

Generowanie wideo przy użyciu głębokiego uczenia

Głębokie uczenie to kategoria uczenia maszynowego, która automatyzuje ekstrakcję cech przy użyciu dużej ilości danych do trenowania złożonych głębokich sieci neuronowych. W ostatnich latach głębokie uczenie stało się jednym z najbardziej popularnych obszarów informatyki i dzięki swojej skuteczności pomogło osiągnąć przełomowe wyniki w wielu różnych obszarach badawczych, m.in. w wizji komputerowej, analizie obrazów satelitarnych i medycznych, analizie dźwięku, rozpoznawaniu mowy, przetwarzaniu języka naturalnego, a także podejmowaniu decyzji przez systemy autonomiczne.

Generative Adversarial Networks (GANs) to modele uczenia głębokiego zdolne do generowania realistycznych, ale syntetycznych danych. GANs cieszą się coraz większym zainteresowaniem ze względu na ich zdolność do modelowania złożonych dystrybucji danych, co pozwala im generować realistyczne obrazy, a także tłumaczyć obrazy pomiędzy domenami. Podczas gdy realistyczne generowanie wideo jest naturalnym rozszerzeniem generowania obrazów, jest znacznie trudniejsze pod względem złożoności i obliczeń, co jest związane z jednoczesnym modelowaniem wyglądu oraz ruchu.

W szczególności, wnioskując i modelując dystrybucję wideo, która przedstawia ludzi, modele generatywne stoją przed trzema głównymi wyzwaniami: (a) generowanie niepewnego ruchu, (b) zachowanie ludzkiego wyglądu w całym generowanym wideo, a także (c) modelowanie spójności przestrzenno-czasowej. Znalezienie odpowiednich metod uczenia się reprezentacji, które są w stanie sprostać tym wyzwaniom, ma kluczowe znaczenie dla ostatecznej jakości wizualnej i wiarygodności renderowanych nowych sekwencji wideo.

Zaproponowany tutaj projekt koncentruje się na generowaniu wideo i obejmuje 3 główne tematy badawcze. Po pierwsze, skupimy się na skomplikowanym problemie generowania wideo bez wcześniejszej wiedzy pod względem wyglądu i ruchu, a jednocześnie ucząc się rozłącznej reprezentacji wyglądu i ruchu, pozwalając na manipulację obydwoma. Reprezentacja rozłączna została tutaj zdefiniowana jako taka, w której pojedyncze jednostki przestrzeni ukrytej są wrażliwe na zmiany w pojedynczych czynnikach generujących, będąc jednocześnie względnie niezmiennymi na zmiany innych czynników. Następnie zamierzamy zbudować podejście do warunkowego generowania wideo, aby umożliwić manipulowanie wygenerowaną treścią wideo. Na koniec skupimy się na probabilistycznym generowaniu wideo, co jest szczególnie ważne w prognozowaniu następujących klatek wideo.