

Aplikacja analizująca emocje streamer'ów na platformie Twitch

Raport z projektu na kurs "Informatyka afektywna"

Michał Ilski, Jan Pawłowski, Patryk Rygiel

Repozytorium: <https://github.com/PatRyg99/informatyka-afektywna-L1>

1. Iteracje projektu

1. Etap 1 (26.10.2022)

- Zastosowanie mediapipe do ekstrakcji 468 landmarków twarzy
- Klasyfikacja emocji na podstawie landmarków przy użyciu reprezentacji w formie chmur punktów (modele DGCNN)
- Demo aplikacji rozpoznającej emocje z kamery

1. Etap 2 (21.12.2022)

- Analiza dodatkowych zbiorów danych
- Modele do uczenia na siatkach 3D (mesh) - FeaSt, SAGE
- Reprezentacje cech siatek 3D - HKS (Heat Kernel Signatures), XYZ
- Augmentacje danych - interpolacja emocji granicznych, generacja twarzy o nowych emocjach pomiędzy istniejącymi

1. Etap 3 (18.01.2022)

- Budowa aplikacji do analizy emocji streamer'ów na platformie Twitch
- Analiza zmian wykrywanych emocji w trakcie trwania stream'u
- Analiza porównawcza rozkładu emocji dla różnych gier

2. Zbiory danych

W ramach zbiorów danych rozważaliśmy trzy poniższe zbiory:

- CK+ (Extended Cohn-Kanade dataset) - 7 emocji + neutral
- AffectNet-HQ - 7 emocji + neutral
- AFEW-VA - valence-arousal (regresja)

2.1 CK+

Zbiór zaproponowany do użycia w ramach listy 1. Zbiór składa się z 593 sekwencji video dla 123 różnych osób. Ze wszystkich sekwencji 327 jest oetykietowanych jedną z 7 emocji: anger, contempt, disgust, fear, happy, sadness, surprise. Jedna sekwencja

przedstawia przejście z emocji neutralnej do zadanej emocji. Poniższa wizualizacja pokazuje sekwencję 15 obrazów przejścia z emocji neutral do happy:



Zbiór jest dosyć tendencyjny ze względu na setting zdjęć: wszystkie są czarno białe, wycentrowane na twarzy oraz twarze są tej samej wielkości. Emocje są za to dobrze otykietowane.

2.2 AffectNet-HQ

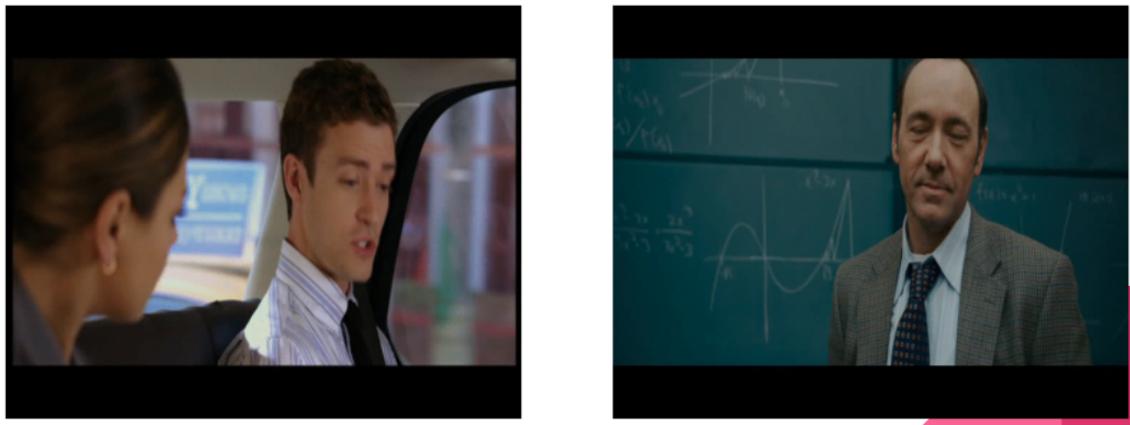
Zbiór AffectNet-HQ był częściowo olabelowany ręcznie, a częściowo automatycznie. Z tego powodu jakość zbioru jest suboptimalna. Przykładem są pokazane zdjęcia.



Mimo tego, że zdjęć jest zdecydowanie więcej niż w CK+, bo aż 31 tysięcy, jakość annotacji jest na tak słabym poziomie, że użycie zbioru do poprawnej klasyfikacji emocji jest dużo cięższym zadaniem niż ma to miejsce dla zbioru CK+.

2.3 AFEW-VA

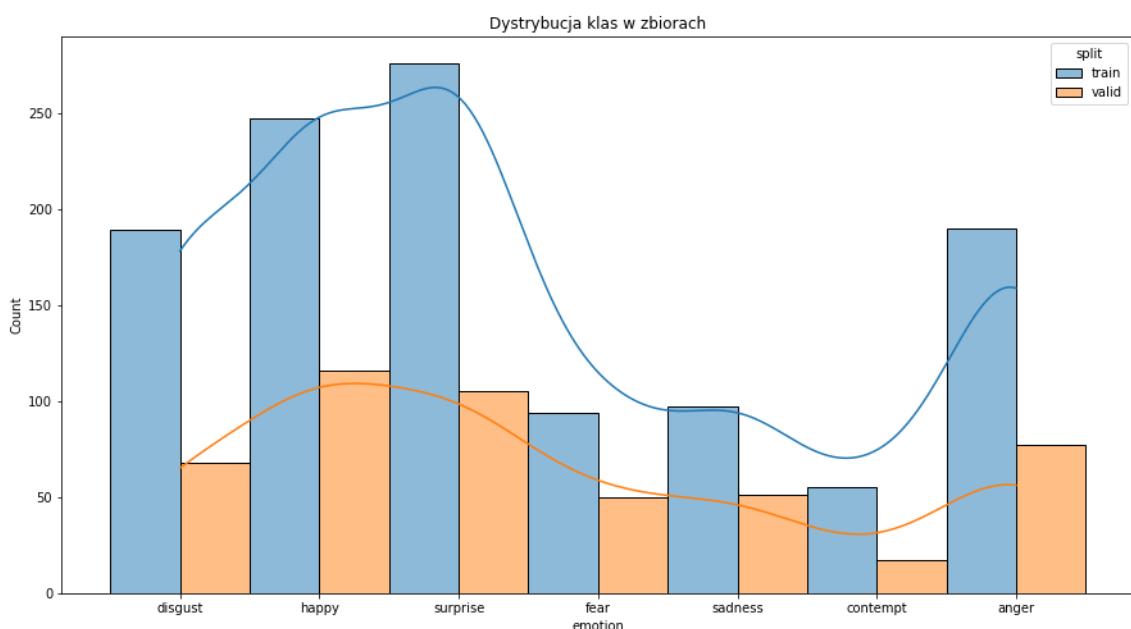
Zbiór AFEW-VA składa się z 330 filmów zaanotowanych przy użyciu valence i arousal. Zbiór jest na tyle problematyczny, że różnica w emocjach na przestrzeni filmów jest dosyć znikoma, oraz emocje są o dosyć małym nasyceniu.



