

```
In [ ]: import numpy as np
import pandas as pd

from visualize.representation_projection import plot_rotation_ablation
```

Klasyfikacja emocji z twarzy z użyciem modeli grafowych

Raport z projektu na kurs "Uczenie reprezentacji"

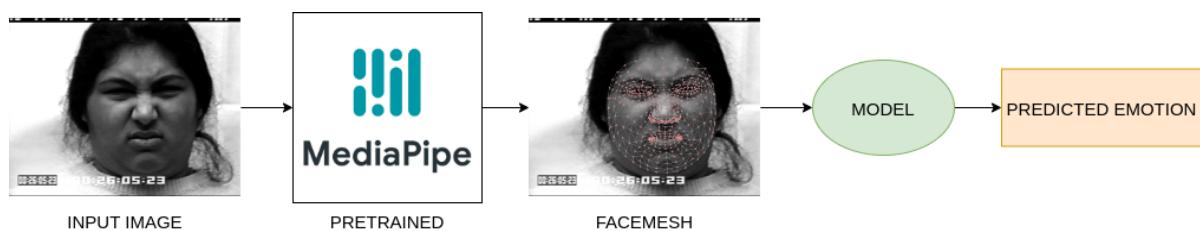
Michał Iłski, Jan Pawłowski, Patryk Rygiel

1. Wstęp

W ramach projektu zajmujemy się klasyfikacją emocji na podstawie wyrazu twarzy. Jest to problem klasyfikacji wieloklasowej, gdzie mamy do czynienia z 7 klasami emocji:

- anger,
- contempt,
- disgust,
- fear,
- happy,
- sadness,
- surprise.

1.1 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). Model nie overfittuje się do tekstur na zdjęciu, jedyne na czym działa to kształt twarzy.

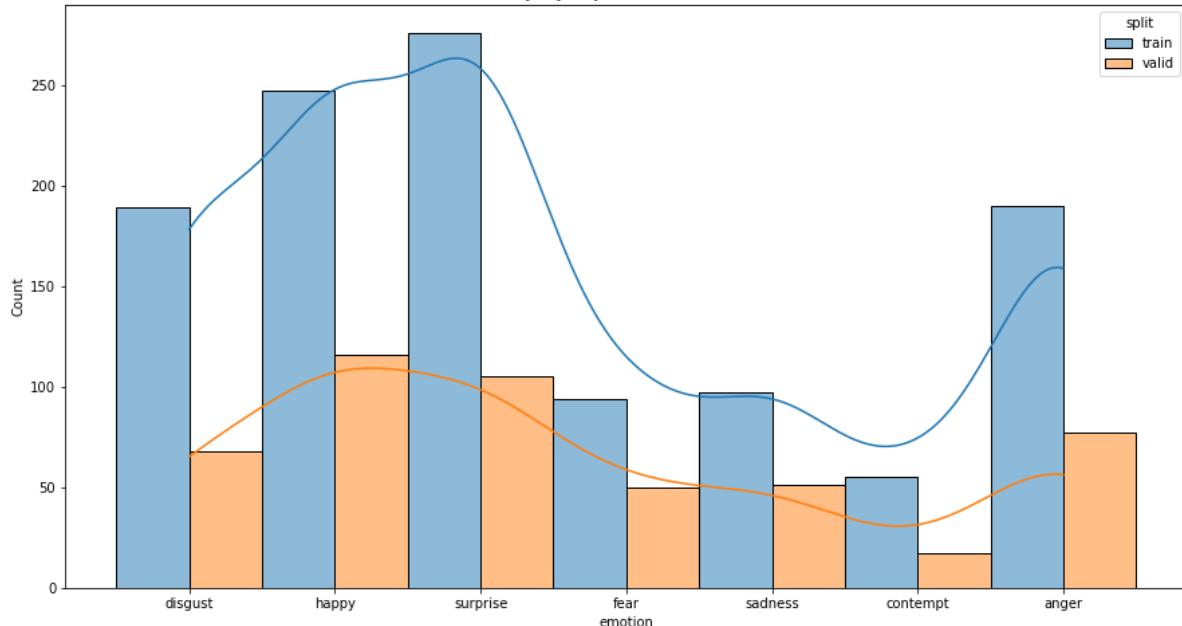
1.2 Zbiór danych

Jako zbioru danych używamy zbioru [CK+ \(Extended Cohn-Kanade dataset\)](#). 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:

