

Widzenie komputerowe jest ważną dziedziną nauki i techniki. Obecnie obserwuje się znaczący przyrost zastosowań technik widzenia komputerowego w wielu obszarach, a w szczególności w medycynie, wojsku, robotyce, systemach monitoringu i nadzoru, rozrywce oraz przemyśle filmowym. Systemy wizyjne będą pełniły kluczową rolę w autonomicznych pojazdach, statkach powietrznych, humanoidalnych robotach oraz systemach sztucznej rzeczywistości. W znaczącej liczbie zastosowań praktycznych, jakość i skuteczność percepcji sceny są czynnikami decydującymi o możliwościach całego systemu. Co więcej, bez znaczącego postępu w obszarze nauk związanych z uczeniem systemów rozpoznających, a w szczególności bez znaczącego rozwoju systemów związanych z podejmowaniem decyzji bez udziału człowieka w oparciu o napływające na bieżąco dane z sensorów, nie będzie możliwe wprowadzenie w szerszej skali robotów usługowych, czy też pojazdów autonomicznych. Zasadniczą przeszkodą w upowszechnieniu się wspomnianych technologii, a także wielu pokrewnych technologii, w których sensory obrazów pełnią istotną rolę, jest to, że obecnie wykorzystywane algorytmy do detekcji, klasyfikacji i rozpoznawania wymagają znaczącej liczby przykładów do zbudowania modelu, czy też nauczania sieci neuronowej. Immanentną cechą obecnie dostępnych algorytmów jest to, że wymagają one setki, a nawet tysiące razy więcej przykładowych obrazów na etapie uczenia w porównaniu do tego co potrzebuje człowiek. Co więcej, większość obecnie dostępnych algorytmów nie jest odporna na zakłócenia, w szczególności na wszelkiego rodzaju artefakty i przekłamania w danych. Innym znaczącym ograniczeniem współczesnych algorytmów jest to, że pozbawione są one zdolności do uogólniania. W konsekwencji, o wiele łatwiej jest zbudować system, który wygra z człowiekiem w warcaby, a nawet w szachy, niż zbudować algorytm, który umożliwiłby robotowi humanoidalnemu uchwycenie nieznanego mu wcześniej przedmiotu w oparciu o obrazy z kamery.

Obecnie dość powszechnie wykorzystuje się głębokie sieci neuronowe w wielu obszarach nauki. W systemach wizyjnych dość powszechnie wykorzystuje się głębokie sieci neuronowe, które łączą sieci konwolucyjne oraz klasyczne sieci z w pełni połączonymi neuronami. Warstwy konwolucyjne odpowiedzialne są za wydzielenie cech o największej sile dyskryminacyjnej, podczas gdy warstwy w pełni połączone odpowiedzialne są za klasyfikację cech wydzielonych tym sposobem. Głębokie sieci neuronowe składają się z wielu warstw, które w kolejnych warstwach grupują cechy na mapach wygenerowanych przez poprzednie warstwy w bardziej ogólne reprezentacje cech obiektów. Koszty i nakłady prac związane z przygotowaniem zbiorów poetykietowanych danych dla potrzeb uczenia głębokich sieci konwolucyjnych, a także innych sieci głębokich są znaczące. Wspomniane koszty pociągają za sobą konieczność rozwoju metod uczenia nienadzorowanego, które dość często wykorzystywane są w pierwszych fazach uczenia.

Obecnie istnieje potrzeba stworzenia nowych lub optymalizacji już istniejących algorytmów dedykowanych do budowy oraz uczenia sieci głębokich. W szczególności istnieje zapotrzebowanie na nowe i bardziej skuteczne rozwiązania dla potrzeb szeroko rozumianego widzenia komputerowego. Badania naukowe, które będą realizowane w niniejszym projekcie, dotyczą budowy modeli oraz uczenia głębokich struktur złożonych z heterogenicznych sieci neuronowych oraz bloków analizy/renderingu obrazów. Opracowane rozwiązania umożliwią bardziej efektywne wykorzystanie niepoetykietowanych danych, a w szczególności zaowocują efektywnymi mechanizmami uczenia sieci, które mimo to, że będą uczone na znacząco mniejszych zbiorach danych, będą charakteryzowały się porównywalną siłą dyskryminującą. Dzięki opracowanym rozwiązaniom powstaną algorytmy, które umożliwią rozpoznawanie nie tylko obiektów, których przykłady zostały użyte do uczenia, ale również obiekty nigdy wcześniej niewidziane, a oceniane wyłącznie na podstawie odkrytych związków i zależności o charakterze bardziej ogólnym. Wynikiem projektu będą nie tylko rozwiązania i opracowania, które dadzą odpowiedź na wiele pytań, ale także algorytmy praktycznie zweryfikowane i przebadane oraz konkretne rozwiązania do widzenia komputerowego w czasie rzeczywistym.