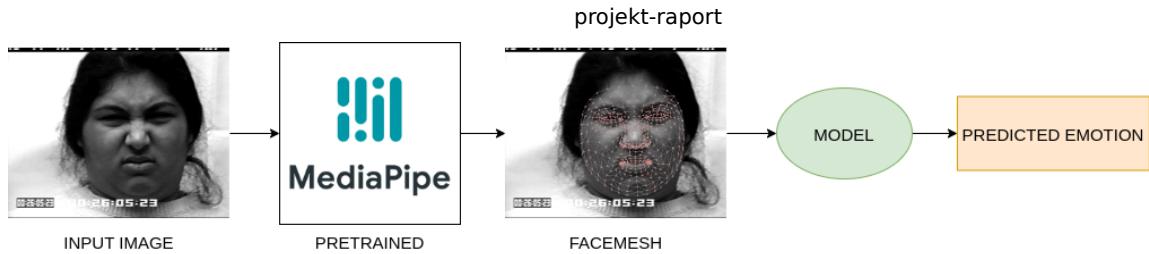
2.4 Wybrany zbiór i podział danych

Ze względu na wyżej wymienione problemy ze zbiorami AFEW-VA oraz AffectNet-HQ, zdecydowaliśmy działać tylko na zbiorze CK+.

Jako obrazy przedstawiające emocje wybrane zostało ostatnie 20% klatek z sekwencji jako, że na nich intensywność emocji jest największa i dostajemy parę różnych przykładów emocji dla osoby. Jako, że dla każdej osoby wybierana jest więcej niż jedna klatka z sekwencji oraz dla jednej osoby istnieje z reguły więcej niż jedna sekwencja (rodzaj emocji), zbiór danych został podzielony na poziomie osób, aby uniknąć przelewu danych treningowych do zbioru testowego. Zbiór został podzielony z uwzględnieniem stratyfikacji emocji (na ile to było możliwe) na zbiór treningowy (85 osób - 1148 zdjęć) oraz testowy (38 osób - 484 zdjęć). Poniższy wykres obrazuje rozkład klas w obu zbiorach:



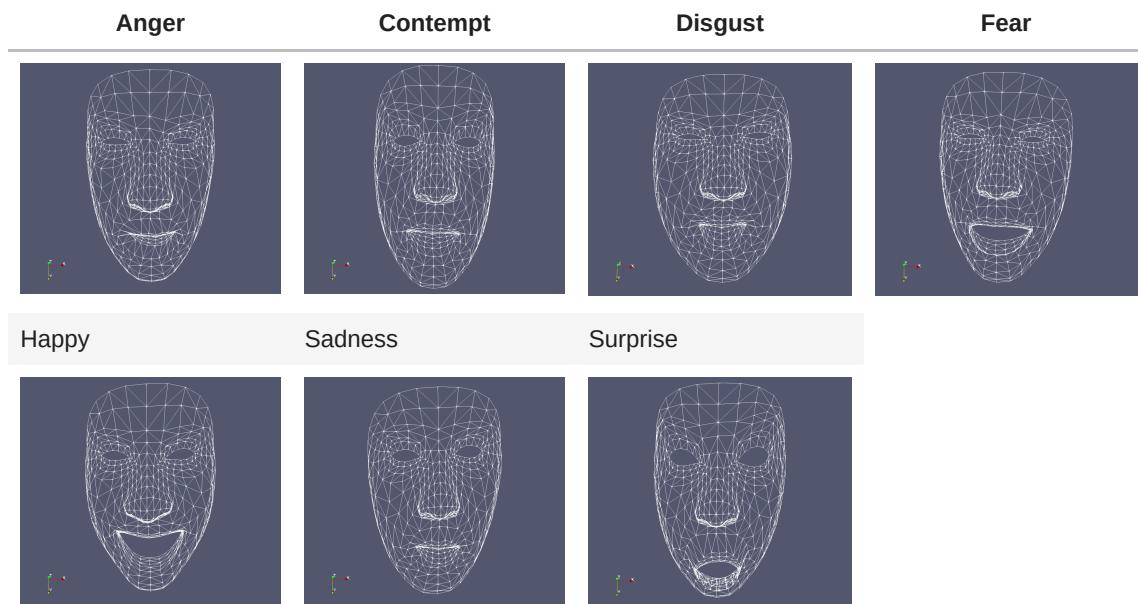
3. Metodologia



Nasza metoda oparta jest ekstrakcją ze zdjęć tzw. FaceMesh przy użyciu pre-trenowanego narzędzia MediaPipe. FaceMesh to reprezentacja twarzy w formie siatki 3D składającej się z 468 punktów charakterystycznych. Tak uzyskane siatki są używane jako zbiór do trenowania i ewaluacji modeli grafowych, których używamy w tym projekcie.

