

dr hab. inż. Dariusz Pazderski, prof. PP
Instytut Automatyki i Robotyki
Wydział Automatyki, Robotyki i Elektrotechniki
Politechnika Poznańska

Poznań dnia 15 marca 2025 r.

SEKRETARIAT
Rady Dyscypliny AEEITK

Wpłynęło dnia 25.03.2025

Zarejestrowano pod nr 510-13.f/24

Podpis dm

RECENZJA ROZPRAWY DOKTORSKIEJ
przygotowana dla Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika
i Technologie Kosmiczne Akademii Górniczo-Hutniczej

na podstawie pisma Przewodniczącego Rady Dyscypliny Automatyka, Elektronika i Technologie Kosmiczne, dr. hab. inż. Ryszarda Sroki, prof. AGH, z dnia 7 stycznia 2025 r.

Tytuł rozprawy: Zastosowanie głębokiego uczenia ze wzmocnieniem w robotycznym montażu przemysłowym

Autor rozprawy: mgr inż. Grzegorz Bartyzel

Dyscyplina naukowa: Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne

Promotor: dr hab. inż. Paweł Rotter, prof. AGH

Promotor pomocniczy: dr inż. Marcin Węgrzynowski

1. Tematyka, teza naukowa i cel rozprawy

Rozprawa doktorska dotyczy obszaru robotyki, uczenia maszynowego i sterowania. Obejmuje zagadnienia modelowania procesów dynamicznych oraz projektowania architektury sterującej robotów przemysłowych, ze szczególnym uwzględnieniem algorytmów uczenia maszynowego wykorzystywanych do percepcji i wyznaczania strategii (tzw. polityki), wspieranych przez klasyczne techniki sterowania pozycyjno-siłowego.

Praca została zrealizowana jako wynik współpracy podmiotu akademickiego (AGH) z partnerem przemysłowym w ramach projektu finansowanego przez NCBR.

Rozważana tematyka jest niewątpliwie aktualna ze względu na rosnące znaczenie systemów zrobotyzowanych w elastycznych procesach produkcyjnych, które umożliwiają ich szybkie dostosowywanie do zmieniających się wymagań wytwórczych. Z perspektywy badawczej kluczowe jest opracowanie nowych architektur sterowania manipulatorami, zwiększających autonomię ruchu oraz wykazujących odporność na zmiany środowiska i różnorodność elementów, na których bezpośrednio operują. Potwierdza to rosnąca liczba prac badawczych w tej dziedzinie robotyki, obserwowana w ciągu ostatnich dziesięciu lat.

Autor rozprawy, wybierając tę tematykę, kieruje się przesłankami praktycznymi, właściwie definiując potrzebę poszukiwania nowych rozwiązań w zakresie inteligentnych metod sterowania robotami. Szczególny nacisk kładzie na zadania precyzyjnej manipulacji, w których dochodzi do

interakcji obiektu ze środowiskiem, co wymaga zastosowania strategii sterowania pozycyjno-siłowego.

Uwzględniając współczesne trendy w robotyce, wyraźnie rośnie znaczenie metod wykorzystujących uczenie maszynowe, które szczególnie dobrze sprawdzają się w rozwiązywaniu zadań charakteryzujących się niepewnością modelu i informacji. Do tej klasy problemów należy manipulacja podczas montażu elementów mechanicznych, wymagająca wpasowania danego elementu do otworu (ang. *peg-in-hole insertion*). Szczególnym przykładem takiego zadania jest montaż elementów elektronicznych w wersji przewlekanej (THT) w otworach na płytkach drukowanych (ang. PCB). Ponieważ elementy te zazwyczaj mają wiele wyprowadzeń, ich montaż można rozpatrywać jako przypadek jednoczesnego dopasowania wielu elementów do otworów (ang. *multiple peg-in-hole insertion*). Jest to szczególnie trudne w warunkach niepewności pozycjonowania i zmienności asortymentu produkcji.

Wybór tematu pracy wyraźnie odzwierciedla zainteresowanie Doktoranta obszarem uczenia maszynowego, a w szczególności paradygmatem uczenia ze wzmocnieniem, znanym jako *Reinforcement Learning* (RL). Choć koncepcja RL nie jest nowa i ma solidne podstawy w teorii sterowania – jej początki sięgają połowy XX wieku – to obecnie rozwija się niezwykle dynamicznie dzięki realizowalności uniwersalnych aproksymatorów, którymi są sieci neuronowe. Naturalnym obszarem zastosowania RL jest robotyka, w której interakcja upostaciowanego agenta – robota – ze światem rzeczywistym z natury wiąże się z dużą niepewnością modelu i informacji.

