Uniwersytet Mikołaja Kopernika w Toruniu

Wydział Matematyki i Informatyki

Szymon Nowacki

nr albumu: 316259

Niestacjonarne Studia Informatyczne

Praca inżynierska

Szablon pracy dyplomowej

Promotor

Dr Marta Burzańska

Toruń 2024

# Wstęp

## 1.1 Wprowadzenie

Sztuczna inteligencja w ciągu ostatnich lat zyskuje coraz większe znaczenie jako dziedzina nauki i technologii. Jej wpływ obejmuje coraz szersze aspekty życia, lecz na szczególną uwagę zasługują te z nich, które intuicyjnie wydawały się nieosiągalne dla algorytmów i komputerów. Jednym z takich zastosowań jest analiza ludzkich emocji na podstawie samych tylko obrazów twarzy. Innowacje osiągnięte w dziedzinie masowego i błyskawicznego wykrywania stanów emocjonalnych bez potrzeby ingerencji innego człowieka mogą doprowadzić do usprawnień w wielu dziedzinach życia opartych na osobistych odczuciach.

Obecnie, detekcja emocji znajduje zastosowanie w wielu dziedzinach. W marketingu pozwala na personalizację kampanii reklamowych, znacząco poszerzając zakres informacji zwrotnych jakie reklamodawca może otrzymać od odbiorców. W opiece zdrowotnej obecność takiej technologii może umożliwić dokładniejszą diagnozę i terapię zaburzeń emocjonalnych. Dostęp do tak ludzkiego aspektu życia ma możliwość poskutkować znacznym postępem w relacjach człowiek-komputer, dając asystentom wirtualnym dostęp do całkowicie nowego sposobu na rozumienie swoich użytkowników.

1.1.1 Podstawy teoretyczne

#### Emocje i ich cechy charakterystyczne

Emocje to reakcje psychofizyczne mające odzwierciedlać ludzkie przeżycia, są podstawą interakcji międzyludzkich. Każda z nich wywołuje specyficzne zmiany w ciele, zachowaniu oraz co najważniejsze w ekspresji mimicznej twarzy. Można je charakteryzować na podstawie intensywności, trwania oraz kontekstu pojawienia się[[1]](#endnote-1). Te cechy umożliwiają klasyfikację emocji, istotny fakt zarówno w psychologii, jak i w podejściach technologicznych.

Ekspresje emocji

Ekspresje emocji to uniwersalny język, którym ludzie komunikują wewnętrzne stany emocjonalne. Paul Ekman zidentyfikował 6 podstawowych emocji, które są powszechne we wszystkich odkrytych kulturach świata. Każdą z nich można scharakteryzować przez wzorce ruchów mięśni twarzy[[2]](#endnote-2):

* **Radość:** Wyrażana przez uniesienie kącików ust, często towarzyszy jej uśmiech oraz zmarszczenie skóry wokół oczu.
* **Smutek**: Przejawia się opuszczenie kącików ust i lekkim uniesieniem wewnętrznych krawędzi brwi. Oczy mogą być lekko zwężone albo zamglone.
* **Złość**: Objawia się ściągniętymi brwiami, co powoduje pojawienie się między nimi zmarszczek. Pojawia się napięcie wokół ust, mogą być zaciśnięte lub przypominać cienką linię.
* **Strach**: Rozpoznawany przez uniesienie brwi, szeroko otwarte oczy oraz otwarte usta.
* **Zaskoczenie**: Charakteryzuje się uniesionymi brwiami, które tworzą zmarszczki na czole, szeroko otwartymi oczami. Usta często tworzą kształt litery „O”.
* **Wstręt**: Wyrażany przez uniesienie górnej wargi, marszczeniem nosa i opuszczeniem brwi.

#### Facial Emotion Recognition (FER)

Rozpoznawanie emocji na podstawie mimiki twarzy to technologia, która potrafi klasyfikować emocje z pomocą uczenia maszynowego i wizji komputerowej (Computer Vision, CV). Proces obejmuje detekcję twarzy, ekstrakcję cech, klasyfikację i opcjonalnie interpretację wyniku.