Takie podejście jest dobrą generalizacją, gdy mamy mało danych uczących, które są tendencyjne (np. czarno białe, twarz zawsze na środku zdjęcia - problemy zbioru CK+). Model nie overfittuje się do tekstur na zdjęciu, jedynie na czym działa to kształt twarzy.

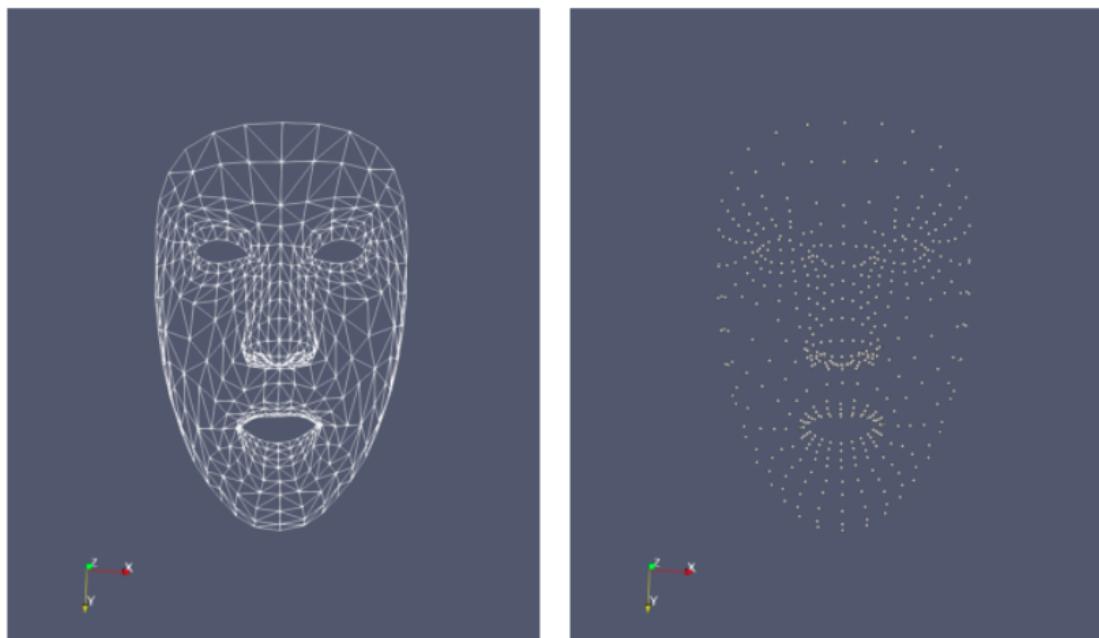
Poniżej przedstawione są przykładowe siatki dla klas emocji na zbiorze CK+:



4. Modelowanie

4.1 Architektury grafowe

Do uczenia na wygenerowanych siatkach 3D, przetestowaliśmy modele na reprezentacji danych wejściowych w postaci chmur punktów (tylko wierzchołki, brak krawędzi) oraz w postaci grafów (wierzchołki i krawędzie) - różnica pomiędzy tymi reprezentacjami jest pokazana na poniższej figurze.



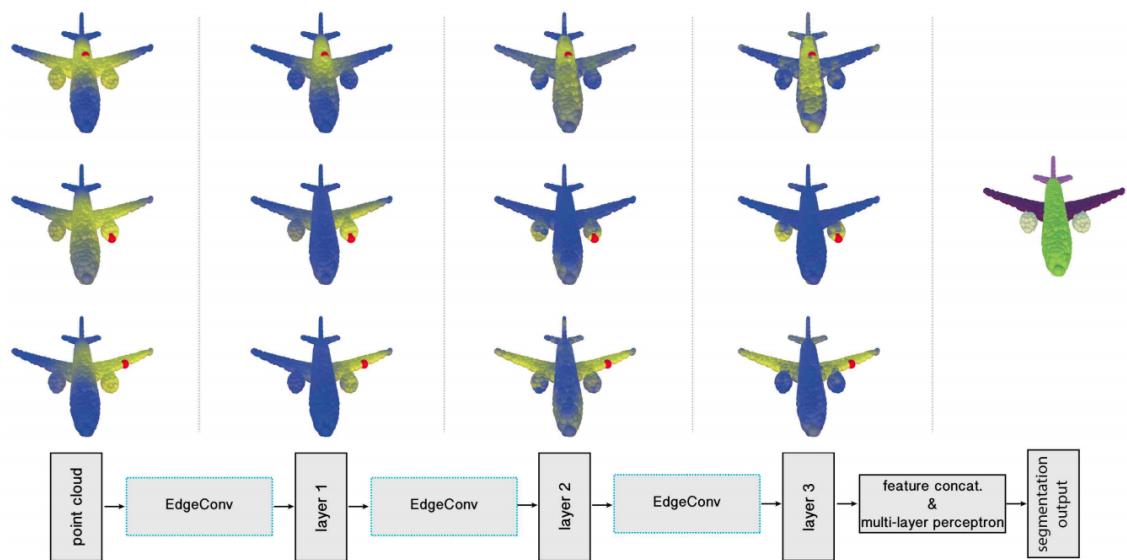
mesh

pointcloud

Przetestowaliśmy następujące architektury modeli grafowych:

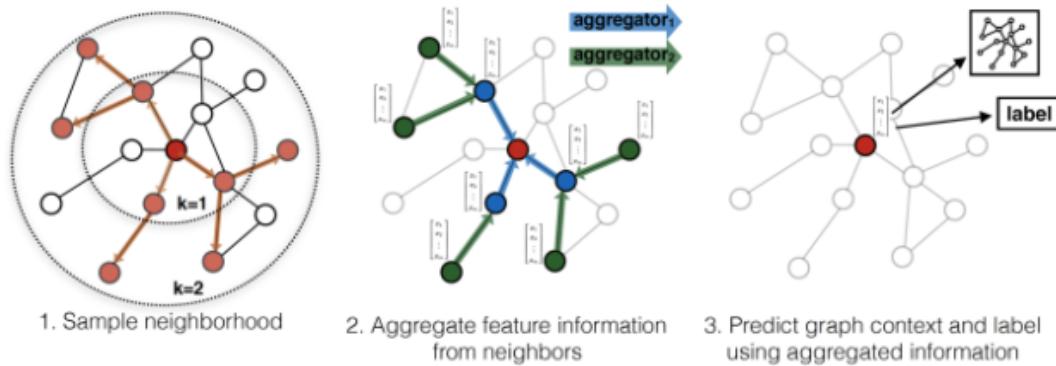
a) DGCNN

Model do przetwarzania chmur punktów oparty na dynamicznym grafie budowanym po odległościach wektorów reprezentacji punktów.



