

Politechnika Warszawska

Warsztaty z Technik Uczenia Maszynowego

**APLIKACJA DO WYKRYWANIA
TWARZY I OKREŚLANIA EMOCJI
WYKRYTYCH OSÓB**

Wojciech Gajda 304494

Jakub Brzóskowski 313180

Aleksy Bałaziński 313173

Jan Cichomski 313201

Prowadzący: dr inż. Janusz Rafałko

Data oddania: **9 marca 2023**

Spis treści

1 Wstęp	3
1.1 Opis	3
1.2 Plan pracy	4
1.3 Podział zadań	4
2 Przebieg projektu	5
2.1 Wybór narzędzi	5
2.2 Wyodrębnianie twarzy ze zdjęć	6
2.2.1 Wybór typu sieci	6
2.2.2 Przegląd gotowych rozwiązań	6
2.2.3 Opis danych treningowych	6
2.2.4 Proces uczenia	6
2.2.5 Ocena efektów uczenia	6
2.3 Rozpoznanie emocji	7
2.3.1 Wybór typu sieci	7
2.3.2 Opis danych treningowych	7
2.3.3 Proces uczenia	8
2.3.4 Ocena efektów uczenia	8
2.4 Aplikacja użytkownika – GUI	9
2.5 Podsumowanie	10
3 Literatura	11

1 Wstęp

Rozpoznawanie twarzy w obrazach wykorzystywane jest obecnie w niezliczonych dziedzinach. Systemy monitoringu przetwarzają dziennie petabajty informacji, które bez odpowiedniego sklasyfikowania stanowią bezużytecznych zbiór nagrani. Aplikację do przechowywania zdjęć starają się porządkować przechowywany kontent, rozpoznając podobne twarze.

Jednocześnie samo wyodrębnienie twarzy nie niesie dużej wartości informacyjnej. Można, co prawda, zliczyć ilość osób objętych danym ujęciem, jednak bez dalszej analizy takie zdjęcia przedstawiają po prostu, ludzi. Więcej informacji uzyskujemy podając dopiero tak przygotowaną i wyodrębnioną twarz dalszej analizie. Obecnie dostępne narzędzia pozwalają oszacować wiele cech rozpoznanej osoby. Twarz skrywa информацию o wieku, płci, rasie, ale również bardziej ulotne cechy jak uśmiech, smutek, zmęczenie itp. W ramach tego projektu skupimy się właśnie tych ulotnych cechach które w ogólności można nazwać emocjami.

1.1 Opis

Celem projektu jest przygotowanie aplikacji do rozpoznawania emocji osób znajdujących się na zdjęciu. Działanie programu można podzielić na kilka etapów. Najpierw do aplikacji wczytywane jest dowolne zdjęcie. W kolejnym kroku na załadowanym zdjęciu wykrywane są ludzkie twarze. Wykrycie twarzy pozwala wizualizować efekt klasyfikacji przez dorysowanie kontrastujących ramek wokół twarzy. Jednocześnie każdą z twarzy można wyekstrahować ze zdjęcia i dodać do zbioru. Każdy z elementów tak utworzonego zbioru jest następnie poddany analizie i przypisywana jest mu jedna z przyjętych klas, reprezentujących emocję.

Powyższa aplikacja do działania wymaga jednak wytrenowania dwóch odmiennych modeli sieci: do binarnej klasyfikacji twarzy i klasyfikacji wyodrębnionych twarzy względem emocji. Proces treningu ww. modelu przedstawiony zostanie w odrębnych skryptach (notatnikach).

1.2 Plan pracy

1. Przegląd literatury.
2. Przygotowanie modułu do wyodrębniania twarzy przy pomocy pretrenowanego modelu sieci.
3. Obróbka danych uczących i wytrenowanie modelu do klasyfikacji emocji.
4. Przygotowanie modułu do rozpoznawania emocji.
5. Połączenie modułów w jednorodną aplikację z GUI.
6. Obróbka danych uczących i wytrenowanie własnego modelu do wykrywania twarzy.

1.3 Podział zadań

Praca nad projektem realizowane są w formie wewnętrznych spotkań. W naturalny sposób powoduje to, że postęp realizowany jest wspólnie, bez większego podziału ról. By łatwiej utrzymywać jakość projektu, do poszczególnych obszarów przydzielana została osoba odpowiedzialna. Podział ten prezentuje się następująco:

- Wojciech Gajda:
 - opracowanie architektury projektu i integracja modułów projektu
 - przygotowanie aplikacji użytkownika (GUI)
- Jan Cichomski:
 - przegląd literatury
 - dobór danych uczących
- Jakub Brzóskowski:
 - opis danych uczących do rozpoznawania twarzy
 - wytrenowanie modelu do klasyfikacji twarzy na zdjęciach
- Aleksy Bałaziński:
 - opis danych uczących do rozpoznawania emocji
 - wytrenowanie modelu do rozpoznawania emocji

2 Przebieg projektu

2.1 Wybór narzędzi

Jako główny język programowania wykorzystany zostanie **Python**. Język ten zdobył olbrzymią popularność w domenie Machine Learningu, Deep Learning i sztucznej inteligencji. Za popularnością Python'a przemawia jego prostota i niski próg wejścia. Dodatkowo ogromna społeczność zgromadzona wokół języka ułatwia poszukiwanie informacji i rad.

