

Częstochowa, dn. 5 kwietnia 2024 r.

prof. dr hab. inż. Rafał Scherer  
Katedra Inteligentnych Systemów Informatycznych  
Wydział Inżynierii Mechanicznej i Informatyki  
Politechnika Częstochowska  
al. Armii Krajowej 36  
42-200 Częstochowa

**S E K R E T A R I A T**  
Rady Dyscypliny AEEiTK

Wpłynęło dnia ..... 10.04.2024 .....

Zarejestrowano pod nr .....

Podpis ..... 

### Recenzja

rozprawy doktorskiej mgr inż. Mariusza Karola Nowaka, pt.: Depth Completion for FMCW Radars.

Niniejszą recenzję opracowano na wniosek Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne AGH. Promotorem jest dr hab. inż. Paweł Skruch, prof. AGH.

## 1. Charakterystyka tematu, celu i tezy badawczej rozprawy

Dane z radaru i LiDARu są niezwykle ważne z wielu powodów, zwłaszcza w kontekście technologii, badań naukowych, oraz bezpieczeństwa. Zarówno radar, jak i LiDAR pozwalają na skuteczne monitorowanie otoczenia, są niezbędne w nawigacji, zarówno w lotnictwie, morskiej, jak i nawigacji samochodowej. Pomagają one w wykrywaniu przeszkód, określaniu odległości i kierunku, oraz w zapobieganiu kolizjom. Rozdzielczość przestrzenna radaru może być ograniczona. Zagęszczanie danych pozwala zwiększyć ilość informacji, co może być istotne w celu uzyskania bardziej szczegółowego obrazu otoczenia służącego do podejmowania decyzji w czasie jazdy autonomicznej.

## 2. Zawartość rozprawy

Recenzowana praca mgr inż. Mariusza Karola Nowaka składa się ze streszczenia w języku angielskim i polskim, spisu skrótów, spisu symboli, ośmiu rozdziałów, bibliografii spisu rysunków oraz spisu tabel. Dokument liczy 169 stron.

Pierwszy rozdział jest wprowadzeniem do tematyki. Autor opisał ideę systemów autonomicznej jazdy polegającą na podejmowaniu decyzji na podstawie sygnałów z czujników, np. kamery,

radaru czy LiDARu oraz informacji o lokalizacji pojazdu. Następnie algorytm percepcji tworzą reprezentację środowiska potrzebną dla algorytmów planowania, a następnie dla algorytmów sterowania, które wydają komendy dla aktuatorów.

Rozdział 2 omawia metody pomiaru odległości stosowane w pojazdach autonomicznych. Pierwszym z nich jest LiDAR mierzący czas powrotu światła odbitego od obiektu zazwyczaj przez skanowanie przestrzeni laserem. Jego odmianą jest kamera typu Time-of-Flight bez części ruchomych. Następnie opisano radar CW oraz dokładnie FMCW używany w badaniach ze względu na możliwość uzyskania informacji od odległości. Omówiono techniki CFAR wykorzystywane w radarach do detekcji obiektów w obecności szumu i zakłóceń. Omówiono własności i budowę RDC oraz sposoby wyznaczania kąta do obiektu przy użyciu wielu anten odbiorczych. Omówiono pokrótce czujniki ultradźwiękowe oraz możliwości szacowania odległości z sensorów nie dających wprost takiej informacji, takich jak stereowizja oraz pojedyncze zdjęcia ze zbiorem KITTI Depth Prediction jako przykładem wokół którego skupione jest wiele metod z literatury.