#### Sieci Neuronowe

Sieci neuronowe to struktury obliczeniowe inspirowane biologicznym mózgiem będące podstawą współczesnych algorytmów sztucznej inteligencji. Składają się z warstw sztucznych neuronów, na których strukturę i działanie składa się:

* **Wejście**: reprezentowanego jako wektor liczb
* **Wagi**: każde z wejść posiada swoją, jest to miara znaczenia tego połączenia dla podejmowanej decyzji. Wagi mogą być dodatnie, ujemne lub zerowe. Są dostosowywane podczas procesu uczenia, a ich modyfikacja jest środkiem do minimalizowania błędu pomiędzy oczekiwanym a rzeczywistym wyjściem neuronu.
* **Suma ważona**: neuron oblicza sumę ważoną swoich wejść. Wynik tej operacji nazywa się aktywacją neuronu.
* **Funkcje aktywacji**: Służy do przetworzenia wartości aktywacji neuronu, decyduje o tym jaki sygnał zostanie przekazany dalej.
* **Wyjście**: wartość, która jest rezultatem zastosowania funkcji aktywacji do sumy ważonej.

Każdy neuron jest połączony z innymi w warstwie poprzedniej i następnej z pomocą wag. Celem sieci jest zminimalizowanie różnicy między przewidywaniem modelu (oznaczanym jako y) a rzeczywistym wynikiem (oznaczanym jako ŷ). Sieci neuronowe adaptują różne architektury, od jedno lub dwuwarstwowych perceptronów do Konwolucyjnych i Rekurencyjnych sieci zaprojektowanych do przetwarzania obrazów, mowy czy pisma. Te struktury są szczególnie skuteczne w wyszukiwaniu wzorców w dużych zbiorach danych[[3]](#endnote-3).

## 1.2 Cel pracy

Celem niniejszej pracy jest opracowanie modelu AI zdolnego do skutecznego wykrywania emocji na podstawie analiz obrazów twarzy. By to osiągnąć zostanie przeprowadzona analiza istniejących metod, która posłuży stworzeniu nowego modelu opartego. Rozwiązanie zostanie poddane szeregowi testów efektywności i wydajności. Ostatecznie, pozwoli to też na identyfikację wyzwań związanych z implementacją takiego modelu w rozwiązaniach komercyjnych.

## 1.3 Zakres pracy

Niniejsza praca inżynierska koncentruje się na analizie, zaprojektowaniu, implementacji i ewaluacji modelu sztucznej inteligencji zdolnego do rozpoznania 6 podstawowych emocji: radości, smutku, strachu, złości, zaskoczenia i wstrętu [[4]](#endnote-4) na podstawie obrazów twarzy. Projekt ten koncentruje się na:

1. **Przegląd literatury i standardów rynkowych:** W pierwszej kolejności zostanie przeprowadzony przegląd badań, analiz oraz rozwiązań w dziedzinie rozpoznawania emocji za pomocą sztucznej inteligencji. Ta część ma na celu identyfikację zalet i wad stosowanych obecnie podejść oraz potencjalnie znalezienia luki w analizowanych źródłach, którą może wypełnić proponowane rozwiązanie.
2. **Projektowanie modelu:** Z użyciem przyswojonych wcześniej podstaw teoretycznych zostanie zaprojektowany nowy model. Na drugą część opracowania złożą się: wybór odpowiedniej architektury sieci neuronowej, definicji warstw, algorytmów optymalizacji i funkcji aktywacji. Kandydaci na docelowe rozwiązanie będą trenowani na małym zestawie danych w celu wyłonienia najbardziej obiecujących kombinacji cech modelu. Każdy znaczący wybór zostanie odpowiednio uzasadniony, a ciekawsze odrzucone przypadki odpowiednio udokumentowane.
3. **Implementacja modelu:** Zaprojektowany model zostanie stworzony, wytrenowany i wdrożony. Kluczowymi narzędziami, które pozwolą zaprojektować wytypowane rozwiązanie są: Python (główny język programowania), TensorFlow (główna biblioteka umożliwiająca tworzenie i trenowanie modeli głębokiego uczenia), Keras (wysokopoziomowe API do TensorFlow) i OpenCV (przetwarzanie i augmentacja obrazów, wykrywanie twarzy). Na tym etapie nastąpi też selekcja i ewentualna modyfikacja zbioru danych służących do treningu i testowania sieci.
4. **Testowanie i analiza:** Zaimplementowany model zostanie poddany szeregowi testów, które dostarczą informacje niezbędne do oceny jego skuteczności i wydajności. Podstawowymi miarami do oceny działania będą: dokładność (accuracy), precyzja (precision), czułość (recall), czas odpowiedzi i zużycie zasobów. W przypadku niewystarczalności analiz ilościowych (specyficzny wzorzec popełnianych błędów, trudne przypadki graniczne) dodatkowo zostaną przeprowadzone wszelkie potrzebne analizy jakościowe. Zwieńczeniem tego etapu prac będzie ustandaryzowany raport prezentujący wspomniane metryki.
5. **Identyfikacja wyzwań i możliwości:** Zwieńczenie pracy będzie zawierać opis napotkanych problemów i wyzwań związanych z komercyjnym użyciem modelu. Możliwości stworzonego narzędzia zostaną omówione od strony zarówno technicznej jak i etycznej/prawnej. Ponadto zostaną przedstawione możliwe kierunki rozwoju modelu z podziałem na potencjalne obszary w których stworzone rozwiązanie dostarczyłoby najwięcej korzyści.