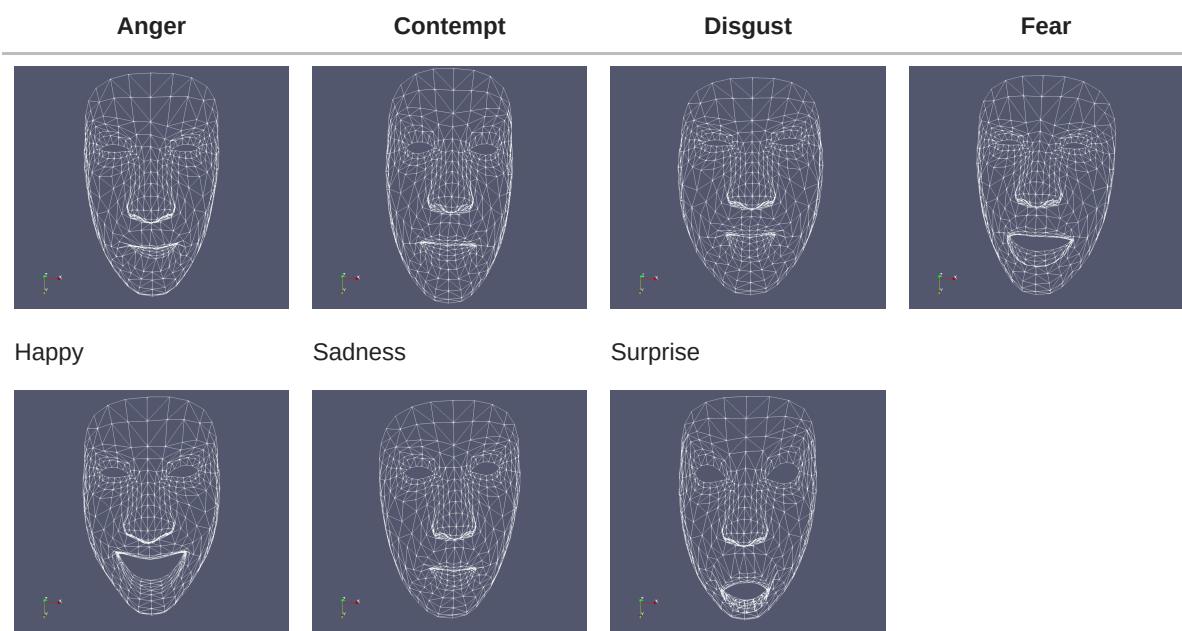


Jako obrazy przedstawiające emocje wybrane zostało ostatnie 20% klatek z sekwencji jako, że na nich intensywność emocji jest największa i dostajemy parę różnych przykładów emocji dla osoby. Jako, że dla każdej osoby wybierana jest więcej niż jedna klatka z sekwencji oraz dla jednej osoby istnieje z reguły 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:

Dystrybucja klas w zbiorach



Tak jak to zostało opisane w ppkt Metodologia , ze zdjęć została dokonana esktrakcja FaceMesh 'y. Poniżej przedstawione są przykładowe siatki dla klas emocji:

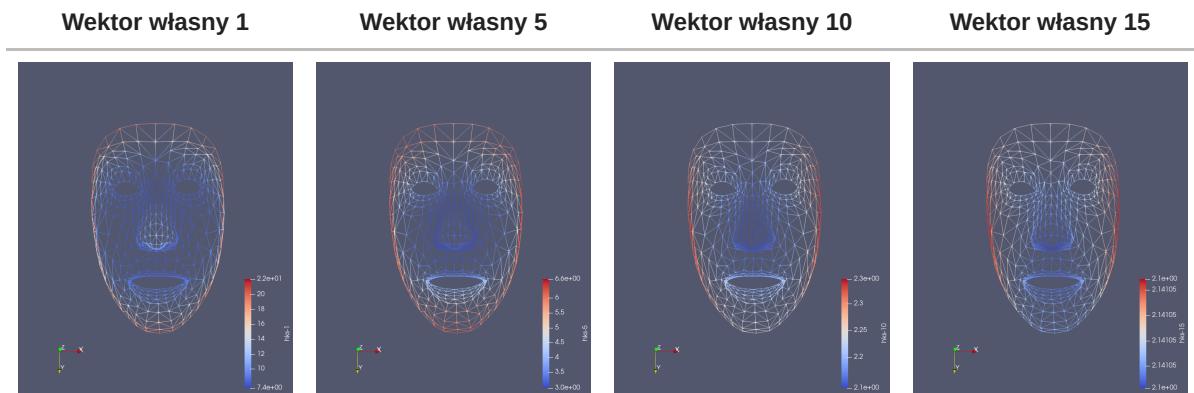


1.3 Modele

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). 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. Sygnatura 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.



Przetestowaliśmy następujące architektury modeli grafowych:

- [DGCNN](#) - model do przetwarzania chmur punktów oparty na dynamicznym grafie budowanym po odległościach wektorów reprezentacji punktów
- [GraphSage](#) - model do przetwarzania grafów bez ukierunkowania na stricte siatki 3D
- [FeaSt](#) - model do przetwarzania siatek 3D w formie grafów

1.4 Badane zagadnienie w kontekście uczenia reprezentacji

1. Z założenia siatki 3D powinniśmy procesować w taki sposób aby ich rotacja w przestrzeni (symetrie grupy $SO(3)$) nie wpływała na budowaną reprezentację:

- Jaki wpływ ma dobór architektury i danych wejściowych (chmura punktów vs mesh) na inwariancję $SO(3)$?
- Jaki wpływ ma wybór cech wejściowych dla wierzchołków (XYZ vs HKS) na inwariancję $SO(3)$?
- Czy nauka modeli z augmentacjami $SO(3)$ pozwala uzyskać invariantną reprezentację?

1. Analiza wyuczonych reprezentacji pomiędzy emocjami:

- Reprezentacje, których emocji są bliskie siebie? Dlaczego?
- Czym różnią się przestrzenie reprezentacji wyuczone przy użyciu różnych cech wejściowych (XYZ vs HKS) ?

1. Analiza rozkładu map aktywacji klas na siatce twarzy:

- Czy aktywowane obszary pokrywają się z klasycznymi deskryptorami emocji? uniesione brwi, otwarte usta itp.
- Które emocje mają między sobą wspólne obszary aktywacji?

*UWAGA: w ramach składanych założeń zadania napisaliśmy, że przetestujemy także model [GEM-CNN](#). Niestety w naszym przypadku ten model nie zadziała, gdyż zakłada on, że mesh jest manifoldem, co nie jest prawdziwe dla naszych danych - krawędzie na obrzeżu maski mają tylko po jednej ścianie, do której należą.

1.5* Skrypty i kod

W folderze `run` znajdują się skrypty:

- `export_meshes` - przeprocesowanie zbioru danych na siatki 3D
- `extract_cams` - ekstrakcja map aktywacji klas w postaci siatek emocji z wygenerowanych reprezentacji
- `train` - nauka wybranego modelu
- `inference` - inferencja wybranego modelu
- `run_preview` - przetestowanie wybranego modelu na kamerze komputera

W folderze `src` znajdują się implementacje modeli, zbioru danych i transformacji.