Do rozwiązania problemu ekstrakcji twarzy z zdjęcia oraz ogólnie pojętej obróbki wykorzystana zostanie biblioteka **OpenCV**. Biblioteka stanowi rozbudowany zbiór narzędzi do obróbki obrazu oraz zbiór podstawowych modeli sztucznych sieci neuronowych ukierunkowany na wykrywanie wzorców i wizję komputerową. Biblioteka zawiera model kaskadowej sieci z funkcjami Haara, która w ostatnich latach uznawana był jako główne narzędzie do klasyfikacji twarzy. W plikach źródłowych projektu znaleźć można pretrenowany model sieci, przeznaczony do odnajdywania twarzy w obrazach. Biblioteka ~~umożliwia również samodzielny trening takiego modelu~~. Niestety wsparcie dla tego rozwiązania zostało porzucone. Przyczyną tego jest przewaga, którą dają współczesne sieci DNN. Najnowsze artykuły wskazują, że rezultaty działania takich detektorów (np. z biblioteki TensorFlow lub YOLO) dają znacznie większą dokładność, przy zachowaniu zbliżonej prędkości obliczeń.

Klasyfikacja emocji na wydzielonych twarzach jest natomiast zadaniem dla którego nie istnieje powszechnie przyjęte rozwiązanie. W literaturze istnieje wiele podejść tego problemu jednak znacząca większość opiera swoje działanie na różnych odmianach konwolucyjnych sieci neuronowych. Większość z tych modeli dostępna jest w popularnej bibliotece **TensorFlow**. Ponadto biblioteka ta umożliwia łatwe wykorzystanie układów graficznych do trenowania modelu. W projekcie wykorzystany zostanie taki model, którego struktura da najlepsze rezultaty.

Oba modele wykorzystywane zostaną we wspólnej aplikacji z interfejsem użytkownika. Do przygotowania interfejsu wykorzystana zostanie biblioteka **Qt**.

Wszelkie źródła projektowe zawarte są w wspólnym repozytorium **GitHub**.



Rysunek 1: Przykład danych

2.2 Wyodrębnianie twarzy ze zdjęć

2.2.1 Wybór typu sieci

2.2.2 Przegląd gotowych rozwiązań

2.2.3 Opis danych treningowych

Dane treningowe do wykrywania twarzy na zdjęciach zostaną użyte ze zbioru **WIDER FACE**. Jest to ogólnodostępny zbiór 32,203 zdjęć, na których znajduje się 393,703 twarzy. Są one wysoce zróżnicowane ze względu na rozmiar, a twarze znajdują się w różnych okolicznościach, pozycjach i rozmiarach. Cały dataset jest podzielony na dane treningowe, walidacyjne oraz testowe w proporcjach 40:10:50. Jednakże dane testowe są przeznaczone do sprawdzania funkcjonalności rozwiązań przez autorów zbioru przez co nie zostały udostępnione dane dotyczące informacji o pozycjach twarzy na tych zdjęciach. Zatem zbiór danych, które będą mogły zostać wykorzystane do testowania autorskiej sieci znacznie maleje.

W kontekście wykrywania twarzy pojawia się zagadnienie negatywnych próbek, czyli otoczenia twarzy, aby móc lepiej wyodrębnić pozycje twarzy. Oznacza to, że będziemy musieli przygotować z dostępnych zdjęć wiele próbek, które będą automatycznie wycinane z tego samego zbioru.

