



„Rozpoznawanie Gestów Polskiego Języka Migowego z Wykorzystaniem Modeli Uczenia Maszynowego”

Adam Ledwoń, Błażej Lambui
Dr hab. inż. Aleksandra Świetlicka prof. PP
Politechnika Poznańska

Cel

Opracowanie inteligentnego systemu rozpoznającego gesty Polskiego Języka Migowego i przekształcającego je na tekst z wykorzystaniem modeli uczenia maszynowego.

Motywacja

- 1.** Zwiększenie niezależności życiowej osób Głuchych oraz poprawa jakości ich integracji ze społeczeństwem słyszącym poprzez ułatwienie samodzielnego funkcjonowania w codziennych sytuacjach (np. w urzędzie, u lekarza czy w sklepie).
- 2.** Zniwelowanie problemu braku wystarczającej liczby profesjonalnych tłumaczy języka migowego w instytucjach publicznych i komercyjnych.

Projekt powstał z inicjatywy Fundacji Świat Głuchych.

Wybrane Słowa

- Chcieć
- Dostępność
- Ja
- Dowód
- Ty
- Osoba
- Przepraszać
- Czy
- Prosić
- Kiedy
- Powtórzyć
- Praca
- Urząd
- Pójść
- Dziękować
- Pisać
- Czekać
- Jak
- Budynek
- Imię
- I
- Dokument
- Gdzie
- Informacja

Wybór tych 24 słów tworzy tzw. **słownik funkcjonalny**, celowo zoptymalizowany pod obsługę w urzędach, przychodniach i punktach informacyjnych. Zestaw łączy obiekty (*dowód, dokument*), akcje (*chcieć, pisać*), pytania (*gdzie, kiedy*) i zwroty grzecznościowe, co pozwala na składanie praktycznych komunikatów. Dzięki takiemu ograniczeniu domeny rozwiązujemy realny problem społeczny, a jednocześnie zapewniamy wysoką dokładność i szybkość działania modelu uczenia maszynowego na starcie.

- Ze względu na specyfikę projektu dane nie pochodziły z gotowych baz, lecz zostały **pozyskane i opracowane od podstaw** przy użyciu dostępnego sprzętu nagrywającego.
- Zarejestrowano **19 materiałów wideo** o łącznym czasie trwania wynoszącym **prawie 3 godziny**, co zapewniło odpowiednią ilość materiału do dalszego przetwarzania.
- W nagraniach wzięły udział **4 osoby**, co pozwoliło na uwzględnienie indywidualnych różnic między uczestnikami takich jak dynamika ruchów czy wygląd i poprawiło zdolność generalizacji modelu.
- Zgromadzony materiał umożliwił wyodrębnienie **około 100 osobnych próbek dla każdego z analizowanych słów**, zapewniając odpowiednią równowagę zbioru na potrzeby efektywnego uczenia głębokiego.

Wybór Detektora

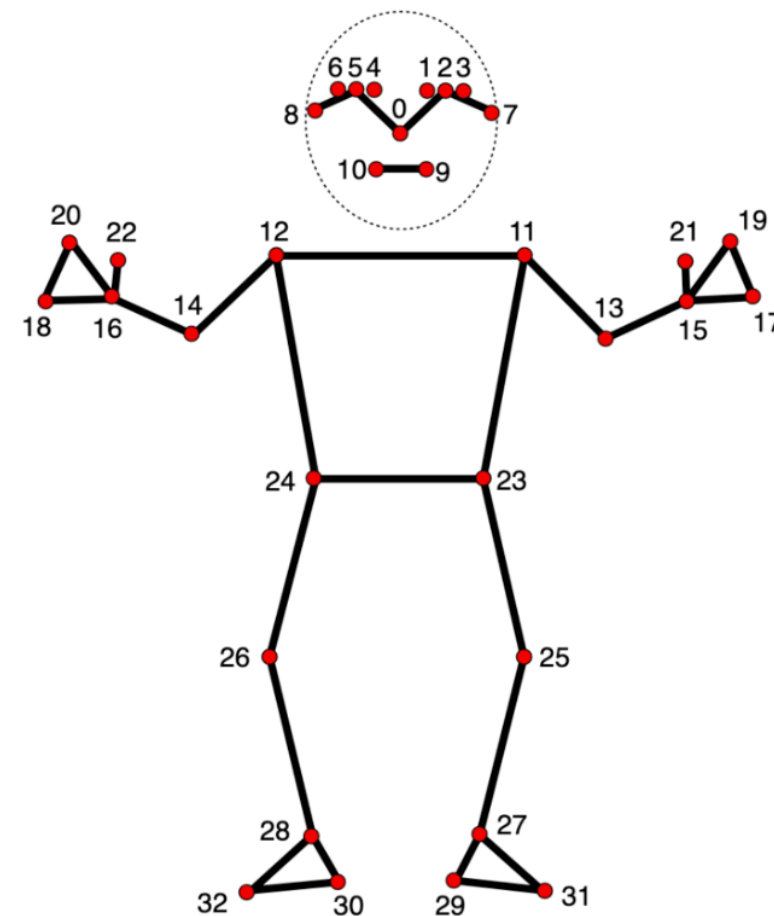
Po analizie najpopularniejszych dostępnych modeli (YOLO11, ViTPose, DETRPose, MediaPipe) wybrano MediaPipe Tasks. Decydujące czynniki:

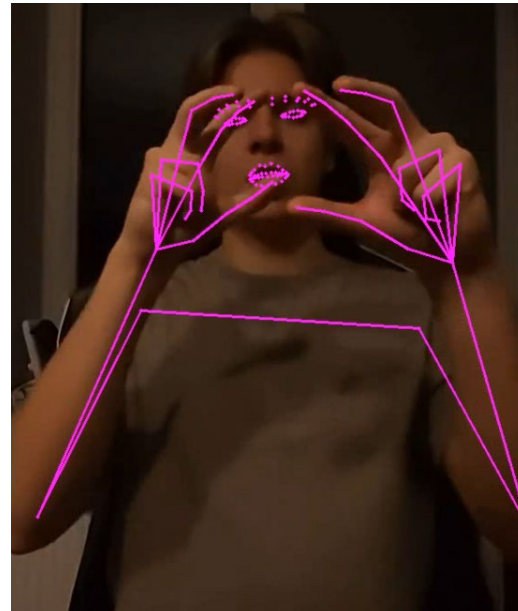
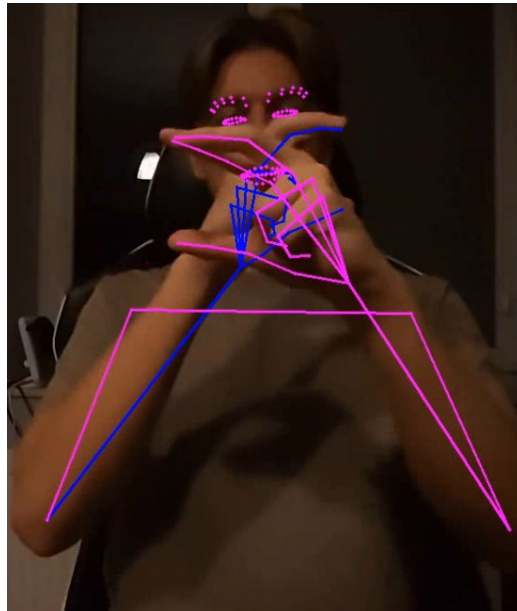
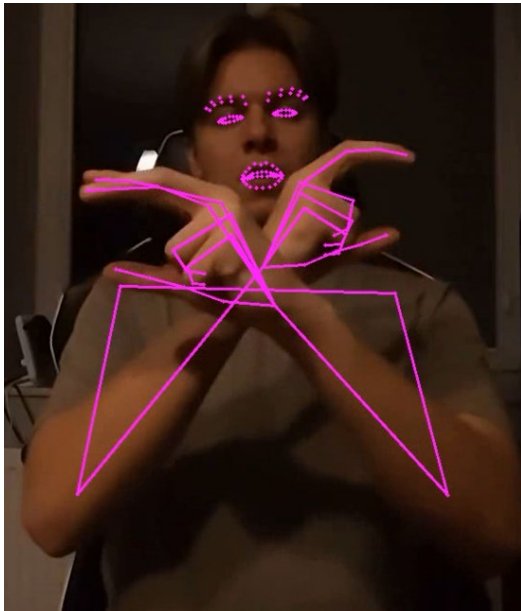
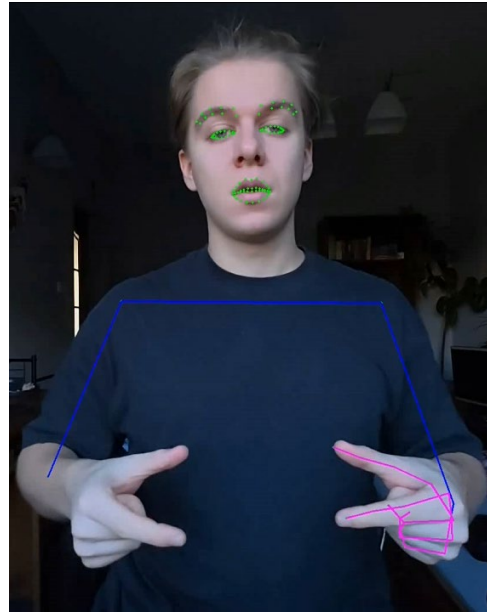
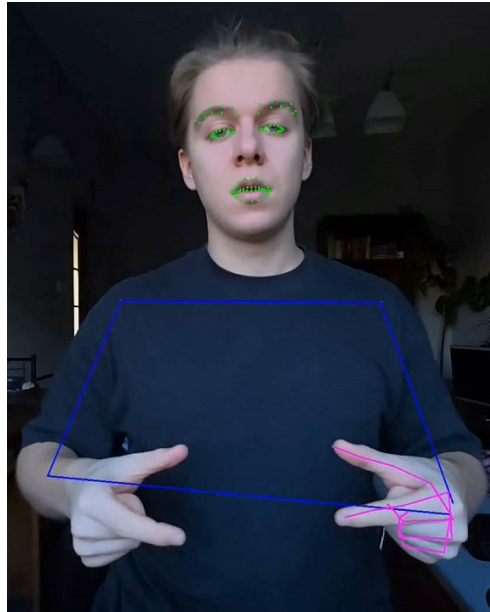
- dedykowana architektura pod urządzenia mobilne,
- gotowe moduły umożliwiające jednoczesną i precyzyjną detekcję rąk, twarzy oraz pozy.