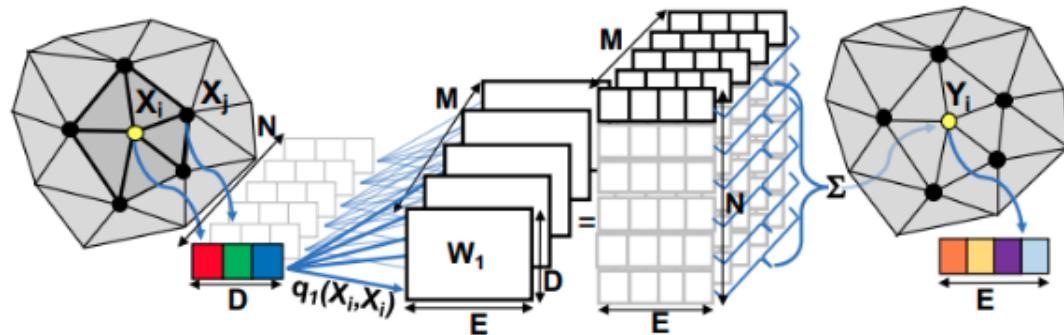
b) GraphSage

Model do przetwarzania grafów bez ukierunkowania na stricte siatki 3D. Agregacje sąsiedztwa jest samplowana z otoczenia - nie muszą być to najbliżsi sąsiedzi.



c) FeaSt

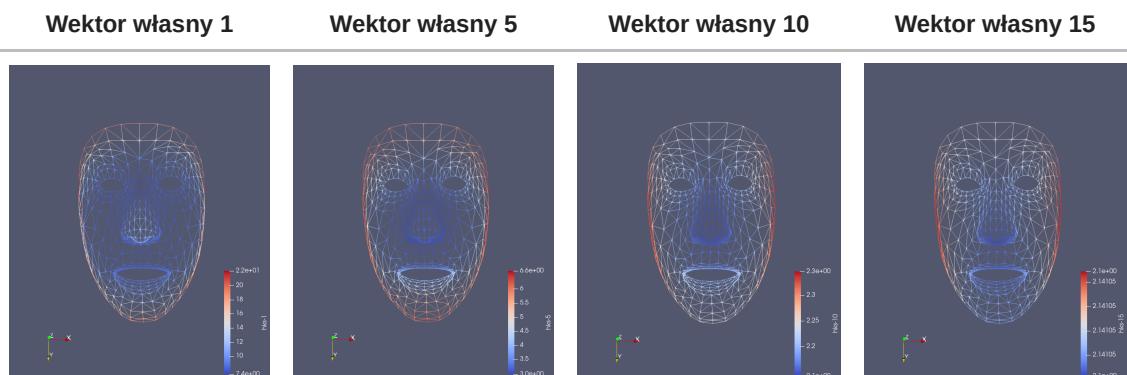
Model grafowy stworzony w szczególności do przetwarzania siatek 3D w formie grafów. Dla wierzchołka agregowanie są tylko cechy z przylegających ścian.



4.2 Cechy wejściowe

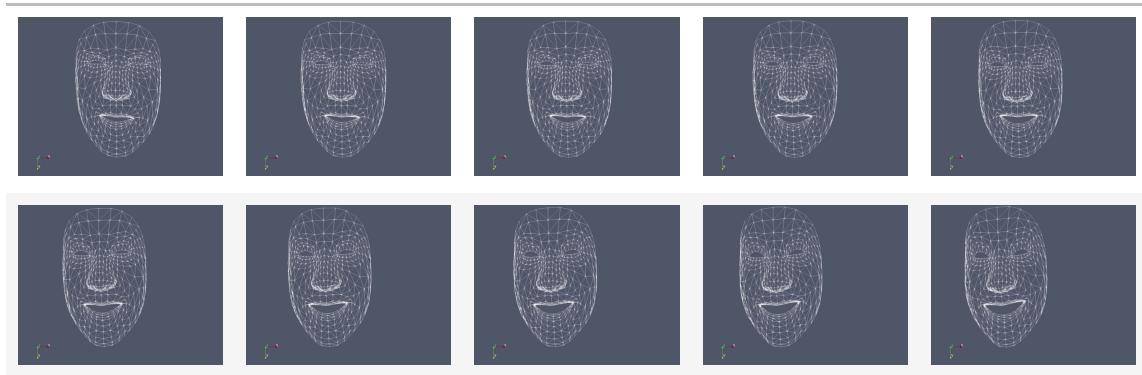
Przetestowaliśmy także dwie metody przedstawienia cech wejściowych wierzchołków:

- XYZ - koordynaty 3D wierzchołków
- HKS - (heat kernel signatures) sygnatury ciepła na siatce, przedstawiają dywergencje gradientu w lokalnym otoczeniu punktu. Sygnatury ciepła uzyskuje się poprzez przejście z bazy cech przestrzennych do bazy cech spektralnych przy użyciu dekompozycji na bazę wektorów własnych operatora Laplace'a-Beltrami'ego. Kolejne wektory własne przedstawiają częstotliwość gradientu. W ramach zadania korzystamy z bazy 16 wektorów własnych. Poniżej przedstawione są wartości konkretnych wektorów własnych w reprezentacji HKS.