Rozdział 3 omawia poprawianie jakości działania sieci neuronowych poprzez drobną zmianę wartości wag w czasie uczenia. Rozpoczyna się od najnowszej historii sieci neuronowych, to jest od czasów pierwszych sieci spłotowych oraz pierwszych eksperymentów z akceleratorami GPU – AlexNet, VGG, ResNet, oraz badań w kierunku zmiany struktury sieci, takich jak Inception czy MobileNet. Autor wspomina metody usuwania wag (pruning), nadmienając, że i tak w praktyce nie zmniejszają one złożoności obliczeniowej sieci neuronowych. Dlatego też, Autor proponuje nową metodę polepszania jakości działania sieci. Następnie opisywane są metody regularyzacji uczenia w sieciach neuronowych oparte na dodawaniu różnego rodzaju szumu do danych uczących. Najpopularniejsze metody wbudowane są w biblioteki służące do budowania i uczenia sieci neuronowych. Dalej omówiono ideę zmniejszania wielkości sieci oraz uczenie podsieci nazwane lottery ticket z 2018, działające głównie w sieciach w pełni połączonych. Podrozdział 3.3 przedstawia autorską metodę uczenia sieci. Analiza literatury pokazała, że może wystąpić sytuacja, kiedy część wag w sieci neuronowej może nie podlegać optymalnemu uczeniu i nie wykształcić użytecznej reprezentacji wiedzy. Jednym z rozwiązań tego problemu może być różnorodności wag co potwierdziły wstępne eksperymenty wykonane przez Doktoranta z uczeniem problemu XOR stosunkowo małej sieci o różnej liczbie neuronów. Dzięki temu wstępnemu sprawdzeniu można było przejść do poważniejszych eksperymentów z użyciem pięciu znanych modeli z lat 2016-2019, a mianowicie: MobilenetV2, DenseNet201, ResNet152V2, Xception, EfficientNet B0 zaimplementowanych w bibliotece Keras, nauczonych wstępnie zbiorem ImageNet. Modele te były uczone zbiorem CIFAR100 używając zaburzeń wag, metody lottery ticket, połączenia tych metod oraz bez żadnej modyfikacji uczenia. Ponadto, zmieniano długość uczenia i współczynnik uczenia. Eksperymenty pokazały, że zaburzanie wag przez dodawanie szumu poprawiało dokładność uczenia i, co za tym idzie, działania testowanych modeli sieci.

Rozdział 4 omawia autorskie rozwiązanie uzupełniania informacji o głębokości danych z LiDARu. Opisano zmodyfikowany zbiór danych KITTI oraz metodę obliczania operacji spłotu dla braków w pikselach Sparsity Invariant Convolutions zaproponowane w literaturze. Dalej omówiono prace używającej tej metody w innych modelach sieci neuronowych, oraz inne metody uzupełniania informacji o głębi. Następnie zaprezentowano autorską sieć neuronową opartą na blokach nazwanych WeaveBlock, złożonymi z operacji spłotu punktowego z małą i dużą liczbą kanałów oraz spłotami z długimi filtrami. Były one inspirowane wspomnianymi Sparsity Invariant Convolutions oraz spłotami „głębokościowymi” z sieci MobileNet. Trzy takie bloki tworzą blok rezydualny znany z sieci ResNet z połączeniem omijającym te trzy bloki. Bloki rezydualne są połączone szeregowo, tworząc się tzw. nienaprowadzaną sieć WeaveNet. Wersja „naprowadzana” tej sieci jest zbudowana podobnie i dostarcza ona w

połączeniu z siecią neuronową będącą koderem obrazu z kamery sygnału dla sieci agregującej oba sygnały, która tworzy gęstą mapę głębokości. Autor uczył zaproponowaną sieć neuronową na trzy sposoby. Pierwszy był standardowym sposobem uczenia składającym się z uczenia wstępnego i standardowego uczenia gęstości wejściowej. Drugi sposób dodawał jeszcze uczenie ze zmienną wejściową gęstością i trzeci dodawał czwarty etap z zadaną gęstością wejściową. Rozdział kończy opis eksperymentów porównujących proponowany model z metodami z literatury dla różnych gęstości danych wejściowych.

Rozdział 5 jest opisem eksperymentów sprawdzających czy sygnał radiowy odebrany z anteny z układem fazowym może być źródłem informacji dla sieci neuronowej do ekstrakcji kąta i amplitudy sygnału. Doktorant użył trzech anten: 5-, 10- i 20-elementowej, co pozwoliło sprawdzić wpływ anteny na informację o kącie sygnału. Wejściami do sieci neuronowej była część rzeczywista i urojona sygnału na każdej z anten odbiorczych. Wyjściami z sieci były kąt i amplituda do źródła większej fali oraz mniejszej fali, oraz kąty zdefiniowane jako rozkład prawdopodobieństwa. Sieć z czterema warstwami ukrytymi była trenowana metodą z perturbacją wag z rozdziału 3. Największa antena okazała się być wyraźnie lepsza w zadaniu predykcji kąta od swoich mniejszych odpowiedników.