Odpalenie przykładowego treningu (model SAGE z cechami XYZ):

```
python -m run.train --model-name sage --features xyz
```

2. Analiza zagadnień

2.1 Inwariantność na rotacje 3D (symetrie grupy SO(3))

Przy procesowaniu siatek 3D zależy nam na inwariancji w obrębie symetrii grupy SO(3) - grupa rotacji w 3D. O ile translacje nie są problem w settingu testowym, gdyż zawsze możemy łatwo wyśrodkować siatkę w punkcie 0.0 i zeskalować do zadanej przestrzeni, o tyle rotacje są wyzwaniem.

W ramach tego eksperymentu sprawdzamy jak różne modele i różne reprezentacje cech radzą sobie z inwariancją na rotacje w zadaniu klasyfikacji emocji.

2.1.1 Trening bez rotacji

W pierwszym teście trenujemy modele bez augmentacji danych w postaci rotacji. Dokonujemy ewaluacji na domyślnym zbiorze testowym (F1-macro) oraz na zbiorze testowym z nałożonymi losowymi rotacjami (F1-macro z rotacjami).

Model	Cechy	F1-macro (%)	F1-macro z rotacjami (%)
DGCNN	XYZ	82.94	7.20
DGCNN	HKS	71.64	72.86
SAGE	XYZ	75.97	6.31
SAGE	HKS	77.86	71.96
FEAST	XYZ	76.36	13.60
FEAST	HKS	73.54	69.23

Zauważamy, że najlepsze wyniki na domyślnym zbiorze testowym osiąga model DGCNN z cechami XYZ. Jednak, gdy testujemy go na zbiorze z rotacjami, jak i każdy inny model na cechach XYZ, zauważamy ogromny spadek w jakości klasyfikacji - praktycznie wyniki są

losowe. Dla cech HKS natomiast widzimy, że inwariantność na rotacje jest zachowana. Nie jest to niespodzianka, gdyż sygnatury ciepła są zależne tylko od kształtu siatki niezależnie od tego w jakim miejscu przestrzeni owa siatka się znajduje.

2.1.2 Trening z rotacjami

W drugim teście trenujemy modele z augmentacją danych w postaci losowych rotacji. Ewaluację dokonujemy w ten sam sposób co w teście pierwszym.

Model	Cechy	F1-macro (%)	F1-macro z rotacjami (%)
DGCNN	XYZ	51.40	43.26
DGCNN	HKS	70.96	74.68
SAGE	XYZ	11.51	8.30
SAGE	HKS	70.35	72.42
FEAST	XYZ	12.16	9.19
FEAST	HKS	76.81	70.29

Tym razem dostrzegamy, że DGCNN z XYZ osiąga lepsze wyniki na zbiorze z rotacjami, jednak są one dużo słabsze niż to co są w stanie osiągnąć modele z cechami HKS. Modele SAGE i FEAST nie są w stanie się zbytnio niczego nauczyć przy użyciu cech XYZ, gdy dokonujemy rotacji. Ponownie bez zaskoczeń, przy wykorzystaniu cech HKS wyniki są takie same jak przy trenowaniu modelu bez rotacji.

Wnioski

Inwariantność na dane symetrie (w naszym przypadku SO(3)) możemy wymusić na dwa sposoby:

- Augmentacja danych w obrębie symetrii grupy
- Budowa reprezentacji invariantnej na symetrie (invariantny model lub reprezentacja danych wejściowych)

Na podstawie przeprowadzonych eksperymentów, ewidentnie zauważamy, że budowa invariantnej reprezentacji jest lepszym podejściem. Prawdopodobnie, zastosowanie augmentacji zbioru mogłoby dać zbliżone wyniki, jednak wymagałoby ono zdecydowanego powiększenia zbioru danych, aby móc dobrze pokryć powiększoną przestrzeń danych wejściowych. Używając modeli / reprezentacji invariantnych na symetrie nie potrzebujemy aż tak wielu danych, gdyż przestrzeń danych wejściowych jest zdecydowanie mniejsza.

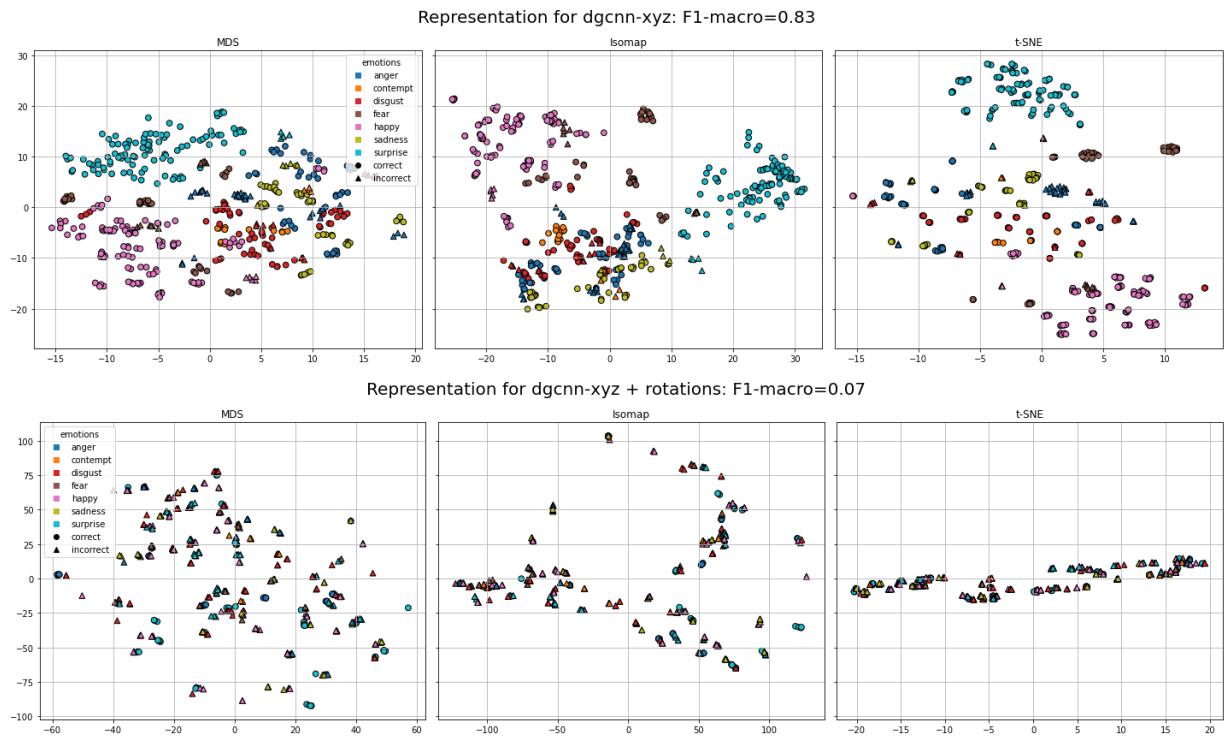
2.2 Analiza wyuczonych reprezentacji klas emocji