Wykrywanie Punktów Kluczowych

Wybrany model oferuje kompleksową grupę rozwiązań, w skład której wchodzi następujące modele do analizy ciała ludzkiego:

- Face Landmarker — służy do oznaczania 478 punktów kluczowych twarzy,
- Pose Landmarker — wykrywa 33 punkty całego ciała,
- Hand Landmarker — oznacza 21 punktów dłoni.

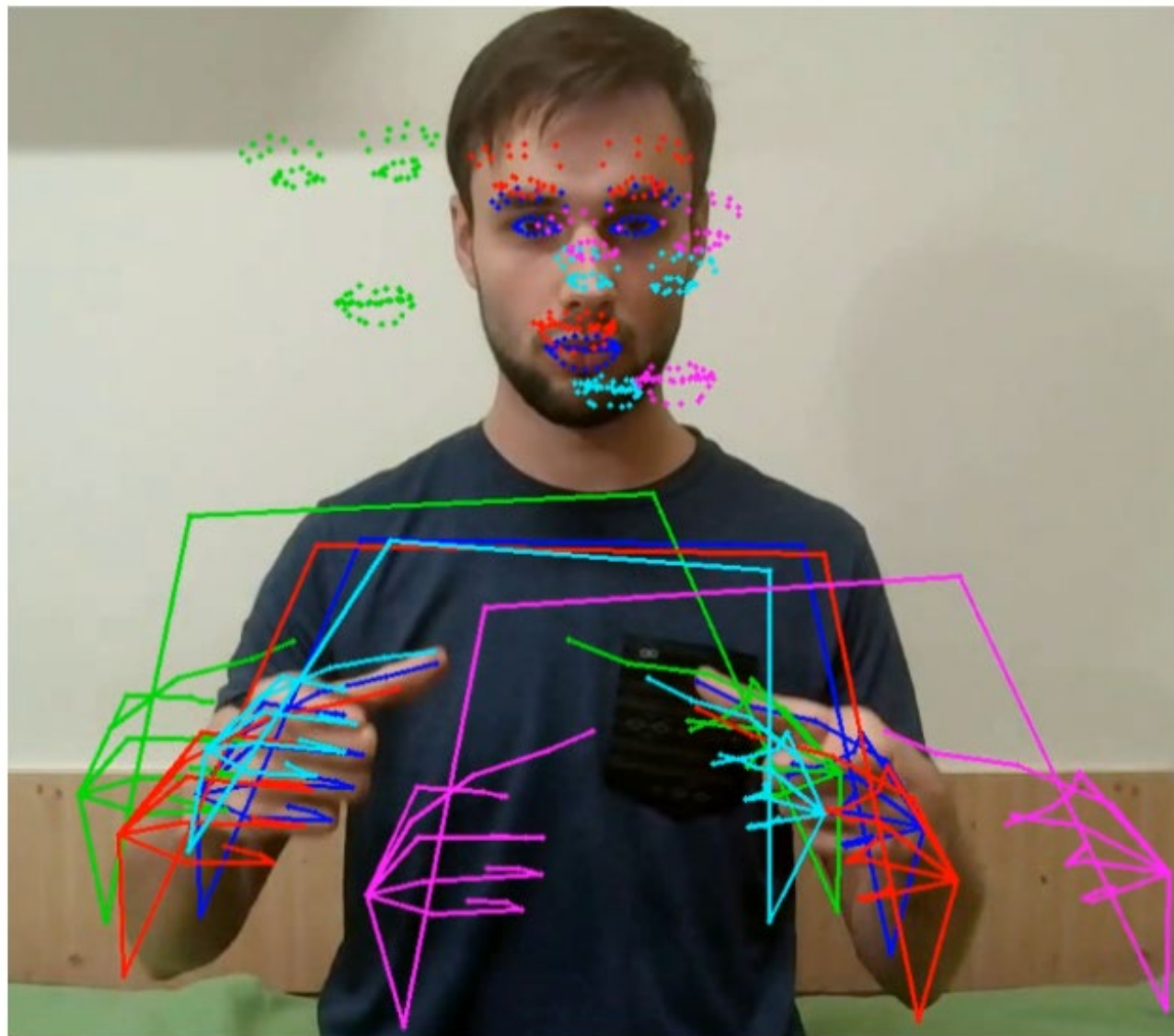


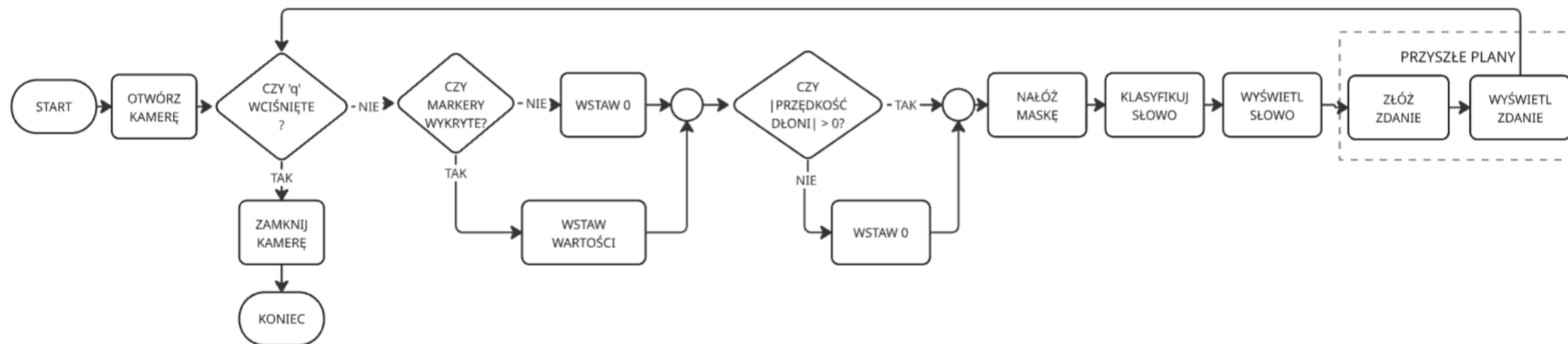


Zastosowano:

- Obrót (w zakresie ± 10 stopni)
- Przesunięcie (o ± 0.1 dla osi x i ± 0.05 dla osi y)
- Skalowanie (w zakresie 0.85 do 1.15 względem środkowego punktu linii barków)
- Szum (± 0.0025)

Po zastosowaniu augmentacji zbiór danych został powiększony pięciokrotnie do około **12 000 próbek**.





Prezentacja działania aplikacji

<https://www.youtube.com/@reaktyw1061>

Dziękujemy za uwagę