5. Augmentacja danych

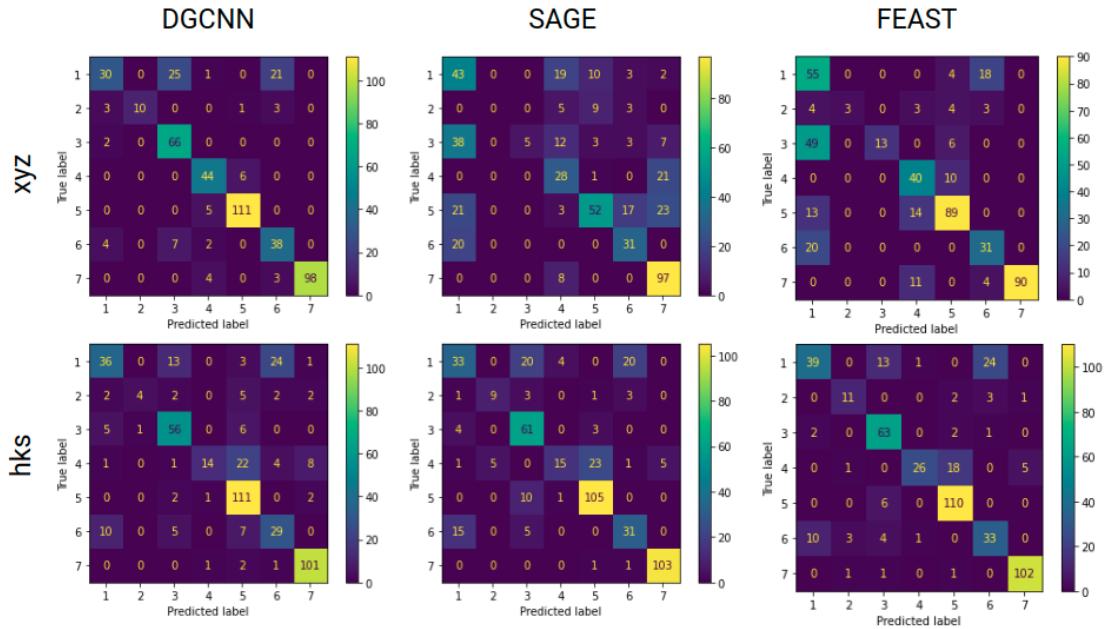
Aby powiększyć zbiór danych oraz dodać bardziej skrajne przypadki w obrębie granic decyzyjnych modelu, zaproponowana została metoda interpolacji pomiędzy emocjami. Nowe próbki są generowane poprzez interpolacje pomiędzy dwoma siatkami - lekkie przesunięcie siatki przedstawiającej jedną emocję w kierunku innej z zachowaniem klasy emocji pierwotnej. W poniższej tabeli pokazane jest przejście z emocji `contempt` do `happy`.



6. Wyniki i ewaluacja modeli

Model	Cechy	Accuracy	F1-macro	Precision-macro	Recall-macro
DGCNN	XYZ	0.82	0.78	0.82	0.78
DGCNN	HKS	0.73	0.62	0.73	0.61
SAGE	XYZ	0.53	0.41	0.52	0.45
SAGE	HKS	0.74	0.66	0.70	0.66
FEAST	XYZ	0.66	0.58	0.59	0.76
FEAST	HKS	0.79	0.74	0.77	0.74

Z naszych eksperymentów wynika, że najlepsze rezultaty osiąga model na reprezentacji w postaci chmur punktów z cechami wejściowymi XYZ. Natomiast dla modeli na siatkach evidentnie widzimy, że cechy HKS poprawiają zdecydowanie wyniki.



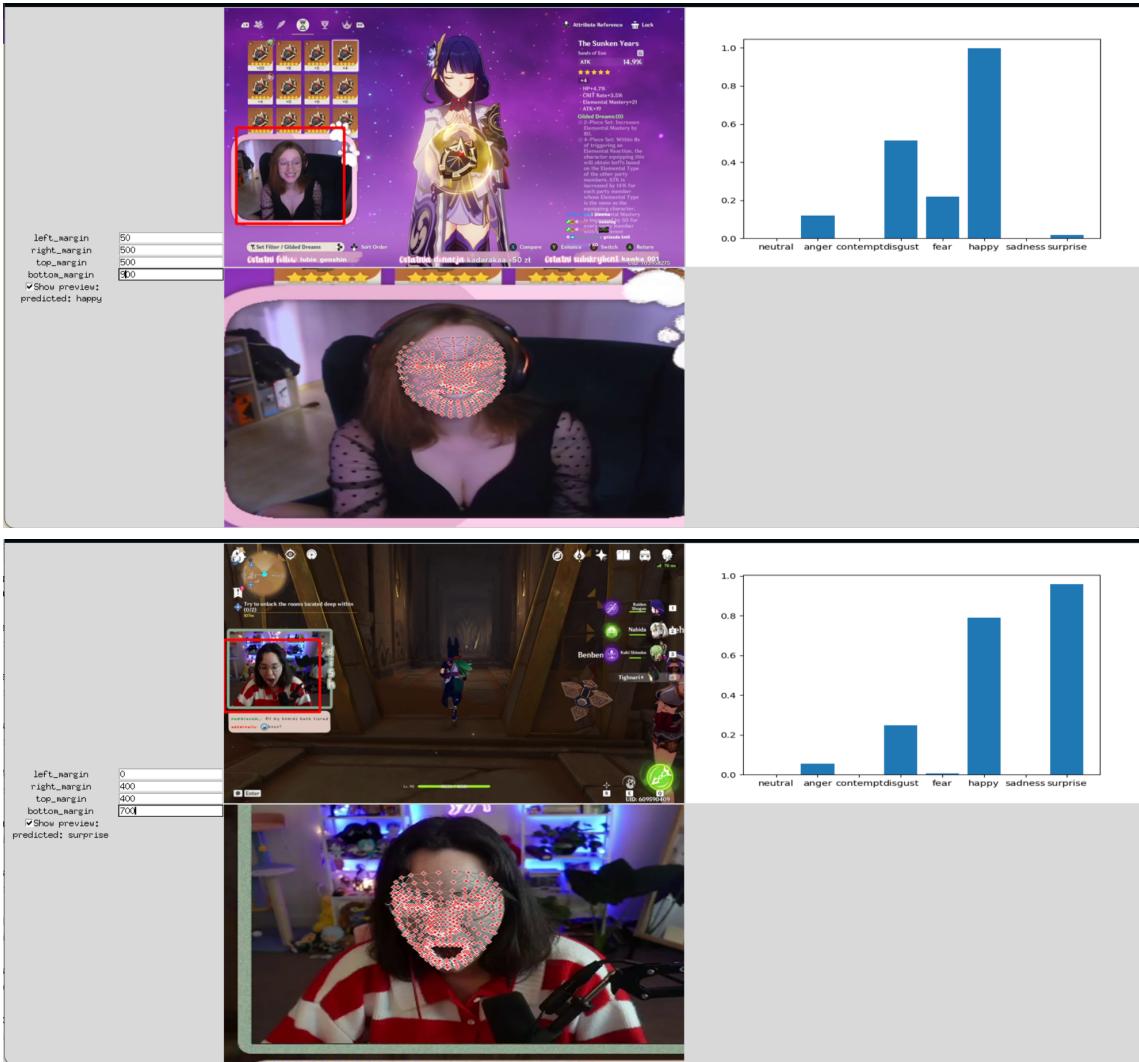
1) anger, 2) contempt, 3) disgust, 4) fear, 5) happy, 6) sadness, 7) surprise.