Jako, że modele operujące na grafie tj. SAGE i FEAST, zachowują się w miarę podobnie w zakresie wyżej przedstawionych badań, w kolejnych analizach, w celu zmniejszenia przytłoczenia wykresami, skupimy się głównie na porównaniu modeli DGCNN oraz FEAST.

2.2.1 DGCNN

a) Cechy XYZ uczone bez rotacji

```
In [ ]: plot_rotation_ablation("dgcnn-xyz")
```

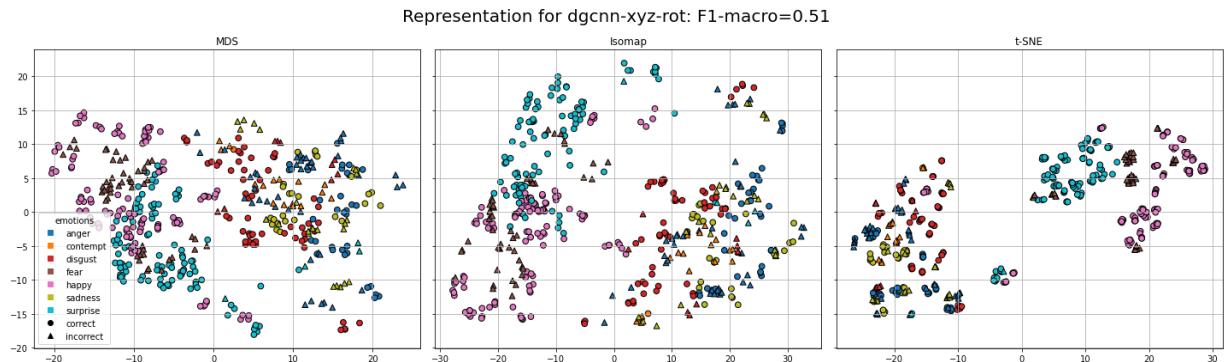


Analizując wyuczoną przestrzeń reprezentacji bez uczenia z rotacjami na domyślnym zbiorze testowym (rys. 1), zauważamy, że ze względu na niektóre emocje przestrzeń jest dosyć dobrze podzielona. Najbardziej charakterystyczne reprezentacje osiągają klasy *surprise* oraz *happy* - prawdopodobnie wpływa na to wysoki poziom ekspresji emocji, który pozwala na łatwą dyskryminację. Ciekawą obserwacją jest także to, że klasy emocji negatywnych *anger*, *contempt*, *sadness* i *disgust*, są dużo bardziej do siebie zbliżone niż do emocji pozytywnych tj. *happy*.

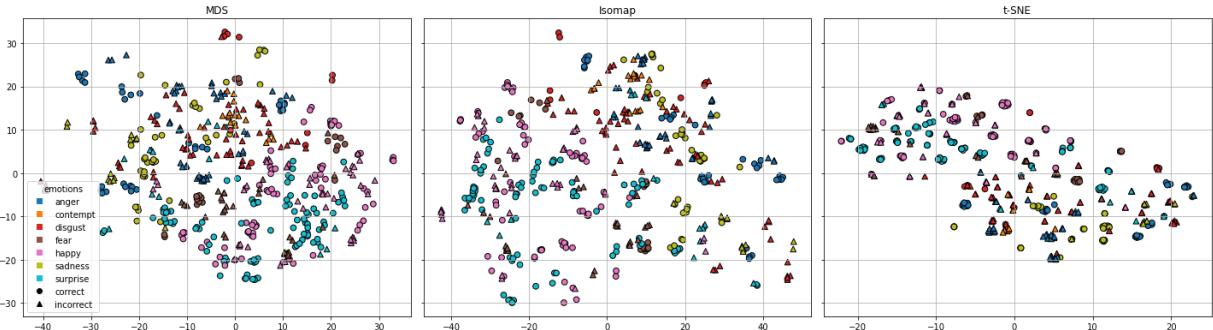
Patrząc na reprezentację zewaluowaną na zbiorze z rotacjami, widzimy, że przestrzeń osadzeń nie posiada żadnych istotnych klastrów klas, które pozwalały by na poprawną dyskryminację. Widzimy, że model zdecydowanie nie jest w stanie generalizować na rotacje siatek i tworzone reprezentacje są nic nie warte.

b) Cechy XYZ uczone z rotacjami

```
In [ ]: plot_rotation_ablation("dgcnn-xyz-rot")
```



