Streszczenie

Jazda autonomiczna to jeden z głównych tematów badawczych w branży motoryzacyjnej. Pełna automatyzacji procesu prowadzenia pojazdu może przynieść znaczne korzyści, obejmujące wyższy komfort użytkowania takiego pojazdu i poprawę ogólnego bezpieczeństwa na drogach. Nowe narzędzia i postęp technologiczny umożliwiają tworzenie coraz bardziej zaawansowanych systemów, które starają się w pełni zrealizować możliwości pojazdów autonomicznych. W tym nieustannie zmieniającym się obszarze, czujniki takie jak kamery, LiDAR i Radar odgrywają istotną rolę w systemach percepcji, odpowiedzialnych za funkcje poznawcze takich pojazdów. Czujniki te pełnią rolę oczu i uszu systemów percepcji, rejestrując kluczowe dane z otoczenia w postaci obrazów czy chmur punktów. W niniejszej rozprawie przedstawiona jest dogłębna analiza takich czujników, koncentrująca się zarówno na ich budowie, jak i formacie dostarczanych przez nie danych. Jest to szczególnie istotne z uwagi na to, że stanowią one dane wejściowe dla dedykowanych algorytmów percepcji, które tworzą cyfrowy model otoczenia.

Wśród tych algorytmów, metody uczenia maszynowego zyskały w ostatnim czasie znaczną popularność w odniesieniu do systemów percepcji. Zwłaszcza w problemach związanych z wykrywaniem obiektów, poprzez wnikliwą analizę danych z czujników, takie rozwiązania dominują obecnie w aplikacjach przemysłowych. Osiągają one doskonałe wyniki i oferują perspektywy dalszych udoskonaleń. Temat badawczy tej pracy skupia się właśnie na systemach percepcji pojazdów autonomicznych opartych o metody uczenia maszynowego, a konkretnie konwolucyjne sieci neuronowe i podejście głębokie do ich projektowania. Te techniki znacznie usprawniły dziedzinę percepcji, umożliwiając pojazdom precyzyjne postrzeganie otoczenia. Dzięki wykorzystaniu nowoczesnych architektury omówionych w tej pracy, takie sieci są w stanie odkrywać skomplikowane wzorce i reprezentacje z danych pochodzących z czujników, co prowadzi do lepszego zrozumienia badanego otoczenia.

Ponadto, metodyka percepcji może znacząco skorzystać z dopełniających się informacji dostarczanych z różnych źródeł. W kontekście takiego rozwiązania, każdy czujnik ma swoje zalety i wady, a połączenie różnych urządzeń w zestawie czujników pojazdu autonomicznego może złagodzić ich niepożądane cechy, co w rezultacie poprawi cały system. Kierując się tą myślą, niniejsza praca badawcza skupia się przede wszystkim na wykorzystaniu fuzji danych z czujników i eksploracji jej korzyści. Koncepcja fuzji danych jest szczegółowo omówiona wraz z różnymi metodami łączenia danych z czujników. Spośród tych metod szczególnie interesującym jest specjalny typ niskopoziomowej fuzji danych, który naturalnie dobrze współgra z podejściem przetwarzania danych z użyciem sieci neuronowych. Głównym celem tej rozprawy doktorskiej jest określenie, czy oparte na metodach uczenia maszynowego rozwiązanie fuzji niskopoziomowej może okazać się korzystne dla systemu percepcji pojazdu autonomicznego w zadaniu związanym z wykrywaniem obiektów na drodze. Fuzja ta jest przeprowadzana na obrazach z kamer oraz danych chmur punktów z czujników LiDAR lub radaru.

Kluczową innowacją w dążeniu do tego celu jest nowatorskie podejście do fuzji niskopoziomowej, nazwane metodą Cross-Domain Spatial Matching. Metoda CDSM oferuje alternatywną technikę fuzji, niespotykaną dotąd w tej dziedzinie badań. Składa się ona z dwóch głównych elementów: dopasowania domen danych z czujników i metod fuzji. Pierwszy komponent zajmuje się ujednoliceniem odczytów czujników związanym z różnymi orientacjami próbek w stosunku do wspólnego układu współrzędnych pojazdu. Po takim dopasowaniu możliwe staje się łączenie danych z różnych czujników bez dodatkowych rzutowań i konwersji. Pozwala to także na bezpośrednie zastosowanie zaproponowanych strategii fuzji. W tej pracy badawczej przedstawiono

trzy takie unikalne strategie, bazujące na różnych podejściach i niosące różnorodne korzyści, które są poddane dalszej weryfikacji w kolejnych rozdziałach. Oba elementy CDSM są zintegrowane w stworzonej architekturze sieci neuronowej. Ta integracja ułatwia nie tylko uczenie sieci w tak zwanym podejściu \textit{end-to-end}, ale także przyczynia się do efektywnych czasów inferencji modelu podczas jego wdrożenia.

