

Prace zrealizowane w semestrze 22L

Rozpoznawanie dyscypliny sportu na podstawie wideo.
Rafał Lewanczyk

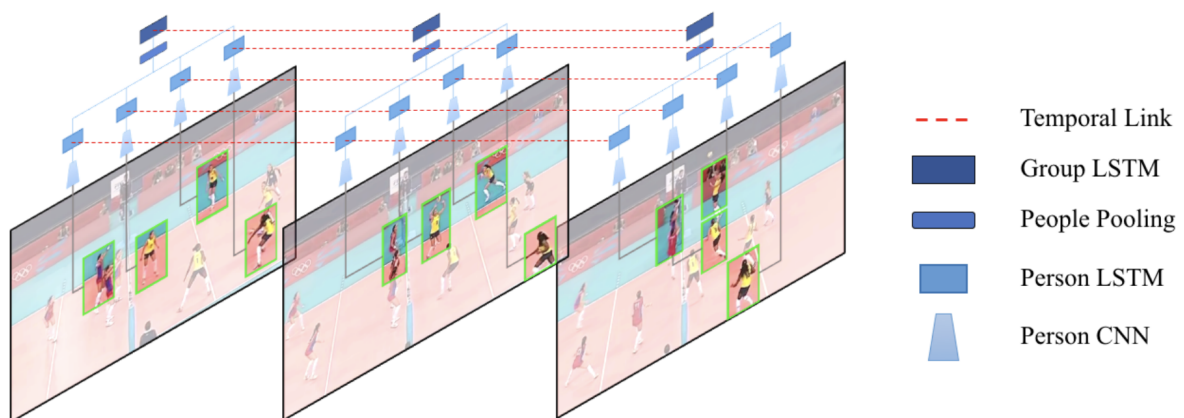
Wzorcowe podejście

W pracy decyduje się na modyfikację podejścia zaprezentowanego w pracy [A Hierarchical Deep Temporal Model for Group Activity Recognition](#)

Podejście to charakteryzuje się zastosowaniem dwóch warstw LSTM:

Pierwsza warstwa składa się z kilku LSTM odpowiada za śledzenie akcji pojedynczej osoby. Na jej wejściu znajdują się cechy wydobyte przez sieć CNN z bounding box'a badanej osoby w danym momencie.

Druga warstwa przyjmuje na wejściu połączone wyjścia z pierwszej warstwy. Swoje wyjście kieruje do warstwy klasyfikującej soft-max, wykonującej predykcję aktywności grupowej.



Model ten był trenowany w 2 krokach. Jako pierwsza wytrenowana została sieć LSTM odpowiedzialna za akcje pojedynczej osoby. Druga warstwa została wytrenowana przy sprawnie działającej pierwszej warstwie.

Model osiągnął skuteczność 81.5% na zbiorze z ogólnymi aktywnościami [Collective Activity Dataset](#).

Realizowane podejście

W pracy magisterskiej, również planuje zbudować 2-warstwowy model, składający się z sieci LSTM. Warstwy pełnią takie same funkcje co w podejściu wzorcowym, jednak będą one przybierały inne parametry na swoim wejściu:

Pierwsza warstwa LSTM na wejściu przyjmuje punkty kluczowe wykrytej osoby, odpowiadające w jakiej pozycji się znajduje.

Druga warstwa LSTM przyjmuje na wejściu połączone wyjścia z pierwszej warstwy. Przy eksperymentach nad tą warstwą, możliwe, że jej wejście zostanie rozszerzone o dodatkowe parametry np. cechy charakterystyczne obrazu danej klatki.

Zrealizowane prace

W semestrze 22L zrealizowane zostały:

1. Przegląd dostępnych rozwiązań klasyfikacji wideo
2. Przegląd zbiorów danych zawierających nagrania aktywności sportowych.

Ostatecznie wybrany został zbiór [Sports Videos in the Wild \(SVW\)](#). Zbiór składa się z 4200 nagrań podzielonych na 30 dyscyplin sportowych. Nagrania pokazują sporty uprawiane przez profesjonalistów jak i amatorów.

3. Przegląd dostępnych rozwiązań wykrywania sylwetki oraz indeksowania sylwetek

Ostatecznie wybrany został model [AlphaPose](#), pozwalający na wykrywanie oraz śledzenie sylwetek w czasie rzeczywistym.