Representation for dgcnn-xyz-rot + rotations: F1-macro=0.43

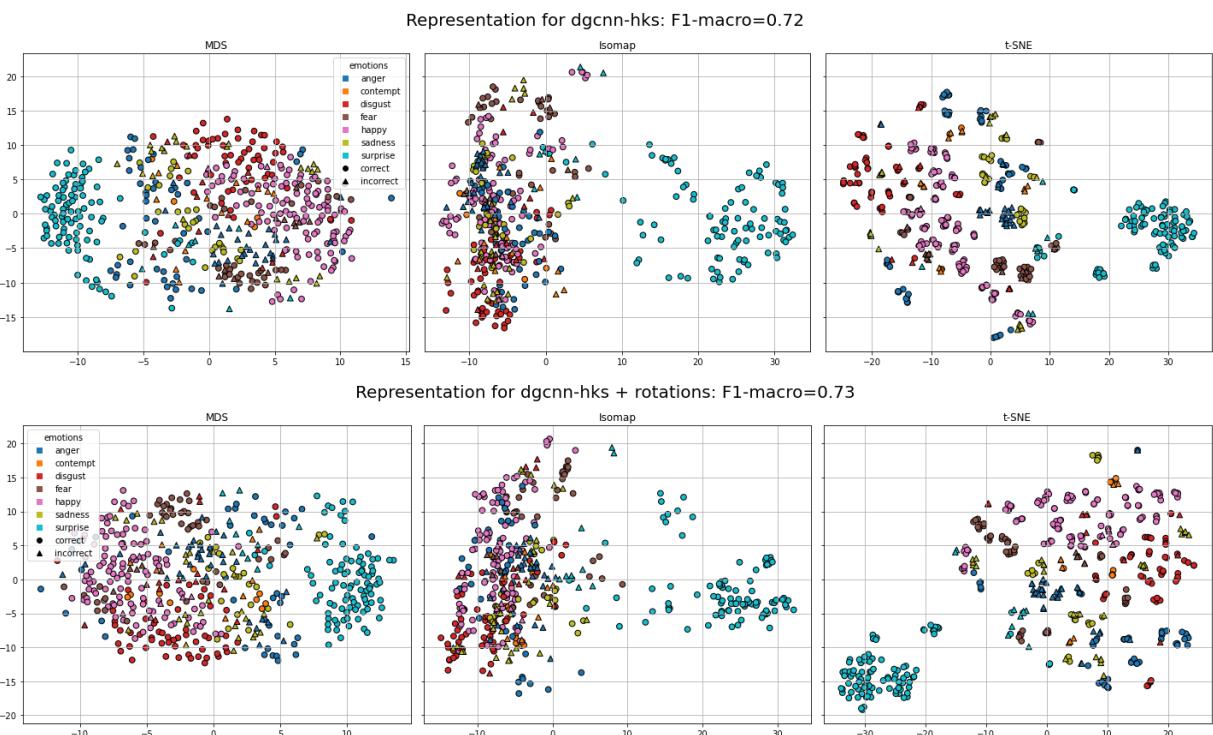


Patrząc na wytworzzone reprezentacje przy uczeniu modelu na rotacjach, zauważamy, że o ile reprezentacje są bardziej przemieszane ze względu na klasę niż w przypadku trenowania bez rotacji, o tyle reprezentacje na zbiorze z rotacjami i bez są dosyć zbliżone. Potwierdza to tezę, że pewien stopień inwariancji na daną symetrię da się osiągnąć poprzez inkorporację augmentacji w jej zakresie podczas uczenia modelu.

Ponownie klasy negatywne i pozytywne są dosyć dobrze od siebie oddzielone. Tym razem klasy happy oraz surprise są bardziej ze sobą zmieszane, a dodatkowo dla klasy fear praktycznie prawie wszystkie predykcje są błędne i jest ona klasyfikowana jako happy lub surprise .

c) Cechy HKS

```
In [ ]: plot_rotation_ablation("dgcnn-hks")
```



Przestrzenie reprezentacji dla cech HKS wydają się być bardziej ciągłe - dla XYZ dostrzegaliśmy więcej małych blobów. Ponownie widzimy, że klasa surprise jest łatwo rozróżnialna w porównaniu do innych klas. Tym razem jednak, klasa pozytywnych emocji w postaci happy jest bardziej przemieszana z innymi klasami - dla XYZ widzieliśmy dosyć jasny podział między emocjami negatywnymi, a pozytywnymi.

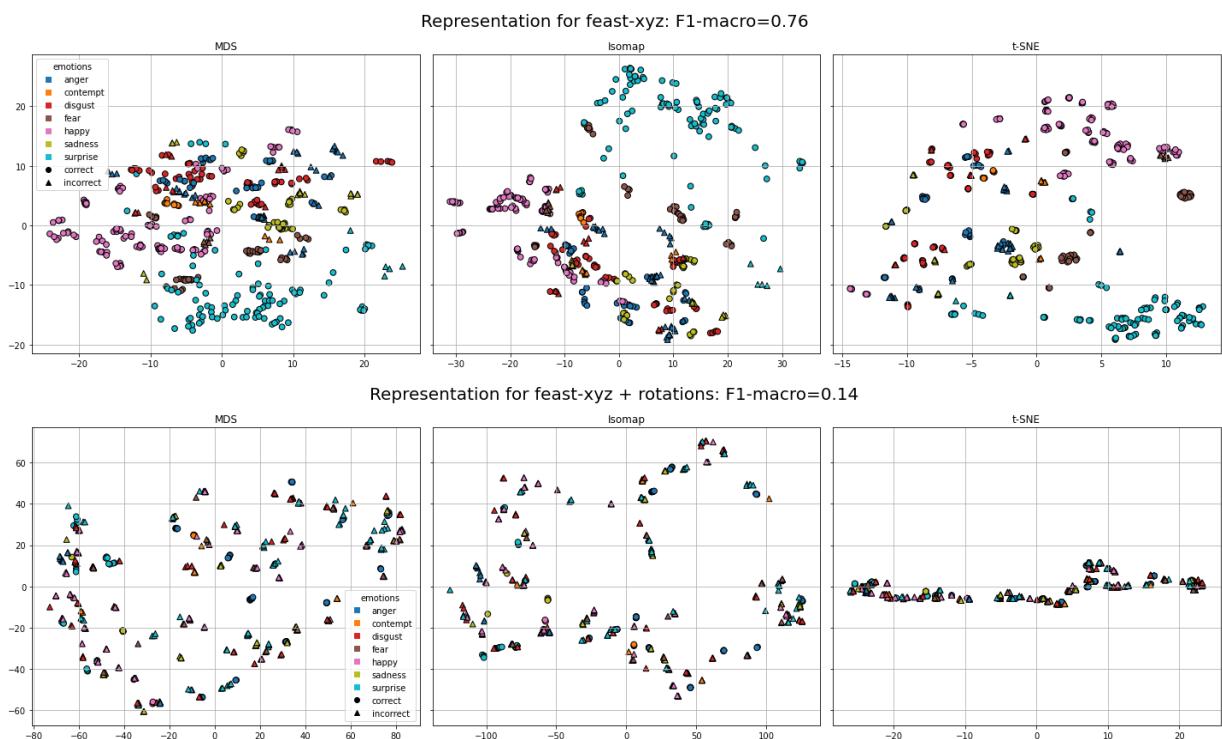
Ponownie porównujemy zbudowaną reprezentację na zbiorze testowym z rotacjami i bez.

