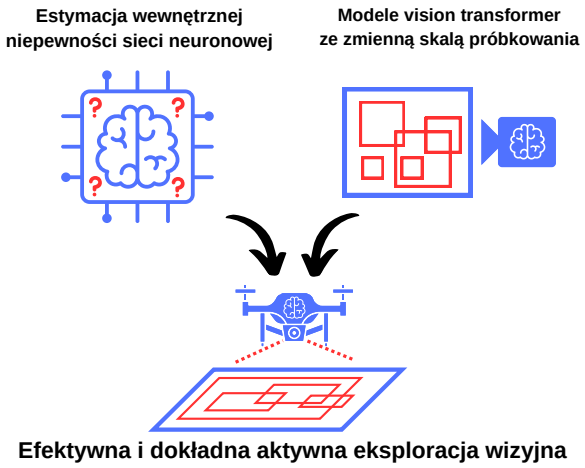
Where to look next - aktywna eksploracja sterowana wewnętrzną niepewnością modelu

Widzenie komputerowe. W ostatnich latach głębokie uczenie zrewolucjonizowało dziedzinę widzenia komputerowego. Obecne modele głębokich sieci neuronowych oparte na architekturze vision transformer (ViT) nie tylko osiągają bezprecedensowo wysoką dokładność, ale także uogólniają się na wiele dziedzin bez potrzeby dodatkowego treningu. Jednak wraz ze wzrostem rozmiaru i złożoności sztucznych sieci neuronowych, wzrosły również ich wymagania obliczeniowe i zużycie energii. Co więcej, tradycyjne rozwiązania wizji komputerowej zwykle zakładają pełny dostęp do danych wejściowych, pomijając wyzwania stawiane przez rzeczywiste aplikacje. Aktywna eksploracja wizyjna (Active visual exploration; AVE) ma być remedium na te problemy.



Rysunek 1: Projekt ten łączy wewnętrzną estymację niepewności modelu z architekturą sieci neuronowej typu vision transformer, tworząc efektywne i dokładne metody Aktywnej Eksploracji wizyjnej. W tym celu, wprowadzamy strategie eksploracji bazujące na niepewności modelu oraz adaptujemy vision transformer do obsługi wejścia w różnych rozdzielczościach.

Aktyna eksploracja wizyjna. AVE pozwala maszynom aktywnie eksplorować swoje otoczenie, podobnie jak robią to ludzie. Zamiast przetwarzać całe pole widzenia w maksymalnej rozdzielczości, AVE umożliwia agentowi selektywne próbkowanie i koncentrowanie zasobów obliczeniowych na kluczowych obszarach. Dzięki inteligentnemu określaniu gdzie i w jakiej rozdzielczości zbierać informacje wizualne, AVE pozwala na bardziej wydajne i dokładne widzenie komputerowe. Obecne rozwiązania dla AVE mają jednak poważne ograniczenia. Po pierwsze, opierają się one na złożonym treningu, co prowadzi do algorytmów wymagających dużej ilości obliczeń i pamięci. Po drugie, używają próbki obrazu o stałym rozmiarze, nie wykorzystując najbardziej podstawowych możliwości platform robotycznych, takich jak zoom optyczny lub swobodnie obracające się kamery. Niniejszy projekt ma na celu zaradzenie tym ograniczeniom poprzez stworzenie nowej gałęzi w badaniach.

Wykorzystanie niepewności sztucznej sieci neuronowej. Pierwszy aspekt tego projektu koncentruje się na wykorzystaniu możliwości modeli typu vision transformer w zakresie szacowania wewnętrznej niepewności. Włączając tę niepewność do procesu AVE, agenci mogą podejmować świadome decyzje dotyczące tego, gdzie eksplorować, aby zmaksymalizować świadomość sytuacji i aktywnie poszukiwać regionów, w których prognozy modelu są niejednoznaczne, co prowadzi do dokładniejszego zrozumienia środowiska. To nowatorskie podejście pozwala uniknąć konieczności stosowania dodatkowych modułów neuronowych i zmniejsza złożoność modelu.

Wykorzystanie skali i elastyczności. Drugim kluczowym celem jest sprawienie, by vision transformer wykorzystywały informację o skali próbkowania i były dostosowane do zadań AVE. Obecnie ViT wykorzystują próbkowanie o stałej siatce, co utrudnia im uchwycenie kluczowych szczegółów w różnych sytuacjach. Planujemy zmodyfikować warstwy wejściowe ViT, aby akceptowały dowolnie próbkowane dane, pozwalając przetwarzać zdjęcia o różnym poziomie powiększenia.

Optymalizacja selekcji obserwacji. Ostatni etap tego projektu ma na celu udoskonalenie procesu selekcji obserwacji. Poprzez integrację metod opartych na wykorzystaniu niepewności sieci neuronowej z modelami vision transformer wykorzystującymi informację o skali, proces selekcji określi, które obszary wymagają wielu fragmentów o wysokiej rozdzielczości, a które można pokryć za pomocą kilku próbek o niskiej rozdzielczości. To zoptymalizowane podejście znacznie zmniejszy liczbę potrzebnych obserwacji do wykonania zadania przy zachowaniu wysokiej dokładności.

Wnioski i perspektywy na przyszłość. Sukces tego projektu może mieć ogromny wpływ na wiele dziedzin, w tym robotykę, pojazdy autonomiczne i analizę obrazu. Ucieleśnieni agenci, jak roboty lub bezzałogowe statki powietrzne, skorzystają z bardziej wydajnej i dokładnej eksploracji wizualnej, pozwalającej im na efektywną nawigację w złożonych i dynamicznych środowiskach. Co więcej, postępy poczynione w architekturze ViT i strategiach aktywnej eksploracji mogą prowadzić do bardziej zrównoważonych pod względem wpływu na środowisko naturalne systemów sztucznej inteligencji, zmniejszając wymagania obliczeniowe i zużycie energii.