Działanie modelu AlphaPose na przykładowym wideo pochodzącym ze zbioru SVW:

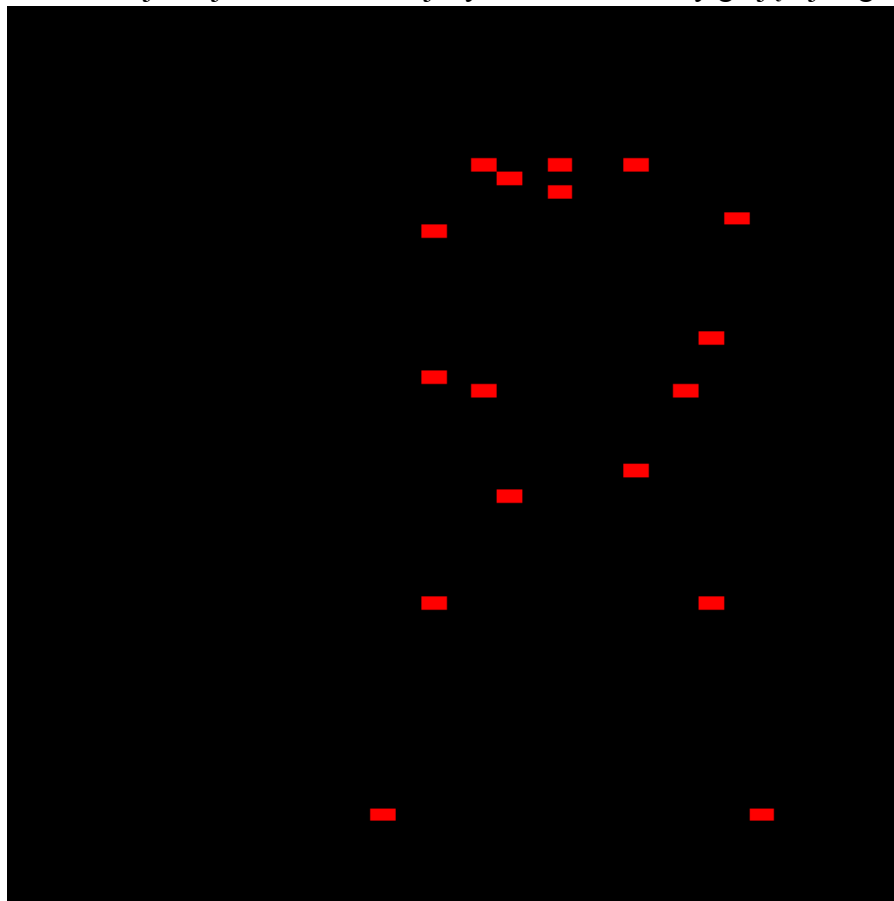


4. Zrealizowany został pierwszy wstępny eksperyment treningu pierwszej warstwy LSTM. Eksperyment składał się z następujących kroków:
- a. Ze zbioru SVW wybrane zostały 3 klasy sportów indywidualnych (łuczniectwo, tenis, golf) oraz po 40 nagrań z każdej z tych klas.
 - b. Z każdego z nagrań wyeksportowane zostały pozy sportowca
 - c. Opracowana została metoda wstępnego przetworzenia danych zawierających pozę uczestnika na formę pozwalającą użycie ich na wejściu modelu LSTM.

Przetworzenie składa się z następujących kroków:

- i. Przesunięcie pozy do punktu (0, 0), w celu usunięcia poruszania się obiektu lub kamery
- ii. Ograniczenie otoczenia pozy do jej największego bounding box'a
- iii. Normalizacja punktów kluczowych pozy w ramach wybranego bounding box'a

Wizualizacja wejścia modelu. Pojedyncza klatka osoby grającej w golfa:



- d. Opracowany został podstawowy model LSTM klasyfikujący indywidualną aktywność sportową. Model osiągnął skuteczność klasyfikacji 94%.

Kod eksperymentu został umieszczony w [repozytorium](#)

Dalsze prace

Na dalsze prace składają się:

1. Opracowanie filtra usuwającego sylwetki nie wnoszących informacji np. wykrytą widownię
2. Rozszerzenie modelu z pierwszego eksperymentu na nagrania, w których bierze udział więcej niż jedna osoba
3. Opracowanie drugiej warstwy LSTM klasyfikującą aktywność grupową.