

Projekt: Wykorzystanie CLIP do rozpoznawania obiektów na zdjęciach satelitarnych

Szymon Laszczyński

February 4, 2025

Na czym polega CLIP?

- ▶ CLIP (Contrastive Language-Image Pretraining): Model stworzony przez OpenAI do łączenia obrazów i tekstów w jednej przestrzeni zanurzeń.
- ▶ Kluczowe cechy:
 - ▶ Trening na dużych, różnorodnych zbiorach danych (obraz+tekst).
 - ▶ Uniwersalność zastosowań: od analizy obrazów po wyszukiwanie informacji.

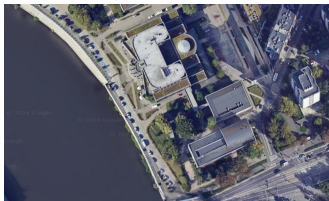
CLIP: Przykład 1



- ▶ a photo of a cat: 0.99
- ▶ a photo of a dog: 0.01
- ▶ —
- ▶ a photo of a cat: 0.91
- ▶ a photo of a tv remote: 0.09

- ▶ a photo of a cat: 0.01
- ▶ a photo of two tv remotes: 0.99
- ▶ —
- ▶ a photo of a cat: 0.00
- ▶ a photo of two tv remotes: 0.08
- ▶ a photo of two cats and two tv remotes: 0.92

CLIP: Przykład 2



- ▶ 0.05 : a photo of a car
- ▶ 0.40 : a photo of a river
- ▶ 0.56 : a photo of a city

- ▶ 0.00 : a photo of a car
- ▶ 0.04 : a photo of a river
- ▶ 0.06 : a photo of a city
- ▶ 0.89 : a photo containing a part of a city with a river, buildings, roads and some cars in the parking lots

CLIP: Przykład 2 : uwaga!

- ▶ 0.00 : a photo of a car
- ▶ 0.01 : a photo of a river
- ▶ 0.01 : a photo of a city
- ▶ 0.11 : a photo containing a part of a city with a river, buildings, roads and some cars in the parking lots
- ▶ 0.88 : a satellite image of a seacoast with industrial buildings and ship containers

- ▶ Cel: Wykorzystanie modelu CLIP do rozpoznawania obiektów na zdjęciach satelitarnych.
- ▶ Rozszerzenie: Analiza fragmentów muzyki w celu rozpoznawania gatunku muzycznego.

Ogólna charakterystyka

- ▶ Przeliczenie zanurzeń dla polskiego LLM i encodera dla obrazów.
- ▶ Dotrenowanie dla specjalistycznych danych (zdjęcia satelitarne).
- ▶ Eksperymenty z użyciem modelu:
 - ▶ Rozpoznawanie obiektów na zdjęciach satelitarnych.
 - ▶ Oznaczanie obszarów z obiektami.
 - ▶ Zliczanie obiektów.
- ▶ Badanie możliwości przeniesienia metod na inne dziedziny (np. muzykę).

Dodatkowe narzędzia i dane

- ▶ Narzędzia:
 - ▶ huggingface
 - ▶ OpenAI CLIP: github.com/openai/CLIP
 - ▶ Narzędzia do przetwarzania danych geograficznych (np. GDAL, QGIS).
- ▶ Dane:
 - ▶ EuroSAT github.com/phelber/eurosat
 - ▶ Sentinel-2 bigearth.net
 - ▶ github.com/Yuan-ManX/ai-audio-datasets

Definicje sukcesu projektu

- ▶ Pełen sukces:
 - ▶ Opracowanie rozszerzenia do analizy muzyki, które działa w praktyce.
 - ▶ Model poprawnie generuje anotacje do zdjęć satelitarnych.
- ▶ Sukces częściowy:
 - ▶ Model jest w stanie rozpoznawać obiekty na zdjęciach satelitarnych.
 - ▶ Identyfikacja ograniczeń CLIP w specyficznych zadaniach.