Jestem przekonany, że wybór tematyki przez Doktoranta jest trafny zarówno z punktu widzenia aspektu badawczego jak i zastosowań. Zagadnienia badawcze, które podejmuje z pewnością wpisują się w dyscyplinę naukową Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne (AEEiTK). Wynika to zarówno z wiodącej tematyki związanej z sterowaniem w robotyce, rozwojem architektur sterujących manipulatorami wykorzystującymi podejście RL, przetwarzaniem danych sensorycznych, a także użytego warsztatu badawczego, którego istotnym elementem jest opis problemu w ujęciu formalnym, implementacja metod dla rzeczywistego manipulatora, prowadzenie badań eksperymentalnych i analiza wyników tych badań. Głównym celem Autora jest implementacja nowej architektury sterowania robotem manipulacyjnym o wysokim poziomie adaptacji na zmienność obiektu i środowiska w zadaniu montażu przemysłowego. Punktem wyjścia dla Doktoranta jest stosunkowo nowe podejście RL wykorzystujące strategię poza polityką (*off policy*) w odmianie aktor-krytyk znany w literaturze jako *Soft Actor-Critic* (SAC).

W rozprawie nie sformułowano jawnie tezy badawczej, ograniczając się do scharakteryzowania głównych celów badawczych. Stwierdzono, że głównym celem jest wykazanie stosowności RL w rzeczywistym zadaniu przemysłowym na przykładzie montażu komponentów elektronicznych przewlekanych. Założono, że algorytmy głębokiego uczenia ze wzmocnieniem, połączone z uczeniem reprezentacji wielomodalnych, mogą być skutecznie wykorzystywane w elastycznych systemach produkcyjnych dla systemów robotycznych, umożliwiając ich szybkie dostosowanie do nowych zadań i manipulowanych obiektów. Zwrócono również uwagę na celowość stosowania fuzji wielomodalnej w celu ekstrakcji cech istotnych dla realizowanego zadania, co powinno poprawić zdolność generalizacji nauczonej strategii RL.

2. Zawartość merytoryczna

Rozprawa została przygotowana w języku angielskim i liczy 94 strony wraz z bibliografią. Zawiera 140 odniesień do referatów konferencyjnych, publikacji w czasopismach oraz materiałów dostępnych na stronach internetowych. Jej zasadniczą część stanowi 5 rozdziałów, wzbogaconych o załączniki, streszczenia w języku polskim i angielskim oraz wykaz rysunków i tabel.

Pierwszy rozdział stanowi wprowadzenie do głównej tematyki dysertacji. Omawia zagadnienie manipulacji pozycyjno-siłowej w procesach produkcyjnych oraz wskazuje – moim zdaniem słusznie – na potrzebę zwiększenia elastyczności tych procesów poprzez zastosowanie nowych architektur sterowania manipulatorami. W szczególności uwzględnia wykorzystanie fuzji sensorycznej, systemów wizyjnych oraz metod uczenia maszynowego, ze szczególnym naciskiem na uczenie ze wzmocnieniem. Doktorant charakteryzuje zadanie manipulacyjne typowe dla montażu komponentów elektronicznych typu THT. Następnie krótko przedstawia cele rozprawy oraz opisuje jej strukturę. Dalej obszernie omawia rozważany problem na gruncie literatury przedmiotu, odwołując się zarówno do kluczowych prac z klasycznej robotyki i sterowania jak i najnowszych publikacji, związanych z przetwarzaniem informacji wykorzystującym uczenie maszynowe. Prowadzi dyskusję m.in. w zakresie zastosowania Reinforcement Learning, metod reprezentacji wiedzy i wielomodalnych modeli generatywnych, odwołując się do koncepcji wykorzystywanych następnie do realizacji celów rozprawy. Rozdział kończy autorski wykaz głównych osiągnięć dysertacji.

W rozdziale drugim Doktorant wprowadza podstawowe pojęcia. Dominującym zagadnieniem jest tutaj uczenie ze wzmocnieniem, którego omówienie ukazuje złożoność problemu oraz szeroki wachlarz metod i narzędzi możliwych do zastosowania. Prezentacji tej towarzyszą precyzyjne odwołania do literatury. Autor analizuje znane podejścia w ramach RL, takie jak metody oparte na funkcji kosztu, gradientu strategii (polityki), wykorzystujące akcje zgodne lub niezgodne z polityką, a także rozwiązania modelowe i bezmodelowe. Następnie przedstawia metodę aktor-krytyk i przechodzi do klasy metod wykorzystujących głębokie uczenie, określanych jako Deep RL. Szczegółowo omawia algorytm z grupy aktor-krytyk typu *Soft Actor-Critic* (SAC), zaproponowany w 2019 r. przez Haarnoja i współautorów, który stanowi podstawę rozwiązań opracowanych w ramach rozprawy. Kolejna, stosunkowo krótka część rozdziału dotyczy zagadnień sterowania pozycyjno-siłowego, kluczowego elementu współczesnych architektur sterowania manipulatorami. Doktorant przypomina klasyczne podejście, oparte na aktywnym przełączaniu między trybem sterowania pozycyjnego i siłowego, oraz omawia sterowanie oparte na koncepcji impedancji/admitancji podatnego układu mechanicznego II rzędu. Ostatni, najkrótszy podrozdział poświęcony jest modelom generatywnym, w szczególności wariacyjnym autoenkoderom (VAEs), które znajdują zastosowanie w głównej metodzie proponowanej przez Autora rozprawy. Warto zauważyć, że w tym rozdziale Doktorant nie tylko charakteryzuje istniejący stan wiedzy, ale także porządkuje pewne pojęcia. Nie ogranicza się przy tym do opisu słownego, lecz stosuje formalne metody przedstawienia zagadnienia, co pozwala na jego bardziej rzetelne ujęcie.