Dostrzegamy, że faktycznie budowana reprezentacja jest invariantna na rotacje w przestrzeni - ten sam fakt wyjaśniliśmy w 2.1.1.

2.2.2 FEAST

a) Cechy XYZ uczone bez rotacji

```
In [ ]: plot_rotation_ablation("feast-xyz")
```

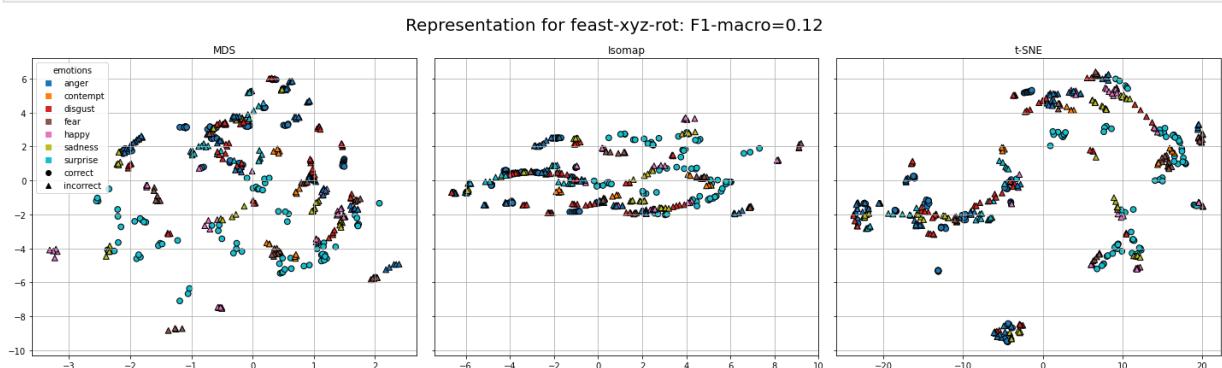


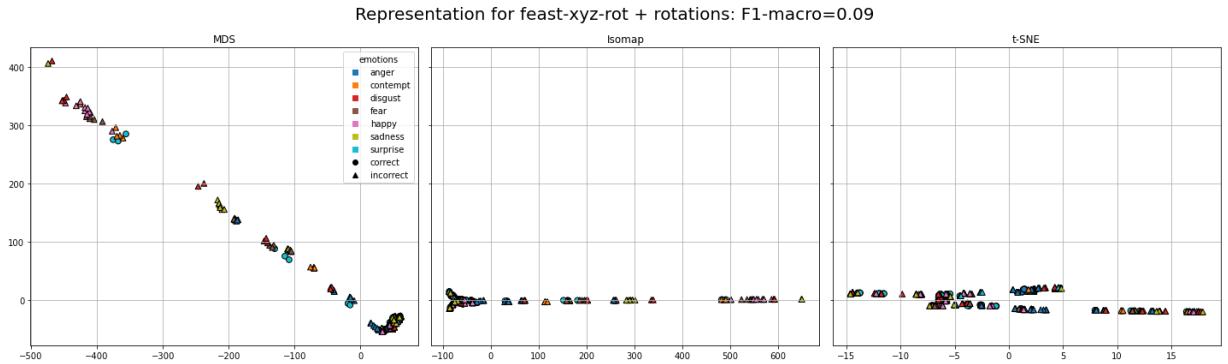
Dla cech XYZ ponownie widzimy, że emocja **surprise** jest łatwo separowalna. Emocja **happy** jest trochę bardziej połączona z innymi reprezentacjami lecz ciągle klasy emocji negatywnych są w jednym miejscu.

Przy testowaniu na zbiorze z rotacjami, wyuczone reprezentacje nie niosą ze sobą w zasadzie żadnej informacji na temat klasy.

b) Cechy XYZ uczone z rotacjami

```
In [ ]: plot_rotation_ablation("feast-xyz-rot")
```