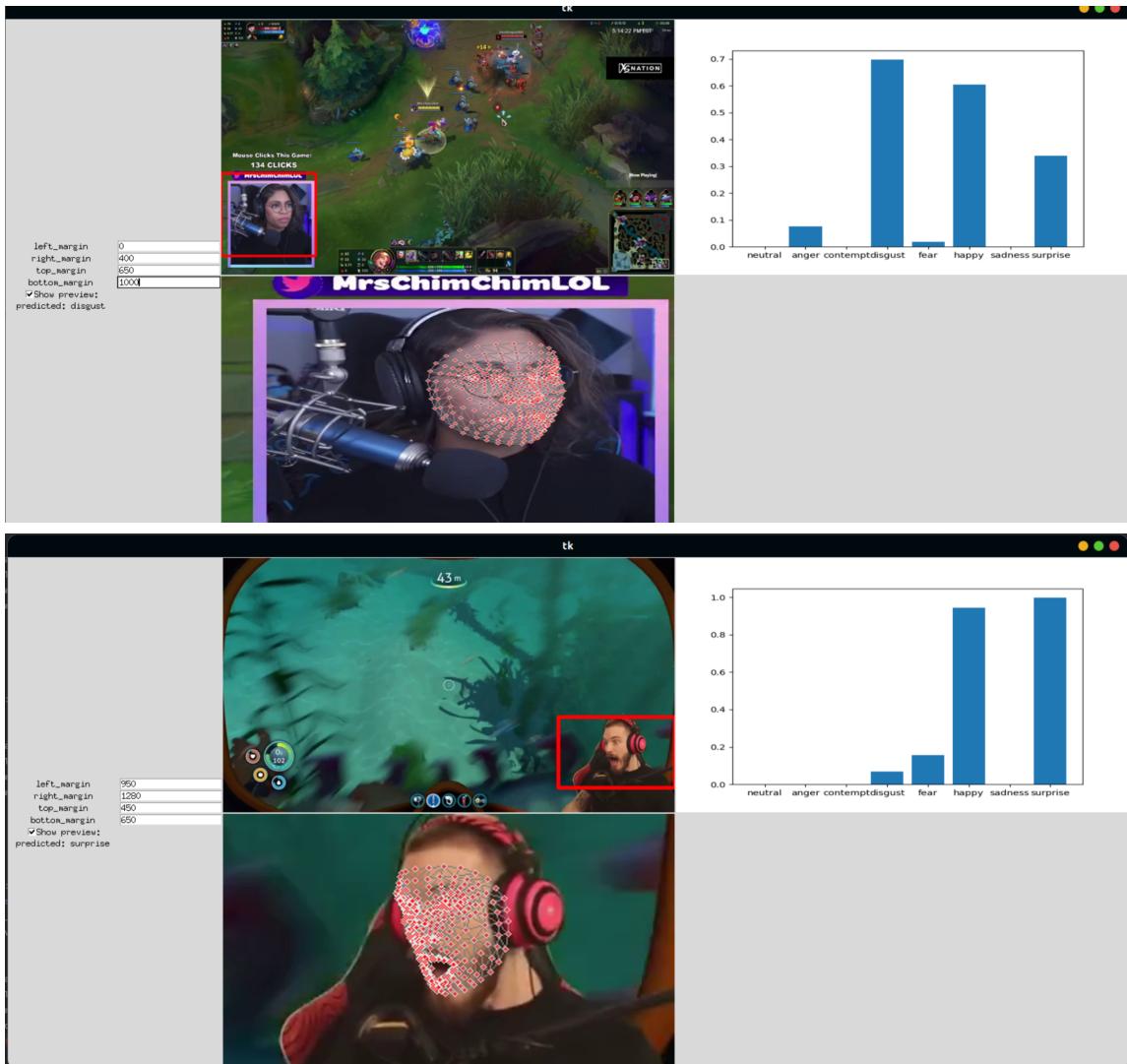
Skupiąc się na analizie macierzy pomyłek dla najlepszych modeli tj. **DGCNN - XYZ** oraz **FEAST - HKS**, widzimy, że emocje **happy** oraz **surprise** są bardzo dobrze wykrywane. Emocje negatywne natomiast, są często mylone między sobą - nie są one aż tak charakterystyczne i nie cechują się tak wysoką ekspresją.