Następne dwa rozdziały są kluczowe, gdyż dotyczą głównych wyników uzyskanych przez Doktoranta. Ich treść częściowo pokrywa się z zawartością dwóch artykułów opublikowanych przez Autora dysertacji w latach 2023-24.

W rozdziale 3 Doktorant omawia stanowisko badawcze oraz strukturę i implementację architektury sterującej, wykorzystującej uczenie ze wzmocnieniem w wersji SAC. Przedstawia zadanie manipulacji, określa przestrzeń zadania i jego geometrię. Charakteryzuje użyte sensory oraz nadrzędną i podrzędną pętlę sterowania. Równolegle wprowadza propozycję funkcji nagrody, wykorzystywaną następnie w algorytmie SAC. Istotnym wątkiem jest omówienie oprogramowania opartego na znanych narzędziach, takich jak ROS2 oraz środowiska Ape-X i Gymnasium, wspomagających głębokie uczenie ze wzmocnieniem. W celu zapewnienia właściwej synchronizacji pomiędzy poszczególnymi podsystemami Doktorant opracował dodatkowe moduły wykorzystujące protokół XML-RPC. Następnie Autor dysertacji przechodzi do przedstawienia narzędzi obliczeniowych oraz demonstrowa elementy i płytki drukowane wykorzystywane w zadaniach manipulacji (w procesie uczenia i walidacji). Opisuje również metodykę zaburzania pozycji zadanej w procesie uczenia. Omawia implementację i parametryzację sterownika admitancyjnego, który odpowiada na sygnał zadany, wyznaczany przez zewnętrzną pętlę sterowania, tj. akcje generowane

przez agenta RL. Po scharakteryzowaniu zaimplementowanej architektury RL SAC Doktorant przechodzi do opisu procedury uczenia oraz określa scenariusze testowe. Następnie ilustruje wyniki badań eksperymentalnych, prezentując efektywność procesu uczenia oraz działanie systemu dla siedmiu zestawów elementów i płytek drukowanych. Część wyników zestawia w dodatku B.1. Doktorant dokonuje również porównania metody SAC z klasycznymi technikami montażu, na podstawie którego wnioskuje o znacząco lepszych wynikach osiąganych dzięki metodzie uczenia ze wzmocnieniem, zarówno w warunkach znajomości stanu manipulatora i pomiaru sił/momentów sił, jak i w przypadku wzbogacenia informacji sensorycznej o dane wizyjne dostarczane przez dwie kamery RGB. Ponadto przedstawia interpretację odpowiedzi układu zamkniętego na podstawie przebiegów pozycji oraz sygnałów pomiarowych sił i momentów sił.

Rozdział 4 ma największe znaczenie badawcze, ponieważ omawia oryginalną propozycję Doktoranta – Multimodal Variational DeepMDP (MVDeepMDP), opracowaną na podstawie idei DeepMDP, wprowadzonej przez Galadę i innych w 2019 r. Dotyczy ona zastosowania modelu przestrzeni ukrytej, który jest trenowany poprzez minimalizację dwóch obliczalnych strat: przewidywania nagród oraz przewidywania rozkładu następnych stanów ukrytych. Doktorant proponuje nowy, wielomodalny schemat przetwarzania informacji, do integracji którego wykorzystuje uogólniony mechanizm Product-of-Experts (gPoE) i odpowiednio dostosowuje go do źródeł informacji w rozważanym zadaniu manipulacji. Omawia włączenie reprezentacji danych w warstwie ukrytej do metody uczenia ze wzmocnieniem SAC. Następnie przedstawia proponowaną architekturę wraz z opisem formalnym dotyczącym aktualizacji funkcji wartości akcji, polityki i entropii. Kolejnym krokiem jest krótkie omówienie procesu uczenia i przedstawienie testów porównawczych efektywności MVDeepMDP z sześcioma innymi metodami RL, w tym z podstawową implementacją SAC. Doktorant demonstruje odporność MVDeepMDP na nowe warunki w środowisku oraz na inny zbiór obiektów, które nie były wykorzystywane na etapie uczenia. Podkreśla łatwość douczania architektury MVDeepMDP oraz pokazuje możliwość zastosowania metody nawet przy znacznie zmienionych warunkach realizacji zadania (tzw. *cross-domain transferability*). Interesujące jest także przedstawienie wpływu cech charakterystycznych dla MVDeepMDP z użyciem metody ablacji. Wykonane badania wskazują na istotność mechanizmu gPoE, którego zastosowanie proponuje Doktorant. Autor próbuje także ocenić cechy zaprojektowanej struktury sterującej w porównaniu z innymi rozwiązaniami, używając wizualizacji t-SNE. Bada także wpływ przetwarzania danych sensorycznych, opisujących stan manipulatora oraz składowe siły i momenty sił, wskazując, że zindywidualizowane przetwarzanie danych użyte w MVDeepMDP daje najlepsze rezultaty. Dodatkowe wyniki badań przedstawione są w dodatku B.2 i B.3.

Zasadniczą część rozprawy kończy rozdział 5, w którym przedstawiono wnioski końcowe oraz dalsze plany badawcze Doktoranta.