Aby zaimplementować kompletną architekturę sieci neuronowej związanej z percepcją obiektów, zaprojektowane są również modele przetwarzające dane z pojedynczych czujników, oparte na najnowocześniejszych rozwiązaniach z odpowiadających im dziedzin. Modele te spełniają podwójną rolę: po pierwsze, pozwalają na ocenę wydajności systemu percepcji, gdy opiera się on wyłącznie na jednym czujniku, a po drugie, stanowią integralne moduły w architekturze fuzji, odpowiedzialne za wyodrębnianie cech z każdej próbki danych wejściowych. Poprzez dokładne eksperymenty i odpowiednie techniki oceny jakości, przeprowadzane są dalsze badania dotyczące skuteczności fuzji, wykorzystujące dwa publiczne zbiory danych z dziedziny motoryzacji - KITTI i NuScenes. Praca doktorska zawiera szczegółową charakterystykę tych zbiorów wraz z dokładnym opisem przebiegu procesu uczenia poszczególnych modeli. Ponadto, wyniki modeli bazujących na pojedynczych czujnikach oraz modeli fuzji zostają przedstawione w sposób, który umożliwia porównanie ich ze sobą. Przeprowadzona zostaję dogłębna analiza wyników predykcji obiektów zarówno wizualna, jak i pod względem przyjętych wskaźników jakości. Dla rozwiązania opartego o fuzję, analiza zostaje rozszerzona o badania pokazujące zyski nad jedno-czujnikowymi modelami oraz przykłady przypadków brzegowych, oferujące wgląd w scenariusze, w których model fuzji odbiega od odpowiedników bazujących tylko na jednym źródle danych. Zapewnione jest również porównanie do wiodących rozwiązań fuzyjnych, co ułatwia umiejscowienie metody CDSM wśród obecnie wykorzystywanych technik. Analiza tych wyników pozwala na wyciągnięcie wniosków dotyczących zaproponowanej metody fuzji w systemach percepcji pojazdów autonomicznych oraz sformułowania odpowiedzi na pytanie, czy fuzja niskopoziomowa może okazać się korzystna do tego celu.

6.10.2023 Daniel Dworak

Data i czytelny podpis kandydata

Abstract

Autonomous Driving is a major research topic in the automotive domain. The promise of fully automating the driving process holds the potential to deliver substantial advantages, encompassing heightened user com- fort and a considerable enhancement in overall safety on the roads. New tools and technological advances enable gradually more sophisticated systems, which try to closely reassemble the entirety of Autonomous Vehicle capabilities. Within this transformative area, sensors such as cameras, LiDAR and Radar play an essential role in perception systems, which corresponds to the cognitive functions of an AV. These sen- sors serve as the eyes and ears of autonomous systems, capturing crucial environmental data in the form of images or pointcloud readings. Throughout this thesis, a thorough exploration of automotive sensors is presented, focusing on both their hardware design and provided data formats, as this data constitutes the input to dedicated perception algorithms, which create a digital model of the surroundings.

Among those algorithms, Machine Learning methods in particular have recently gained significant recognition within the scope of perception systems. Especially in Object Detection problems, through the analysis of sensor data, those solutions tend to dominate in the current industrial applications. They achieve outstanding performance and offer perspectives for further improvements. This thesis research topic is cen- tred exactly around AV perception systems, which are based on ML methods, and more precisely on Con- volutional Neural Networks and Deep Learning approaches. These techniques have advanced the field of perception, enabling vehicles to sense their surroundings with remarkable accuracy. By utilizing modern ar- chitecture designs, reviewed in this research, such networks can decode intricate patterns and representations from sensor data, resulting in a high-level understanding of the environment.

Moreover, a comprehensive perception may benefit significantly from complementary information pro- vided by various sources. Each sensor has its advantages and disadvantages for perception purposes and the combination of different devices in the AV sensor suite could mitigate their undesirable traits, consequently improving the whole system. Addressing that claim, this research focuses primarily on the utilization of sen- sor data fusion and the exploration of its benefits for such systems. The concept of data fusion is discussed in detail and different methods of fusing sensor data are presented. Among those, the special type of low- level data fusion is particularly interesting, as it naturally pairs well with the Neural Networks processing approach. The main goal of this thesis is to determine whether the ML-based Low-Level Fusion solution could prove to be beneficial for an OD task in an AV perception system. The target fusion is performed on camera images and pointcloud data from either LiDAR or Radar.

The key innovation in the pursuit of that goal is a novel approach to Low-Level Fusion, called the Cross-Domain Spatial Matching method. This method offers an alternative methodology, not yet seen in this research domain. It comprises two main elements: sensor data domain alignment and fusion methods. The former component addresses the challenge associated with disparate orientations of data samples in relation to the shared host vehicle coordinate system. Once the data is aligned, it facilitates the integration of samples from various sensors without the need for additional explicit projections. It also allows for the latter fusion strategies to be applied directly to the domain alignment output. In this research, three unique fusion strategies are proposed to be further verified, each built upon a different approach, posing distinct bene- fits. Both CDSM elements are seamlessly integrated into the Neural Network architecture. This integration not only

facilitates end-to-end training but also contributes to efficient inference times during operational deployment.

In order to implement a complete OD network architecture, several single-sensor models are also cre- ated, based on State-Of-The-Art solutions from corresponding domains. These models serve a dual purpose: firstly, to evaluate the efficiency of the perception system when reliant solely on one sensor, and secondly, as integral submodules within the fusion architecture, responsible for extracting feature maps from each input sample. Through extensive experimentation and proper evaluation techniques, further exploration of fusion validity is conducted, using two open-source automotive datasets - KITTI and NuScenes. The thesis contains a detailed description of those datasets together with a training process overview. Furthermore, the results of single-sensor and fusion models are shown and compared to each other. The in-depth analysis of models' predictions is performed in terms of both visual and KPI metrics performance. For the fusion solution, the efficiency gain is highlighted and examples of corner cases are presented, offering insight into scenarios where the fusion model diverges from the predictions of single-sensor counterparts. The com- parison to SOTA fusion solutions is also provided, facilitating the positioning of CDSM among currently leading techniques. Finally, these results analysis allows for drawing conclusions regarding such a fusion methodology in AV perception systems and whether LLF could be beneficial for it.

06.10.2023 Daniel Dworak

Data i czytelny podpis kandydata