Abstract

The development of data-efficient and explainable machine learning models is important for the enhancement of the safety and reliability of autonomous driving systems. Visual perception, a core component of autonomous vehicles, relies heavily on deep neural networks to interpret complex and dynamic environments. However, traditional deep neural networks often require extensive labeled datasets to achieve high accuracy, posing challenges in scenarios where data collection is limited or expensive. Additionally, the inherent opacity of these models hinders comprehension of their decision-making processes, raising concerns about safety and trust. This dissertation addresses these challenges by proposing methods that enhance data efficiency and explainability within the context of visual perception for autonomous vehicles.

This research introduces several contributions aimed at improving the performance and interpretability of deep neural networks in autonomous driving applications. One of the primary contributions is a guided learning framework that enhances model training by using targeted feedback from visualizations of the networks attention. Moreover, novel neural network architectures was developed for the estimation of geometric features from monocular images. These architectures incorporate in its loss functions 3D object models and environmental geometric constraints to enhance the accuracy of keypoint localization to make predictions that are consistent with the physical structure of the objects being observed and providing a robust solution even in scenarios with limited training data.

Furthermore, the dissertation investigates the integration of uncertainty estimation within neural network architectures. A dedicated neural network was designed to predict the uncertainty of 2D and 3D keypoint coordinates, offering a probabilistic measure of confidence in each prediction. Additionally, the dissertation presents a processing pipeline for estimating the pose of surrounding vehicles using monocular images. This pipeline integrates the Unscented Transform algorithm to propagate uncertainties from 2D and 3D keypoint estimations, providing a measure of the overall uncertainty in vehicle pose estimations.

The proposed methods were evaluated in two real-world scenarios. The first scenario involved a docking maneuver to a charging station using an electric city bus. In this application, the system successfully estimate the pose of the bus relative to the charging station, enabling precise docking with an error margin of less than 30 cm from a distance of over 30 meters. The second scenario focuses on the pose estimation of surrounding vehicles in a city environment. The used dataset is an ApolloCar3D that provides images of an urban environment to test the pose estimation capabilities of the developed models. The proposed method achieved state-of-the-art results on the aforementioned dataset.

The methods and their experimental evaluation presented in this dissertation validates research theses: Firstly, architectures that extract and visualise meaningful intermediate features can enhance learning by augmenting existing datasets. Secondly, they accurately describe the uncertainty of the produced geometric features. Thirdly, utilising available 3D models of observed objects facilitates the learning of geometric features from 2D images without exact labeling, and significantly improves the detection accuracy of these features. Lastly, knowledge of geometric constraints from known object models helps reduce false feature detection and increases the precision of feature localization.

Streszczenie

Rozwój modeli uczenia maszynowego, które są i wydajne w kontekście danych uczących i łatwe w interpretacji, jest ważny dla zwiększenia bezpieczeństwa i niezawodności autonomicznych pojazdów. Moduły percepcji wizualnej w pojazdach autonomicznych w celu interpretacji złożonych i dynamicznych środowisk wykorzystują w dużej mierze głębokie sieci neuronowe, które często wymagają obszernych zbiorów danych, aby osiągnąć wysoką dokładność. Stanowi to wyzwanie w scenariuszach, w których gromadzenie danych jest problematyczne lub kosztowne. Dodatkowo, brak transparentności tych modeli utrudnia zrozumienie ich procesów decyzyjnych, budząc obawy o bezpieczeństwo i wiarygodność. Niniejsza rozprawa doktorska odnosi się do tych wyzwań, proponując metody, które zwiększają efektywność wykorzystania danych uczących i wyjaśnialność decyzji w kontekście percepcji wizualnej dla pojazdów autonomicznych.

Badania te wprowadzają kilka rozwiązań mających na celu poprawę wydajności i możliwości interpretacji głębokich sieci neuronowych w zastosowaniach związanych z autonomicznymi pojazdami. Jednym z głównych wkładów jest metoda sterowanego uczenia sieci, która usprawnia trening modelu poprzez wykorzystanie informacji zwrotnej z algorytmu wizualizującego uwagę sieci. Ponadto opracowano nowe architektury sieci neuronowych do szacowania cech geometrycznych z obrazów monokularnych. Architektury te uwzględniają w funkcjach kosztu modele 3D obiektów i ograniczenia geometryczne środowiska w celu zwiększenia dokładności lokalizacji punktów kluczowych, zapewniając predykcje zgodne z fizyczną strukturą obserwowanych obiektów i stabilne rozwiązanie nawet w scenariuszach z ograniczoną ilością danych treningowych.

Ponadto rozprawa bada metody szacowania niepewności predykcji poprzez sieci neuronowe. Dedykowana sieć neuronowa została zaprojektowana do estymacji niepewności współrzędnych punktów kluczowych 2D i 3D, oferując probabilistyczną miarę niepewności każdej predykcji. Dodatkowo, przedstawiony został potok przetwarzania do szacowania pozycji otaczających pojazdów przy użyciu obrazów monokularnych. Potok ten integruje algorytm Unscented Transform w celu propagacji niepewności estymowanych punktów kluczowych 2D i 3D, w celu uzyskania niepewności estymowanej pozycji pojazdu.

Proponowane metody zostały zweryfikowane na dwóch scenariuszach. Pierwszy scenariusz obejmował manewr dokowania do stacji ładowania przy użyciu elektrycznego autobusu miejskiego. W tej aplikacji system z powodzeniem szacował pozycję autobusu względem stacji ładowania, umożliwiając precyzyjne dokowanie z marginesem błędu mniejszym niż 30 cm z odległości ponad 30 metrów. Drugi scenariusz koncentruje się na estymacji pozycji otaczających pojazdów w środowisku miejskim. Dokładność szacowanej pozycji pojazdów została zweryfikowana wykorzystując zbiór danych ApolloCar3D, który zawiera obrazy pochodzące ze środowiska miejskiego. Zaproponowana metoda osiągnęła wyniki state of the art na wyżej wymienionym zbiorze danych.

Metody i ich eksperymentalna ewaluacja przedstawione w niniejszej rozprawie potwierdzają tezy badawcze: Po pierwsze, architektury, które ekstrahują i wizualizują interpretowalne cechy, mogą usprawnić trening sieci poprzez rozszerzenie istniejących zbiorów danych. Po drugie, sieci neuronowe są w stanie opisać niepewność wytworzonych cech geometrycznych. Po trzecie, wykorzystanie dostępnych modeli 3D obserwowanych obiektów ułatwia trening sieci do estymacji cech geometrycznych z obrazów 2D bez dokładnego etykietowania i znacznie poprawia dokładność wykrywania tych cech. Wreszcie, znajomość ograniczeń geometrycznych ze znanych modeli obiektów pomaga zmniejszyć liczbę fałszywych detekcji cech i zwiększa precyzję ich lokalizacji.