3. Ogólna ocena rozprawy i uwagi dyskusyjne

Uważam, że rozprawa doktorska obejmuje istotne zagadnienia związane ze współczesną robotyką manipulacyjną, w której dąży się do zwiększenia stopnia autonomii poprzez wprowadzenie inteligentnych systemów przetwarzania informacji sensorycznej oraz sterowania w warunkach tylko częściowej znajomości modelu środowiska i obiektów, które podlegają manipulacji, przy jednoczesnym uwzględnieniu ograniczeń w przestrzeni zadania. Autor podjął się dobrze uzasadnionego zadania badawczego, które obejmowało implementację zaawansowanej architektury sterowania, w tym przygotowanie oprogramowania integrującego różne podsystemy obliczeniowe i urządzenia fizyczne (sensory, manipulator). Przede wszystkim skoncentrował się na opracowaniu koncepcji oraz projektowaniu nowych algorytmów, a także przeprowadzeniu testów eksperymentalnych, które weryfikują właściwości badanych metod.

Warto podkreślić wysoki poziom warsztatu badawczego Doktoranta, pożądany w dyscyplinie AEEiTK. Myślę tutaj przede wszystkim o sposobie przedstawienia problemu z użyciem podejścia formalnego, uwzględniającego modelowanie procesu i opis zagadnienia w języku matematyki, odwołującym się przede wszystkim do modeli probabilistycznych. Chciałbym docenić taką formę, zwłaszcza w kontekście metod uczenia maszynowego, gdzie w wielu pracach nakierowanych przede wszystkim na wyniki praktyczne spotyka się pobeżny, często nie poparty żadną analizą opis rozwiązania problemu, co może sprawiać wrażenie trywializacji podejścia opartego o standardy naukowe. Tymczasem metody uczenia maszynowego to nie tylko masowe przetwarzanie informacji i implementacja, ale przede wszystkim wiedza mająca solidne podstawy. Dlatego rozważania Autora pracy, poparte szczegółowymi odwołaniami do kluczowych pozycji literatury przedmiotu wyraźnie wzmacniają badawczy charakter tej dysertacji.

Docenić należy również dyskusję, którą Doktorant prowadzi w kontekście wyjaśniania wpływu nowych rozwiązań proponowanych w swojej metodzie Multimodal Variational DeepMDP, w tym badanie ablacyjne stosowane w metodach sztucznej inteligencji. Wartością dodaną są wyniki ilustrujące efektywność reprezentacji wiedzy w MVDeepMDP, szczególnie w kontekście adaptacji wstępnie wyuczonej i nastrojonej architektury do nowych warunków, wynikających z niepewności pozycjonowania, występowania dodatkowych elementów na obrazie, czy też zmiany środowiska i operacji na znacząco innych obiektach w zadaniu manipulacji typu *peg-in-hole insertion*.

Moim zdaniem poziom merytoryczny rozprawy jest wysoki, co wynika przede wszystkim z rozdziałów 3 i 4, w których Doktorant przedstawia główne osiągnięcia. Rozdział 3 dotyczy zagadnień implementacji i przygotowania stanowiska badawczego, które stanowi ważny element wspierający metodykę badawczą, wymagającą efektywnego przeprowadzenia serii eksperymentów rzeczywistych zarówno na etapie uczenia, jak i walidacji. Ponadto warto podkreślić, że choć wykorzystywana jest tutaj znana metoda RL typu SAC, jej implementacja na obiekcie rzeczywistym, wymagająca adaptacji do właściwości robota i systemu pomiarowego w danym zadaniu montażu przemysłowego, stanowi nietrywialne, oryginalne osiągnięcie inżynierskie. Uzyskane wyniki walidacji, przeprowadzone na zdefiniowanych zbiorach płytek drukowanych oraz elementach w warunkach zastosowania percepcji jedno- i wielomodalnej, mogą być postrzegane w kontekście badawczym. Dowodem na to jest publikacja z 2023 roku w czasopiśmie *Journal of Intelligent & Robotic Systems*, w której Doktorant jest pierwszym autorem i w której przedstawiono wyniki zastosowania zaimplementowanej architektury w rozważanym zadaniu manipulacji.

Niewątpliwie głównym wynikiem o charakterze naukowo-badawczym jest autorska propozycja architektury MVDeepMDP, której nowe elementy zostały właściwie przedstawione na tle aktualnej wiedzy. Należą do nich:

- zastosowanie uogólnionego mechanizmu Product-of-Experts (gPoE) do obliczania wspólnej reprezentacji ukrytej, w celu poprawy ekstrakcji cech istotnych na poziomie percepcji;
- włączenie uczenia dynamicznej reprezentacji oraz przewidywania nagrody dla każdej modalności obserwacji, co skutkuje poprawą transferowalności agenta RL i uogólnienia wyników;
- użycie niezależnego przetwarzania z różnych sensorów, co prowadzi do bardziej informatywnej reprezentacji ukrytej w porównaniu do podejścia polegającego na łączeniu danych sensorycznych i ich wspólnym przetwarzaniu.