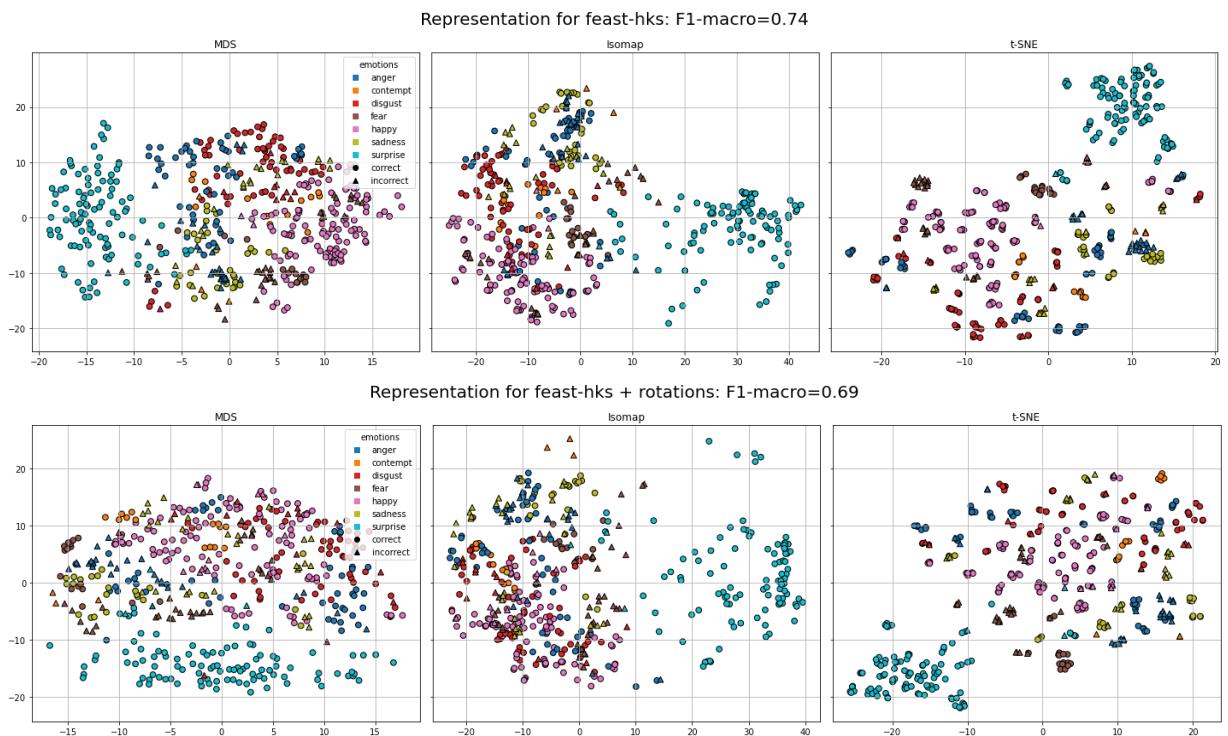




W porównaniu do DGCNN, model FEAST nie był w stanie wyuczyć się na zbiorze z rotacjami. Reprezentacje dla zbioru testowego z rotacjami i bez są podobnie słabej jakości. W zasadzie jedyna emocja jaka jest poprawnie klasyfikowana w zbiorze bez rotacji to surprise.