Rozdział 6 omawia dane zebrane do badań z około 14 godzin jazdy samochodem testowym. Radar posiadał 12 anten odbiorczych. Około miliona ramek zostało podzielonych na dane uczące i testowe w stosunku 90/10. Do uzupełnienia głębokości z LiDARu przy użyciu wstępnie wytrenowanej sieci neuronowej na danych głębokości ze zbioru KITTI Depth Completion. Następnie wyjaśniono problemy wynikające z różnic działania radaru i LiDARu, co powoduje między innymi różnice średniej odległości w rejestrowaniu obiektów. Między innymi, było to też powodem różnic jakościowych zbiorów rejestrowanych w jeździe miejskiej i autostradowej.

Rozdział siódmy jest opisem autorskiej metody uzupełniania informacji o odległości danych z radaru. Rozdział zaczyna się opisem badań światowych, poczynszyszy od możliwości uzupełniania informacji o odległości z obrazowania dwuwymiarowego. Następnie omówiono rozwiązania łączące dane z radaru z danymi dwuwymiarowymi, geometryczne oraz oparte o sieci neuronowe.

Autorskie rozwiązanie jest oparte o sieć neuronową o trzech wejściach: zestawione (stacked) RDC z trybu dalekiego zasięgu radaru, zestawione RDC z trybu krótkiego zasięgu radaru, zestawione odpowiadające im obrazy z kamery RGB (trzy kolorowe obrazy 720×1280). Jedna wersja sieci wykorzystująca tylko dane radarowe niskiego poziomu, a drugą łączyła dane radarowe niskiego poziomu z obrazami RGB na płaszczyźnie azymut-elewacja. Zbadana została również wersja sieci przewidująca głębokość wyłącznie na podstawie obrazów RGB, aby określić w jakim stopniu radar jest korzystny gdy jest scalony z obrazami RGB z kamery. System był złożony z modułu przewidywania kąta, enkodera zakresu i względnej prędkości RDC, modułu niepodlegającego uczeniu do zmiany rozmiaru wejścia enkodera RDC do mapowania go na dwuwymiarową płaszczyznę azymutu-elewacji, enkodera obrazów z kamery, modułu łączącego wyjścia powyższych modułów oraz modułu jednocześnie przetwarzającego zakodowane dane RDC z zakodowanym obrazem z kamery.

W module przewidywania kąta dane z RDC przetwarzane są w kilku warstwach splotowych z filtrami o rozmiarze 1×1. Sygnał z jednej z poprzednich warstw trafia do ostatniej warstwy jako tzw. skip-connection. Enkoder prędkości i zasięgu na podstawie danych z RDC, tworzy gęstą reprezentację z użyciem sieci opartej na MobileNetV3, następnie kilku warstw splotowych 2D z tzw. „strided convolutions”, i ostatecznie przycięcia i warstwy typu „pointwise convolution”.

Moduł zmiany danych w płaszczyznę azymutu-elewacji nie podlega uczeniu i jest zestawem operacji na danych. Moduły enkodowania danych z kamer oparte są na sieciach MobileNetV3

i tworzą mniejszą reprezentację danych. Moduł łączenia danych z czujników i czasowych łączy dane z RDC dalekiego i krótkiego zasięgu oraz przetworzony obraz z kamer. Połączone dane trafiają na MobileNetV3, a aspekt czasowy analizowany jest przez warstwy rekurencyjne GRU. Wyjście jednej warstw trafia na wyjście jako skip-connection. Reprezentacja tworzona w module przetwarzania płaszczyzny azymutu-elewacji tworzona przez warstwy tzw. „transposed convolution” służy jako wejście do warstw klasyfikujących i do warstw regresji, tworzących tzw. dwie „głowy” sieci (heads).

Powyższy system testowany był w czterech konfiguracjach: sieć wykorzystująca jedynie radar, z użyciem trzech RDC krótkiego zakresu oraz trzech RDC długiego zakresu, taka sama konfiguracja z 4+4 RDC, oraz dwie konfiguracje z RDC i ramkami wizyjnymi.

Rozdział 8 jest podsumowaniem rozprawy z wyszczególnieniem autorskich osiągnięć.

Dalej następuje podsumowanie pracy oraz bibliografia składająca się ze 120 pozycji oraz spis rysunków i tabel.