Algorytm MVDeepMDP został zaprezentowany przez Doktoranta w jednoautorskiej publikacji w renomowanym czasopiśmie *IEEE Robotics and Automation Letters*, co niewątpliwie wzmacnia argument o nowatorskich elementach zawartych w rozprawie oraz uzyskaniu nowych wyników w obszarze inteligentnego sterowania robotami manipulacyjnymi.

Biorąc pod uwagę zawartość rozdziałów 3 i 4 mogę stwierdzić, że cele badawcze zarysowane w podrozdziale 1.3 zostały spełnione. Główne wyniki całej rozprawy dotyczą:

- propozycji nowego algorytmu Multimodal Variational DeepMDP, zaprojektowanego w celu poprawy transferowalności agentów RL do nowych obiektów na liniach produkcyjnych, który opiera się wielomodalnym mechanizmie uczenia reprezentacji dynamicznych w przestrzeni ukrytej z bezmodelowym algorytmem RL działającym poza polityką, Soft Actor-Critic;
- przeprowadzenia analizy wpływu mechanizmu fuzji wielomodalnej na wydajność polityki w RL, zdolności do generalizacji oraz zdolności transferu wyuczonego agenta dla manipulacji na innych obiektach oraz zmienionego środowiska;
- opracowania i implementacji wydajnej modułowej architektury do trenowania agentów RL na rzeczywistych systemach robotycznych, wspartej klasycznymi technikami sterowania pozycyjno-siłowego.

Chciałbym podkreślić, że rozprawa jest starannie zredagowana i napisana poprawnie w języku angielskim. Układ pracy jest logiczny, a sposób prezentacji wyników jest czytelny. Warto też docenić szczegółowe opisy pod rysunkami, które ułatwiają przyswajanie treści dysertacji. Sam dokument nie jest obszerny, ale treść została przedstawiona w sposób zwarty, co daje wrażenie zamkniętej i przemyślanej całości. Oczywiście, ta zwartość ma również pewne ujemne konsekwencje, do których wrócę w dalszej części recenzji. Moim zdaniem, Doktorant w swojej rozprawie odpowiednio zrównoważył aspekty badawcze i techniczne, co w naukach inżynierjno-technicznych jest szczególnie pożądane. Prawidłowo operuje opisem teoretycznym, jednocześnie mając świadomość pewnych uproszczeń stosowanych w celu implementacji metod i zmniejszenia kosztów obliczeniowych. Trafnie formułuje scenariusze eksperymentalne i przeprowadza rzetelną, krytyczną ocenę wyników.

W trakcie lektury rozprawy nasunęły się następujące uwagi krytyczne:

1. Biorąc pod uwagę wyniki rozprawy, nie jestem przekonany, czy decyzja o rezygnacji ze sformułowania tezy badawczej była najwłaściwsza. W podejściu akademickim teza pełni istotną rolę, umożliwiając wprowadzenie myśli przewodniej, wokół której koncentruje się cały proces badawczy. Uwzględniając osiągnięte przez Doktoranta wyniki, uważam, że sformułowanie takiej tezy i jej następne poparcie byłoby w pełni możliwe. Oczywiście pewnym substytutem tezy są kluczowe zagadnienia zdefiniowane w podrozdziale 1.3, jednak mają one raczej charakter ostrożnego wyrażenia intencji rozprawy niż jednoznacznej hipotezy badawczej.
2. Tak jak zaznaczyłem wcześniej, dokument rozprawy nie jest obszerny, choć obejmuje szeroki zakres tematyczny. Widać to szczególnie w odniesieniu do rozdziału 4, w którym Doktorant przedstawia swoje główne osiągnięcie badawcze. Opiera się tutaj wyraźnie na publikacji w IEEE RAL, która, ze względu na wymagania czasopisma, musi być ograniczona do kilku stron. Niestety, w rozdziale 4 Doktorant nie poszerzył istotnie opisu proponowanego rozwiązania, utrzymując zwięzły styl narracji. W rezultacie odnosi się wrażenie, że architektura MVDeepMDP jest przedstawiona zbyt skrótowo. Jest to szczególnie widoczne, gdy rozdział 4 porównamy z rozdziałem 3, w którym znajduje się opis i analiza prostszej wersji RL, w której nie zastosowano dodatkowej augmentacji danych obserwacyjnych. Mimo że algorytm MVDeepMDP stanowi rozwinięcie DeepMDP i korzysta z wprowadzonej w rozdziale 3 metody RL SAC, można oczekiwać bardziej szczegółowej dyskusji. Przykładowo, w opisie metody, pseudokodzie Algorytmu 7 oraz zależnościach (4.9)-(4.14) nie jest widoczne wykorzystanie estymat poszczególnych funkcji nagrody, oznaczonych na rys. 4.1 przez $\tilde{r}_t^1, \dots, \tilde{r}_t^M$. Ponadto,