c) Cechy HKS

```
In [ ]: plot_rotation_ablation("feast-hks")
```

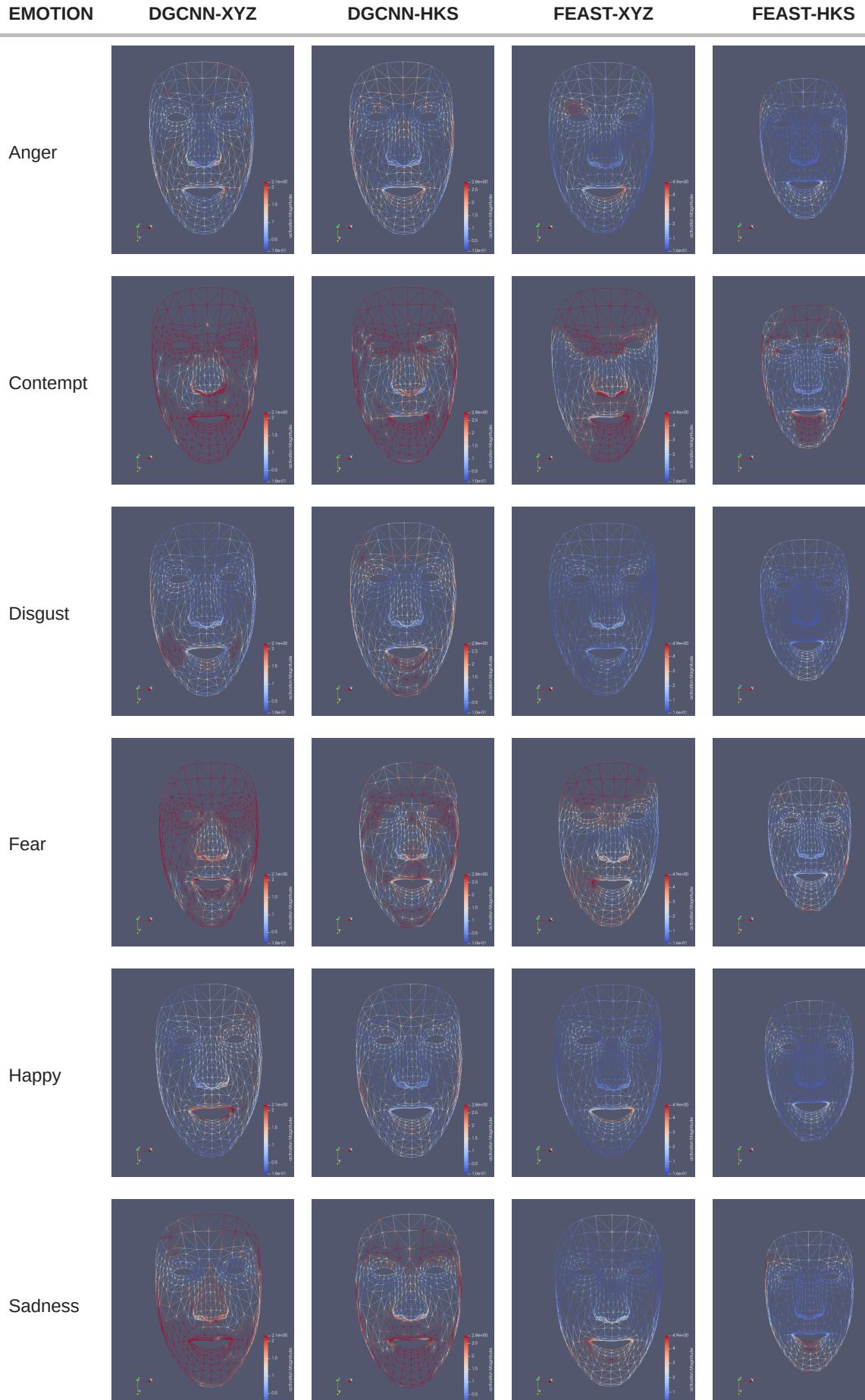


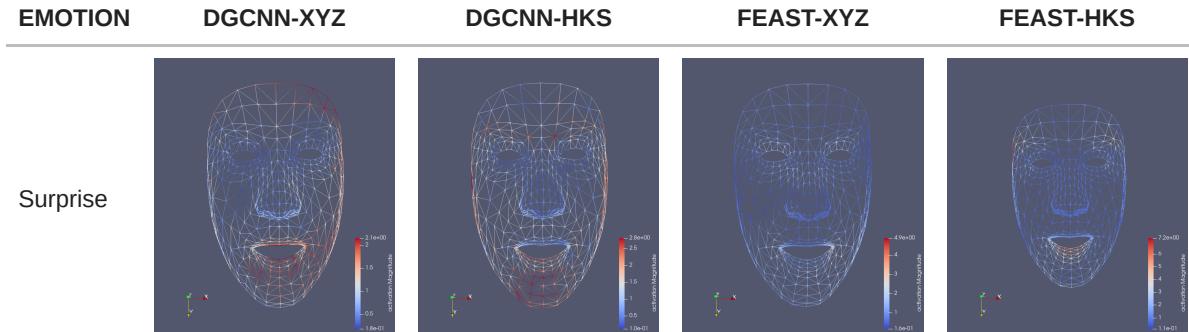
Dla modelu FEAST z cechami HKS dotrzegamy, że reprezentacja jest bardzo zbliżona do tej, którą ma model DGCNN z tymi samymi cechami.

2.3 Analiza map aktywacji klas

Mapy aktywacji klas zostały wygenerowane poprzez projekcję wektorów osadzeń wierzchołków przy użyciu PCA na 3 wektory główne. Następnie projekcje zostały uśrednione w obrębie klas (tylko zbiór testowy). Poniższa tabela przedstawia długość wektora projekcji PCA jako poziom aktywacji - skale są różne dla modeli ale takie same w obrębie jednego modelu.

EMOTION	DGCNN-XYZ	DGCNN-HKS	FEAST-XYZ	FEAST-HKS
---------	-----------	-----------	-----------	-----------





1. Anger

Obszary aktywacji dla DGCNN nie są zbyt charakterystyczne. Dla FEAST widzimy aktywację w zakresie dolnej części ust (szczególnie dla cech HKS).

2. Contempt

Dla wszystkich modeli główne aktywacje występują na czole oraz pod ustami. Może to charakteryzować zmarszczenie czoła, które występuje przy tej emocji.

3. Disgust

Dla tej emocji nie ma żadnej zbyt charakterystycznej aktywacji, która występuje pomiędzy modelami. Dla DGCNN aktywacje są na dolnej wardze i kącikach ust, lecz dla FEAST same aktywacje są relatywnie małe w porównaniu do innych emocji.

4. Fear

Aktywacje pomiędzy modelami dosyć mocno różnią się od siebie. Trudno wskazać jeden charakterystyczny obszar pomiędzy modelami. Dla DGCNN-XYZ prawie cała twarz oprócz nosa się aktywuje, dla DGCNN-HKS występują nieregularne aktywacje wokół oczu. Dla FEAST-XYZ aktywuje się za to czoło, a dla FEAST-HKS niezbyt da się wskazać coś charakterystycznego.

5. Happy

Główne miejsce aktywacji to usta, co jest charakterystyczne, gdyż wskazuje prawdopodobnie na to, że usta z reguły przy tej emocji są otwarte. Najbardziej charakterystyczna aktywacja powstaje dla FEAST-HKS, gdzie aktywuje się dolna warga oraz kąciki oczu - są to części twarzy, które z reguły uczestniczą w tej emocji.

6. Sadness

Dla tej emocji dostrzegamy różnice pomiędzy modelem DGCNN a FEAST. Dla pierwszego z nich aktywuje się głównie dolna część twarzy oraz lekko czoło. Dla drugiego natomiast, aktywacja występuje głównie w zakresie brody.

7. Surprise

Najciekawsze aktywacje widać dla modelu FEAST. Jako charakterystyczne punkty aktywują się usta oraz obszary wokół brwi. Z reguły ruchy w tych obszarach pojawiają się podczas ekspresji tej emocji.

Wnioski

- Najbardziej charakterystyczne aktywacje występują dla modelu FEAST-HKS

- Dla części z emocji (np. happy , surprise) obszary aktywacji pokrywają się obszarami, które identyfikujemy z ekspresją danej emocji - głównie są te emocje, które dobrze się separowały na wizualizacji przestrzeni reprezentacji

3. Podsumowanie

W ramach projektu dokonaliśmy analizy jakości reprezentacji tworzonych w zadaniu klasyfikacji emocji z twarzy przedstawionych w formie siatki 3D. W badaniach sprawdziliśmy inwariantność modeli na rotacje 3D w zależności od architektury modelu, sposobu uczenia (augmentacje) oraz rodzaju podawanych cech na wejściu. Porównaliśmy także rozkład przestrzeni reprezentacji dla wszystkich podejść oraz przeanalizowaliśmy interpretowalność wyuczanej reprezentacji na mapach aktywacji klas.