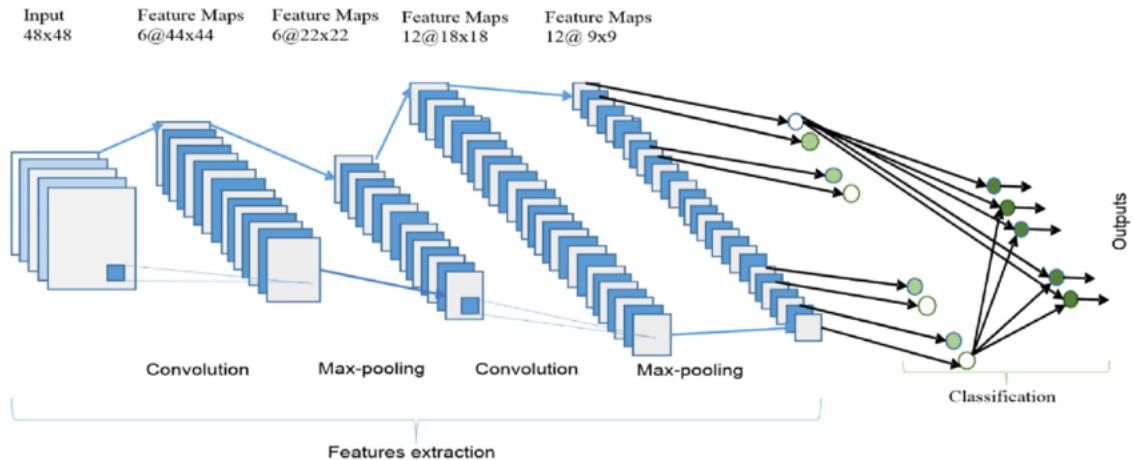
2.2.4 Proces uczenia

2.2.5 Ocena efektów uczenia

2.3 Rozpoznanie emocji

2.3.1 Wybór typu sieci

Planujemy wykorzystać konwolucyjną sieć neuronową do rozpoznawania emocji. Architektura sieci będzie się składać przed wszystkim z: convolutional layer, pooling layers oraz dropout layers. Jako funkcje aktywacyjne do warstw konwolucyjnych wykorzystamy funkcje ReLU ,



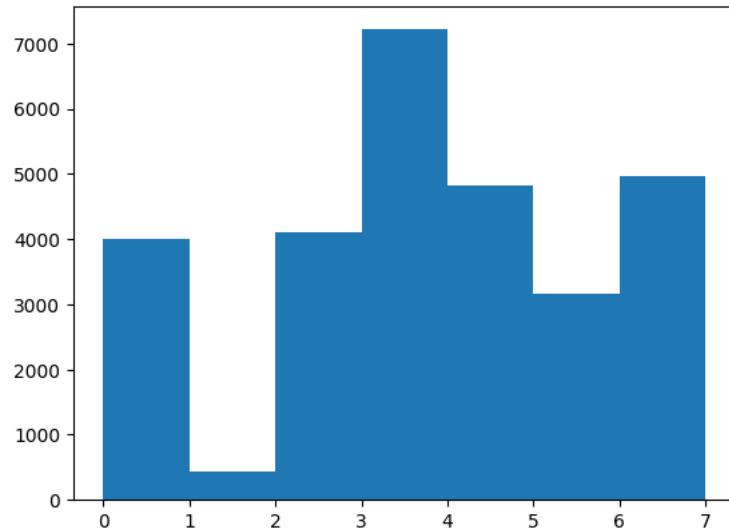
Przykładowa architekta sieci wyglądać następująco (Tensorflow):

```
model = tf.keras.Sequential([
    tf.keras.layers.Conv2D(32, (3,3), activation='relu', input_shape=(48, 48, 1)),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D((2,2)),
    tf.keras.layers.Conv2D(64, (3,3), activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D((2,2)),
    tf.keras.layers.Conv2D(128, (3,3), activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D((2,2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(7, activation='softmax')
])
```

2.3.2 Opis danych treningowych

Dane treningowe pochodzą ze zbioru FER-2013, który był częścią zadania konkursowego zamieszczonego na platformie *kaggle* w 2013 roku. Zbiór treningowy składa się z 28709 zdjęć czarno-białych wymiaru 48×48 przedstawiających twarze ludzkie wyrażających jedną z

sześciu emocji: 0=złość, 1=zniesmaczenie, 2=strach, 3=szczęście, 4=smutek, 5=zaskoczenie, 6=neutralność. Na poniższym wykresie przedstawiono liczęność wymienionych grup.



Dane są zawarte w pliku `train.csv`. Pierwsza kolumna zawiera liczby z zakresu od 0 do 6 włącznie, które odpowiadają emocji przedstawianej na danym zdjęciu. Kolumna druga zawiera napis otoczony cudzysłowami dla każdego zdjęcia. Napis ten zawiera rozdzielone spacjami wartości pikseli zapisane "wiersz po wierszu". Przykładowe zdjęcia ze zbioru danych (happy i angry).



2.3.3 Proces uczenia

2.3.4 Ocena efektów uczenia

2.4 Aplikacja użytkownika – GUI

2.5 Podsumowanie

3 Literatura

- https://www.ripublication.com/ijaer18/ijaerv13n8_119.pdf