- w podrozdziale 4.1.3 nie omówiono szczegółowo przetwarzania danych wejściowych (stan układu, sygnały pomiarowe sił/momentów sił) – opis jest bardziej lakoniczny niż ten w rozdziale 3.4, gdzie takie wyjaśnienia również nie są bardzo szczegółowe. Warto byłoby także wyjaśnić, jakie były wymiary wejść i jak reprezentowano te sygnały. Dość pobieżnie omówiono również problem uczenia architektury MVDeepMDP w podrozdziale 4.2.1.
3. W badaniach wykorzystujących uczenie maszynowe powszechnie pojawia się problem falsyfikacji wyników. Biorąc pod uwagę pewne niedopowiedzenia w opisie algorytmów (odwołanie do komentarza 2) oraz brak szczegółowych uwag implementacyjnych (często o charakterze technicznym), mam wątpliwości, czy architektura (w kontekście oprogramowania) przedstawiana w dysertacji jest w pełni replikowalna. W związku z tym można zapytać, czy Doktorant rozważa udostępnienie swoich rozwiązań oraz bibliotek programowych na zasadzie wolnego dostępu lub na warunkach komercyjnych?
 4. Wydaje się, że porównanie schematu SAC z metodami klasycznymi przedstawione w podrozdziale 3.5.5 powinno zostać uzupełnione o informację, czy algorytm SAC był testowany z wykorzystaniem systemu wizyjnego, czy też bez niego. Użycie sprzężenia zwrotnego z wizji może znacząco poprawić jakość realizacji zadania manipulacji, dając istotną przewagę w procesie decyzyjnym. Najprostsze strategie, zwłaszcza w wersji *Drop Down*, bez wsparcia wizyjnego mogą nie zapewnić oczekiwanych wyników, jeśli informacja o położeniu celu manipulacji jest niedokładna. W tym kontekście wyniki przedstawione na rys. 3.13 wydają się przewidywalne. Warto również zauważyć, że w analizie porównawczej nie podano dokładnych wartości parametrów ruchu dla strategii poszukujących. Ważniejszym zagadnieniem jest jednak kwestia, czy w przypadku zastosowania precyzyjnego systemu wizyjnego, dokładnego określenia współrzędnych celu oraz użycia strategii *Drop Down* w połączeniu ze sterownikiem admitancyjnym, jakość realizacji zadania mogłaby dorównać strategii opartej na RL SAC. Rozszerzenie tej analizy mogłoby dostarczyć informacji na temat realnych korzyści wynikających z zastosowania metod uczenia ze wzmocnieniem w porównaniu do klasycznych podejść.
 5. W rozprawie widać rzetelne próby interpretacji algorytmów wykorzystujących sieci neuronowe, co wpisuje się we współczesny nurt znany jako wyjaśnialna sztuczna inteligencja (ang. *Explainable AI*). Zrozumienie metody oraz parametrów strojenia jest niewątpliwie istotne w kontekście sterowania procesami dynamicznymi. Moim zdaniem w rozprawie zabrakło jednak pewnych szczegółowych uwag dotyczących wyboru parametrów, w tym hiperparametrów sieci neuronowych. Choć wartości tych parametrów zostały podane, pojawia się pytanie, czy Doktorant przeprowadził analizę doboru parametrów oraz wymiarów sieci, czy też opierał się na ogólnych zaleceniach z literatury lub swoim doświadczeniu? Przykładowo, na s. 42 Doktorant wspomina, że docelowa wartość entropii została przyjęta równa $-\frac{1}{2} \dim(\mathcal{A})$, zamiast $-\dim(\mathcal{A})$, jak sugeruje literatura. Podobnie, standardowo przyjęto współczynniki zbieżności uczenia równe 0,0003 dla optymalizatorów Adam i AdamW (por. s. 42 i s. 57), z kolei na s. 40 podano, że w procesie uczenia wybrano $\log(\sigma) = [-11,5; 1,5]$. Na s. 56 zapisano, że parametr $\beta = 1$ – co sprowadza regułę (2.31) do (2.28) – oraz $\sigma = 0,2$. Z jakich powodów dokonano tych wyborów?
 6. Sądzę, że dobre wyniki uzyskane przez zaimplementowane metody RL były możliwe dzięki zastosowaniu pomocniczego sterownika admitancyjnego. Wartości nastaw tego sterownika, zestawione w tabeli 3.1 na s. 38, sugerują, że zostały one ustalone jednorazowo. Czy ich dobór był poprzedzony testami oceniającymi jakość działania sterownika? Wydaje się, że w kontekście zastosowanego podejścia można by dodatkowo rozważyć problem przestrajania tego regulatora oraz wprowadzenie dodatkowej polityki kształtującej wartości parametrów. Czy taka możliwość była brana pod uwagę przez Doktoranta?

