji występujących między nimi.
2. Opracowanie metody automatycznego generowania instrukcji stanowiskowych, na którą składa się 6 autorskich algorytmów umożliwiających identyfikację/weryfikację realizowanych czynności.
3. Zaplanowanie oraz przeprowadzenie szeregu eksperymentów weryfikujących efektywność opracowanej metody.
4. Opracowanie systemu wspomagania procesów szkoleń przeznaczonego dla rzeczywistego przedsiębiorstwa produkcyjnego.

Uwzględniając wymienione osiągnięcia naukowe uważam, że Doktorant zrealizował cel rozprawy. Uzyskane rezultaty potwierdzają kompetencje Doktoranta z zakresu metod sztucznej inteligencji, technik programowania i planowania/realizacji eksperymentów komputerowych. Dowodzą również, że Doktorant potrafi podejmować i samodzielnie realizować zaplanowane cele badawcze.

4. Uwagi i komentarze dotyczące rozprawy

Lektura rozprawy skłania do kilku uwag, tak ogólniejszej, jak i bardziej szczegółowej natury:

1. Zaproponowane podejście dwustopniowej analizy obrazu umożliwia wykrywanie relacji zachodzących między obiektami, które odpowiadają określonej czynności analizowanej procedury serwisowej. Interesującym w tym kontekście wydaje się być pytanie, w jakim



stopniu proponowana metoda może zostać zaadoptowana do projektowania/rozwijania ontologii reprezentującej wiedzę (know-how przedsiębiorstwa) w zakresie realizowanych usług?

2. Prezentowane wyniki eksperymentów (rozdziały 5 i 6) pozwoliły Doktorantowi ocenić efektywność zaproponowanej metody dla konkretnej usługi serwisowej. Rodzące się w tym kontekście pytanie dotyczy istnienia stosownych warunków (np. warunków określających parametry ilościowe charakteryzujące oferowane usługi serwisowe), spełnienie których gwarantuje zadaną dokładność weryfikacji.
3. Interesujące jest również pytanie o możliwość rekonfiguracji wykorzystanych sieci neuronowych (YOLOv3; CNN, CNN + SVM; R-CNN) w zakresie: liczby warstw, funkcji aktywacji itp., w celu poprawy uzyskanej dokładności weryfikacji.

Uwagi szczegółowe:

1. Brak spisu wykorzystywanych symboli i oznaczeń prowadzi do kolizji oznaczeń i błędów w prezentowanych wyrażeniach, np.: brak opisu oznaczeń X_i , W_i w wyrażeniu (3), błędy w opisie oznaczeń wyrażenia (7), błędna definicja funkcji $\tanh(x)$ w tabeli 5; równania układu (11) są takie same; wielkości liter symboli w wyrażeniach (18), (19) są różne; brak definicji hiperparametrów str. 59, itp.
2. str. 63 – czy można zdefiniować „*podobieństwo środowiska pracy*”?
3. str. 63 – z jakiego powodu wykluczono sytuację, w której obiekty są identyfikowane poprawnie, a testowane klatki nie są wystarczająco podobne ?
4. str. 76. – czy można doprecyzować zwrot „*pobierać obraz w czasie rzeczywistym*”?
5. str. 77. – czy można doprecyzować zwrot „*wyznaczonych ręcznie klatek referencyjnych*”?; jakie kryteria wykorzystywane są w tym procesie?
6. str. 83. - punkt „*Dataset do eksperymentu ...*”, nie jest celem.
7. str. 84 – tabele 19, 20, itd. - czy można zdefiniować parametr „*skuteczność*” i czym różni się od „*dokładności*” ze strony 87?
8. Większość z rysunków nie posiada legend i opisu stosowanych oznaczeń. Część z nich ze względu na swój rozmiar jest nieczytelna (np. rys. 18, 19) - przyjęcie takiego sposobu prezentacji utrudnia lekturę pracy.
9. W rozprawie występują błędy językowe np. „*pięciu*” (str. 4); „*oby typów*” (str. 16); „*traktowane jako problemem regresji*”; „*500 obrazków*”, „*datasecie*” (str. 90) itp.

Pomijając powyżej wymienione uwagi należy zaznaczyć, że rozprawa jest zredagowana poprawnie, posiada właściwą strukturę i proporcje. Błędy stylistyczne występują sporadycznie.

5. Konkluzja

Reasumując stwierdzam, że w recenzowanej rozprawie doktorskiej pana mgr inż. Daniela Halikowskiego rozwiązany został oryginalny problem badawczy, polegający na opracowaniu metody automatycznego generowania instrukcji stanowiskowych wykorzystujący paradygmat głębokiego uczenia (deep learning). Doktorant wykazał się znajomością podstawowej literatury przedmiotu rozprawy, umiejętnościami budowy i komputerowej implementacji metod sztucznej inteligencji, a także umiejętnościami prowadzenia eksperymentów obliczeniowych.

Uważam, że opiniowana rozprawa spełnia warunki stawiane przez obowiązującą ustawę o stopniach i tytule naukowym w określeniu do rozpraw doktorskich (Ustawę z dnia 3 lipca 2018 r. Przepisy wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce) w dyscyplinie w dyscyplinie **informatyka techniczna i telekomunikacja** i wnioskuję o jej dopuszczenie do publicznej obrony.