### 3. Ocena rozprawy

W ramach rozprawy doktorskiej Doktorant zaproponował oryginalne rozwiązanie z analizą danych dotyczących sceny widzianej przez samochód autonomiczny z wykorzystaniem bieżących metod uczenia maszynowego. Tezy pracy wspomniane wcześniej są trafnie i jasno sformułowane, tematyka pracy jest aktualna. Doktorant wykorzystuje bieżące metody związane z głębokim uczeniem, sieciami neuronowymi i przetwarzaniem sygnału. Oryginalny dorobek autora polega na stworzeniu

- metody wykorzystujący perturbacje wag o środku w zerze podczas treningu sieci neuronowych,
- sieci neuronowej WeaveNet do uzupełniania głębokości LiDARu przy danych o zmiennej gęstości oraz wykazaniu eksperymentalnie, że sieci neuronowe są zdolne do znajdowania kątów do dwóch celów, gdy anteny radarowe odbierają superpozycję fal odbitych,
- zbioru danych służącego do trenowania i walidacji algorytmów uzupełniania głębokości radarowej, składającego się on z ponad 1 000 000 RDC, wraz z odpowiadającymi im obrazami z kamery i gęstymi mapami głębokości wygenerowanymi przez sieć WeaveNet,
- architektury sieci neuronowej zdolnej do przekształcania niskopoziomowych danych wejściowych RDC w abstrakcyjne kanały w płaszczyźnie azymutu-elewacji,
- architektury sieci neuronowej zdolnej do łączenia danych pochodzących z RDC z radaru i danych pochodzących z obrazów kamery RGB w celu uzupełniania informacji o głębokości z radaru.

Zaprezentowany materiał pokazuje, że Doktorant zrealizował cel pracy.

Rozprawa doktorska uwidacznia ogólną wiedzę teoretyczną i praktyczną mgr inż. Mariusza Nowaka. Doktorant opublikował dwie prace naukowe – jedną w czasopiśmie z listy JCR oraz jedną w materiałach konferencji. Rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej mgr inż. Mariusza Nowaka. Rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego. Zaproponowane metody mają bardzo duże znaczenie aplikacyjne dla nauk technicznych i przemysłu.

Niezależnie od mojej bardzo wysokiej oceny pracy, wykraczającej poza poziom przeciętny, poniżej wymieniam drobne uwagi, które nasunęły się w czasie czytania:

Dlaczego sieć unguided WeaveNet ma akurat pięć bloków rezydualnych, a jeden blok ma trzy elementy WeaveBlock? (str. 69)

Sieć neuronowa w Rozdziale 5 uczona jest 10 milionami próbek. Skąd wiadomo, że trzeba aż tylu przykładów i dlaczego przyjęto akurat 128 neuronów w warstwach ukrytych i funkcję aktywacji swish?

Jak wygląda dokładnie struktura sieci z rysunku 7.2, 7.3, 7.4, 7.5, 7.6 (nie jest podana liczba filtrów, rodzaj nieliniowości itd.). Dlaczego liczba warstw omijanych przez skip-connection jest akurat taka? W jaki sposób zaimplementowane są zaprezentowane sieci neuronowe i jakie są możliwości działania ich w czasie rzeczywistym w samochodzie?

Być może dobrze byłoby sprawdzić pracę pod kątem językowym, np. „Input to this module has dimension [3, 420, 256, 24] and output has dimension [3, 420, 256, 96]. “ -> “The input to this module has dimensions [3, 420, 256, 24] and the output has dimensions [3, 420, 256, 96].”

str. 56: In case of all networks -> In the case of all networks

Liczebniki o małych wartościach piszemy często słownie.

Znak myślnika w zdaniu to znak typograficzny pauza lub półpauza, a nie w dywiz.

Może ładniej jest używać znaku  $\times$  przy oznaczaniu rozdzielczości, np. na str. 56: 32x32 px, shapes 31x1 and 1x31.

Niektóre rozdziały rozpoczynają się bezpośrednio podrozdziałami; być może ładnie byłoby zacząć je kilkoma zdaniami wstępu.

#### 4. Wnioski końcowe recenzji

Podsumowując recenzję stwierdzam, że Pan mgr inż. Mariusz Nowak w rozprawie doktorskiej „Depth Completion for FMCW Radars” zrealizował cel rozprawy. Zaprezentowane rezultaty stanowią oryginalny wkład Autora w rozwój dyscypliny Informatyka Techniczna i Telekomunikacja. Pan Mariusz Nowak wykazał się umiejętnością samodzielnej pracy badawczej, znajomością literatury światowej i wiedzą w zakresie uczenia maszynowego. Recenzowana praca spełnia wymagania ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, w dyscyplinie naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne. Wnoszę o jej przyjęcie i dopuszczenie do dalszych etapów postępowania doktorskiego.