7. Tytuł rozdziału 3.1 jest nieco mylący ponieważ na początku Autor przedstawia opis zrobotyzowanego stanowiska i jego poszczególnych elementów – nie ma to zatem wiele wspólnego z używaną strategią sterowania. Dopiero w dalszej części pojawiają się odniesienia do RL. Myślę, że warto byłoby przedstawić podrozdział dotyczący samego stanowiska, a w kolejnym scharakteryzować algorytm i założenia dotyczące układu sterowania robotem. Pozwoliłoby to na lepszą separację pomiędzy opisem zrobotyzowanego systemu, a samą metodą sterowania.
8. Kontynuując uwagę 7, uważam że funkcję nagrody w rozdziale 3.1 można by przedstawić bardziej formalnie (mniej opisowo), obejmując wszystkie przypadki, włączając w to zależność (3.1) oraz dodatkowe warunki (np. niepowodzenie z uwagi na przekroczenie zakresu kąta obrotu). Nie jest też jasne dlaczego wybrano postać (3.1). Czy testowano również inne funkcje nagrody, które w RL mogą istotnie modyfikować właściwości algorytmu? Wreszcie, czy przeprowadzone testy potwierdziły hipotezę, że w przypadku uszkodzenia elementu jego końcowe dopasowanie będzie niewystarczające i będzie wykrywane przez system jako brak sukcesu w realizacji zadania?
9. Analizując trajektorie przejściowe przedstawione na rys. 3.14, zasadniczo można potwierdzić wnioski Doktoranta przedstawione na s. 48. Mimo to można zauważyć, że w niektórych przypadkach występuje pewne błędzenie oraz pojawiają się niepotrzebne ruchy końcówki w warunkach braku kontaktu elementu ze środowiskiem. Czy według Doktoranta są one wynikiem ograniczonej dokładności pomiaru pozycji, wpływu szumu w pomiarze sił/momentów sił, czy też wyuczonej polityki agenta RL? Jeśli dominującym czynnikiem jest ten ostatni efekt, to możliwa byłaby dalsza poprawa jakości – przynajmniej w początkowej fazie – poprzez zastosowanie prostszego algorytmu sprowadzającego końcówkę w otoczenie punktu docelowego, a następnie przełączenie w tryb sterowania pozycyjno-siłowego. Oczywiście wymagałoby to rozszerzenia możliwości agenta RL lub wprowadzenia dodatkowych reguł decyzyjnych.
10. System sterowania robotem opisany w p. 3.2 zakłada częstotliwość pracy agenta RL z częstotliwością 10 Hz. Czy przyjęta wartość wynika wprost z ograniczonych zasobów obliczeniowych, czy też dobrana jest z dużym zapasem? Wreszcie, czy obserwowano zwolnienie pętli zewnętrznej w wybranych przypadkach?
11. Czy przeprowadzano testy, które pozwalałyby oceniać co stanowi największe ograniczenie w procesie uczenia architektury sterującej? Czy podstawowym czynnikiem wydłużającym czas uczenia jest dynamika manipulatora, czy złożoność algorytmów, a więc czas obliczeń?
12. W opisie architektury na s. 31 Autor wskazuje na celowość zastosowania rozwiązania Ape-X, pisząc, że pozwala ona na rozproszenie uczenia agenta RL z użyciem wielu aktorów. Tymczasem w rozwiązaniu proponowanym przez Doktoranta rozważa się przypadek jednego aktora, odpowiadającego pojedynczemu manipulatorowi. Na czym zatem dokładnie polega zaleta użycia Ape-X, poza rozszerzalnością systemu na przypadek wielu urządzeń?

Część z podanych uwag ma charakter dyskusyjny i nie umniejsza głównych wyników uzyskanych przez Doktoranta w rozprawie. Niektóre z nich mogą stanowić przyczynek do dyskusji podczas obrony, a inne mogą stanowić źródło przyszłych prac badawczych poszerzających dorobek publikacyjny Autora.

4. Uwagi szczegółowe

Dokument jest bardzo dobrze zredagowany i jego struktura jest przemyślana. Język użyty w dysertacji jest precyzyjny, choć można znaleźć pewne uchybienia stylistyczne, błędne sformułowania lub naruszenie właściwego stylu narracji.

Poniżej przedstawiam uwagi szczegółowe, dotyczące wybranych stwierdzeń lub opisów występujących w rozprawie:

- Metryki wykorzystywane do oceny jakości architektury, opisane słownie w podrozdziale 3.5.2, mogłyby zostać przedstawione w bardziej sformalizowany sposób – z użyciem wzorów matematycznych oraz zestawione w formie listy lub tabeli. Takie przedstawienie mogłoby ułatwić odwoływanie się do nich w kolejnych częściach rozprawy.
- Na s. 22, wiersz 6 od dołu, stwierdzono że «The SAC algorithm utilizes a Gaussian policy to generate stochastic actions, which, by default, can result in unbounded action values». Sądzę, że można by precyzyjniej określić, co oznaczają nieograniczone wartości akcji. W tym kontekście można domniemać, że opisują one akcje, których wynik wychodzi poza dopuszczalną przestrzeń stanu.
- Na s. 28 stwierdzono: «The robot operates within a cylindrical workspace with a radius of R_{limit} and an infinite height», tymczasem w rozpatrywanym zadaniu wysokość walca jest raczej skończona.
- Na s. 28 znajduje się komentarz dotyczący pozycjonowania końcówki w osi Z: «Furthermore, the initial Z-axis position can be adjusted during the training process to enhance the agent's learning». Sformułowanie to jest niejasne, gdy uwzględni się, że w procesie uczenia zaburzana jest pozycja początkowa pozycja również w pozostałych współrzędnych – por. podrozdział 3.3.1.
- Sądzę, że sformułowanie na s. 32: «In a deterministic policy, the agent selects the action with the highest probability, whereas in a stochastic policy, the agent selects the action based on the probability distribution over the action space» nie wyjaśnia dokładnie różnicy między strategią (polityką) deterministyczną a stochastyczną.
- Określenie «workspace» używane w podrozdziale 3.1 jest nieco mylące, ponieważ odnosi się ono do przestrzeni roboczej manipulatora (por. «The robot operates within a cylindrical workspace»), przestrzeni zadania oraz nazwy układu odniesienia (por. «workspace frame {w}»).
- Na s. 64 moim niepotrzebnie moim zdaniem znajduje się komentarz dotyczący PoE : «is not a new concept in mathematics, but its first application in multimodal variational autoencoders was introduced by Wu and Goodman [125]». Ten opis w naturalny sposób pasuje do punktu 4.1.1.
- Tytuł rozdziału 2.2.3 nie w pełni odpowiada jego treści. Można by go zmienić np. na: «Impedance and admittance control». Brakuje też krótkiej dyskusji kiedy używać pierwszego lub drugiego schematu sterowania.
- W opisie pod równaniem (2.17) pojawia się niepoprawne oznaczenie $\mathcal{V}_f$ zamiast $V_{\tilde{a}}$.
- Równanie (2.26) – jak należy interpretować składnik przyspieszenia $\Delta \ddot{x}_r$? Czy nie powinien on być zastąpiony przez $\Delta \ddot{x}$?
- Analogicznie do uwagi powyżej, w równaniu (3.3) po prawej stronie jest x i $\tilde{x}$, a po lewej $\tilde{x}_r$. Z punktu widzenia zadania sterowania, śledzenie pozycji zadanej zapewnia $x \approx x_r$. Czy wykorzystywano tę własność przy wyznaczaniu trajektorii referencyjnej? Zapis (3.3) nie jest też spójny z notacją użytą w pracy w *Journal of Intelligent & Robotic Systems*.
- Symbol « $\odot$ » opisujący mnożenie blokowe w (2.30) nie został wcześniej zdefiniowany.

- Przykładowe błędy językowe:
 - Powszechnie przyjmuje się, że formuły matematyczne stanowią część zdania i podlegają regułom interpunkcji. W tej dysertacji ta zasada nie jest przestrzegana.
 - Autor powszechnie używa skrótu «let's», które w formalnych dokumentach powinien być zastąpiony przez «let us».
 - s. 39: podpis pod rysunkiem 3.8: «The robot wad commanded» zamiast «The robot was commanded»
 - s. 39, pierwszy paragraf podrozdziału 3.4: «was designed only for state-based observationsonly» zamiast «was designed only for state-based observations» lub «was designed for state-based observations only»
 - s. 53, szósty wiersz nad równaniem (4.7): wektor $\mathbf{e}_i$ jest napisany fontem zwykłym, a nie pogrubionym.

5. Wnioski końcowe

Pomimo uwag, które zawarłem w tej recenzji stwierdzam, że nie są one krytyczne i nie umniejszają wartości wyników badawczych zawartych w dysertacji. Rozprawa potwierdza wiedzę Doktoranta w zakresie metod sterowania manipulatorów, metod uczenia maszynowego ze szczególnym uwzględnieniem paradygmatu RL oraz projektowania i implementacji zaawansowanych architektur dla robotów integrujących sensorykę, percepcję i system sterujący/decyzyjny. Można też zauważyć umiejętność Doktoranta w zakresie łączenia klasycznych metod z nowymi technikami tzw. sztucznej inteligencji. Jego warsztat badawczy i wyniki uzyskane w rozprawie są związane bezpośrednio z dyscypliną AEEiTK. Autor dostarczył oryginalnych wyników badawczych i wykazał się zdolnością do samodzielnego rozwiązania problemu badawczego w naukach technicznych oraz krytycznej oceny uzyskanych wyników.

Duża część rezultatów przedstawionych w rozprawie, została opublikowana w prestiżowych czasopismach w obszarze związanym z robotyką i automatyką. Warto przypomnieć, że w jednej z tych publikacji Doktorant jest jedynym autorem. Dodatkowo należy zwrócić uwagę na kwalifikacje inżynierskie Autora rozprawy, które mają bardzo duże znaczenie w kontekście rozwiązywania problemów w robotyce, których inspiracja jest praktyczna i mogą być cenne w kontynuowaniu Jego działalności badawczej. Wydaje się, że tematyka doktoratu ma duży potencjał i może być kontynuowana w kierunku poszukiwania nowych rozwiązań wspierających zadania manipulacji z użyciem wielu robotów oraz innej klasy sensorów. Ponadto wydaje się, że proponowane rozwiązania mogą być przeniesione również na obszar robotyki mobilnej lub zautomatyzowanych/autonomicznym pojazdów.

Biorąc pod uwagę powyższe stwierdzam, że rozprawa doktorska mgr. inż. Grzegorza Bartyzela pt. „Zastosowanie głębokiego uczenia ze wzmocnieniem w robotycznym montażu przemysłowym” spełnia w stopniu wystarczającym wymagania stawiane rozprawom doktorskim przez aktualnie obowiązującą ustawę „Prawo o szkolnictwie wyższym i nauce” (Dz.U. z 2024 r. poz. 1571). Wniosuję o jej przyjęcie i dopuszczenie do dalszych etapów postępowania doktorskiego, w tym do jej publicznej obrony.

Pawel Pardała