7. Aplikacja analizująca emocje streamer'ów na platformie Twitch - Proof of Concept

7.1 Wczytywanie logów z monitorowanych streamów

Zebraliśmy dane monitorując streamerów na platformie Twitch.tv za pomocą napisanej przez nas aplikacji znajdującej się w tym repozytorium. Aplikacja wyświetla stream oraz predykuje emocje z twarzy osoby znajdującej się w zaznaczonym prostokącie. Wyniki zostają zapisane do określonego pliku **.csv** w celu dalszego przetwarzania.

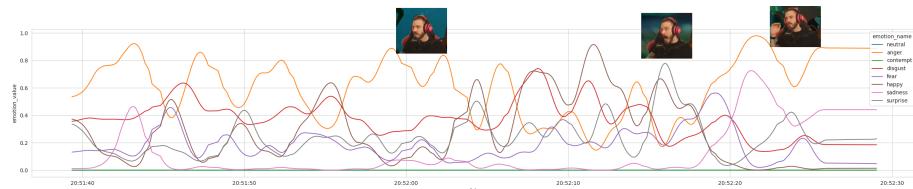




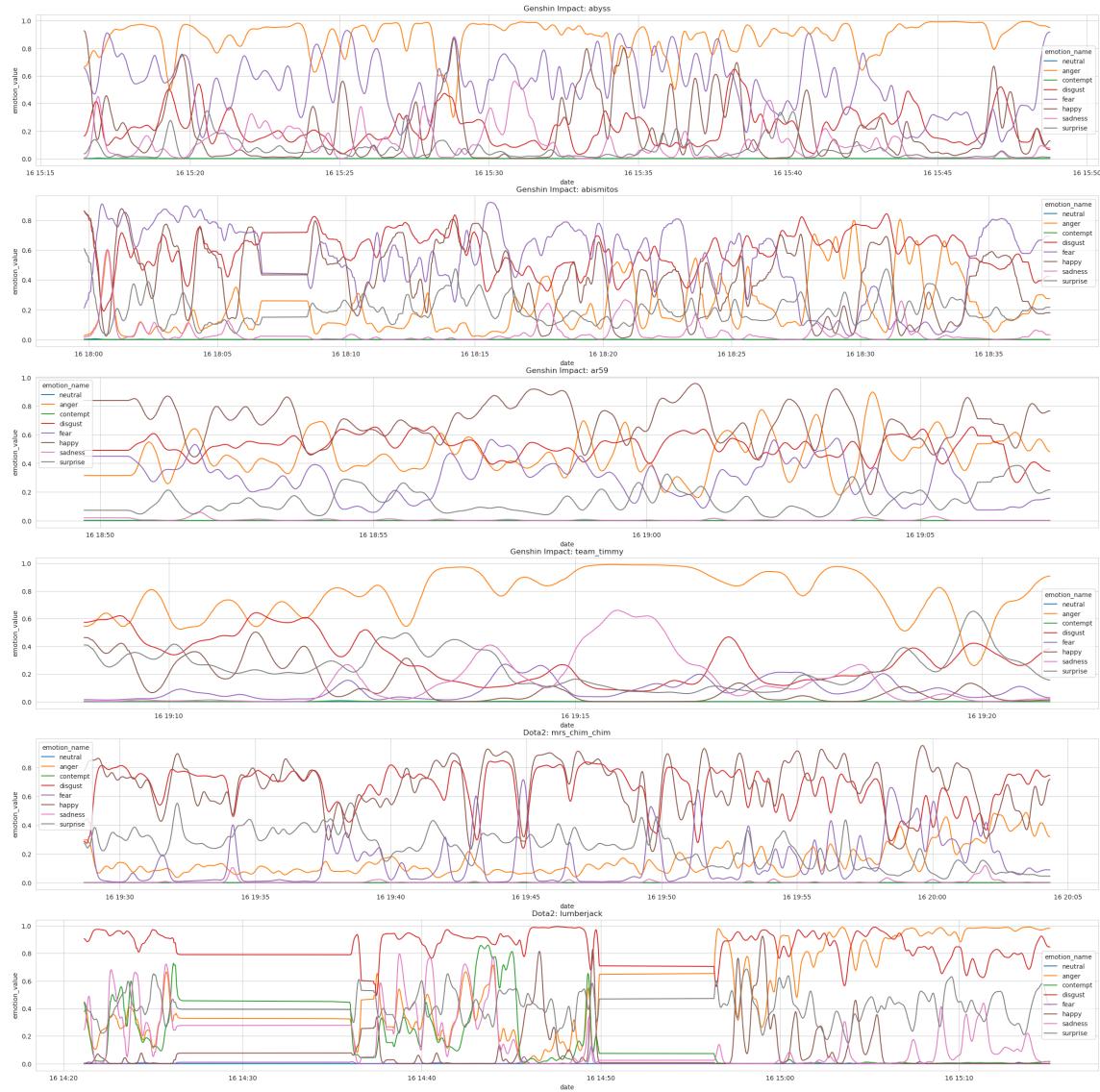
7.2 Analiza emocji w czasie

Można zaobserwować, że różne osoby grające w te same gry mają mocno różniące się od siebie przebiegi emocji spedykowane z twarzy. Jest to prawdopodobnie częściowo zależne od faktu, że kształt twarzy ma duży wpływ na działanie modelu, który był uczyony na niewielkim zbiorze danych. Co więcej model nie nauczył się poprawnie wykrywać emocji **contempt**.

Pomimo tych dużych niedociągnięć na przebiegach można zaobserwować zmiany emocji. Przykładowy przebieg emocji dla krótkiego klipu Youtubera PewDiePie grającego w Subnautica. Pod koniec klipu youtuber zostaje wystraszony, a następnie siedzi niezadowolony. Na modelu widać opisany skok emocji, jednak model niepoprawnie interpretuje przestraszony wyraz twarzy jako połączenie szczęścia i zaskoczenia (i częściowo rozumiemy jego decyzję). Pod sam koniec klipu model całkiem słusznie wskazuje mieszankę złości i smutku.

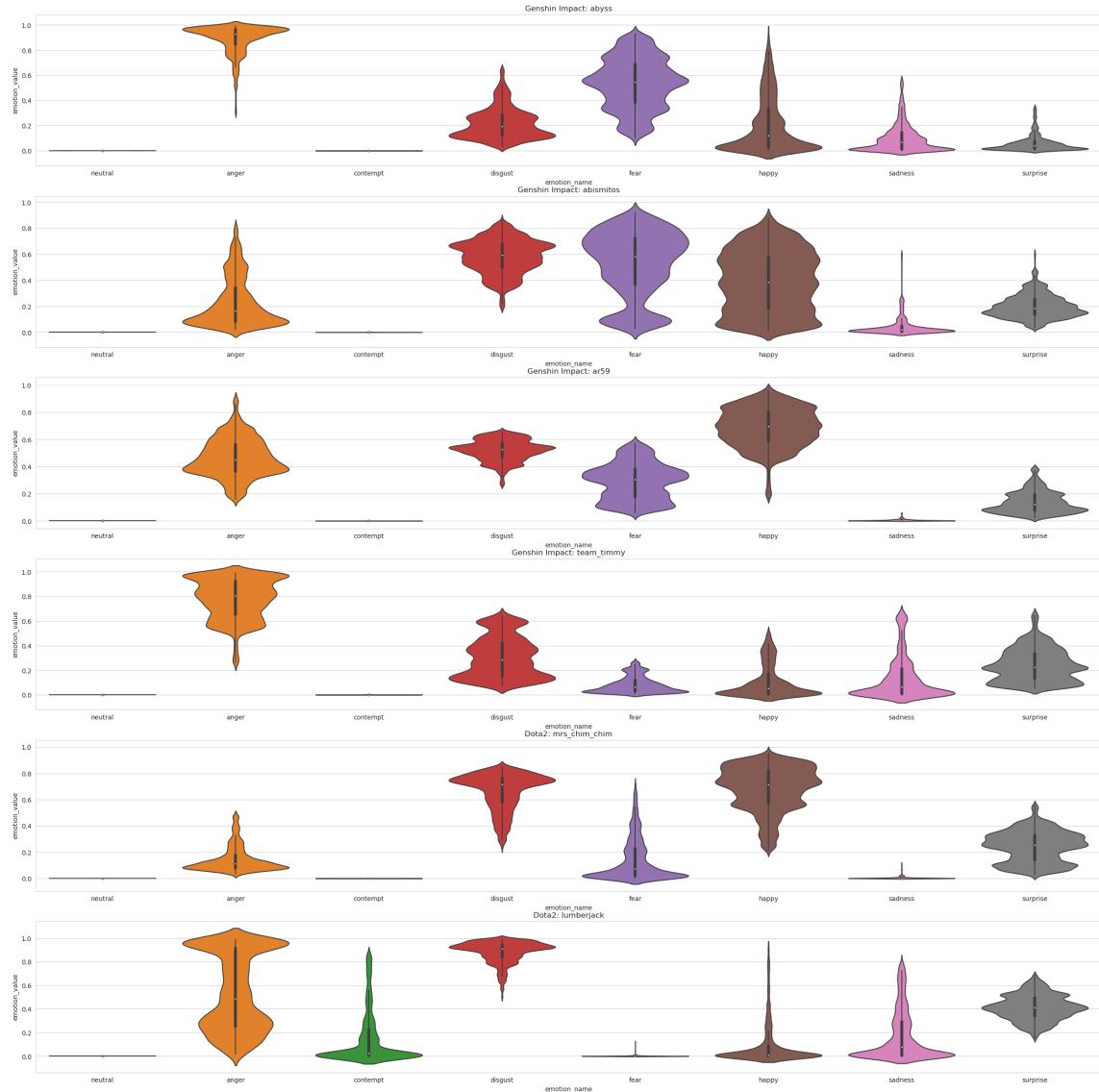


Wykresy dla analizowanych streamerów znajdują się poniżej:



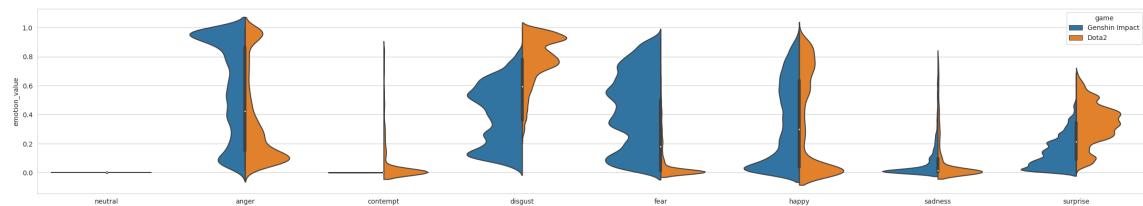
7.3 Analiza rozkładu emocji

W tej części przedstawiony jest rozkład emocji dla streamerów. Można jeszcze lepiej zaobserwować to co było widać w poprzedniej części, a mianowicie, bardzo dużą różnicę w rozkładzie emocji między ludźmi. U niektórych model wykrywa głównie `anger`, u innych głównie `happy`.



7.4 Zestawienie wyników między grami

Analizowani streamerzy grali w grę **Genshin Impact** lub **Dota2**. Rozkład emocji jednej oraz drugiej grupy znajduje się na poniższym wykresie. Należy zaznaczyć, że przy tak dużej wariancji emocji między gracząmi jednej gry należałoby zebrać ogromne ilości danych i włożyć znaczne ilości pracy w dalszą automatyzację tego procesu co wychodzi poza zakres tego projektu.



8. Podsumowanie

W ramach projektu zaproponowaliśmy zastosowanie reprezentacji geometrycznej twarzy w postaci chmur punktów i siatek 3D, do klasyfikacji emocji. Przetestowaliśmy różne architektury i podejścia do reprezentacji danych wejściowych. Jako finalny produkt, zbudowaliśmy PoC aplikacji do zbierania i analizy emocji streamer'ów z platformy Twitch.