## 1.4 Przegląd istniejących rozwiązań

### 1.4.1 Klasyczne metody rozpoznawania

#### *Metody oparte na analizie cech geometrycznych twarzy*

Wymaga identyfikacji kluczowych punktów, na przykład krawędzi ust, kącików oczu, linii brwi. Na ich podstawie algorytm tworzy zestaw cech geometrycznych, zawierający kąty i odległości pomiędzy kluczowymi elementami, co pozwala na klasyfikację emocji. Do konkretnych algorytmów wpisujących się w taki paradygmat działania zaliczamy Active Shape Models (ASM) [[5]](#endnote-5), i jego rozszerzenie – Active Appearance Models (AAM) [[6]](#endnote-6). AAM nie bazuje wyłącznie na kształcie, lecz także na teksturze twarzy, co umożliwia dokładniejszą analizę poprzez uwzględnienie wyglądu skóry twarzy.

#### *Metody oparte na klasyfikatorach*

Nie ograniczają się jedynie do cech geometrycznych, lecz mogą wykorzystać szerokie spektrum cech. Pozwala to na tworzenie bardziej ogólnych modeli rozpoznawania. W połączeniu z metodami ekstrakcji cech, takimi jak Local Binary Pattern (LBP) czy Histogram of Oriented Diagrams (HOG) metody te okazują się efektywniejsze w przypadku danych o wysokiej zmienności. Najbardziej sprawdzonym przykładem stosowanym w rozpoznawaniu emocji na podstawie twarzy w tej kategorii rozwiązań jest Support Vector Machines (SVM) [[7]](#endnote-7). W procesie klasyfikacji SVM tworzy hiperpowierzchnię w przestrzeni cech. Jej zadaniem jest oddzielić klasy przynależne do poszczególnych emocji, z czym radzi sobie zarówno w przestrzeni liniowej jak i nieliniowej. Pozwala to na lepszą klasyfikację w przypadkach, gdzie granice pomiędzy emocjami nie są proste do wyznaczenia (różnice w naświetleniu, kącie kamery).

### 1.4.2 Nowoczesne metody oparte na głębokim uczeniu

#### Głębokie uczenie

Głębokie uczenie, będące poddziedziną uczenia maszynowego, opiera się na użyciu wielowarstwowych sieci neuronowych. Głównym usprawnieniem w stosunku do tradycyjnych algorytmów jest automatyczna identyfikacja szukanych wzorców na różnych poziomach abstrakcji. W procesie propagacji wstecznej wagi należące do poszczególnych neuronów są aktualizowane tak, by minimalizować błąd pomiędzy oczekiwanym a rzeczywistym (dostarczonym przez dane treningowe) wynikiem.

#### Convolutional Neural Networks (CNN)

Najczęściej stosowana architektura głębokiego uczenia w przetwarzaniu obrazów. Potrafi automatycznie wyodrębnić

[https://www.researchgate.net/publication/374967891\_Algorithms\_used\_for\_facial\_emotion\_recognition\_a\_systematic\_review\_of\_the\_literature 2.8.2](https://www.researchgate.net/publication/374967891_Algorithms_used_for_facial_emotion_recognition_a_systematic_review_of_the_literature%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%202.8.2)

<http://neuralnetworksanddeeplearning.com/chap1.html>

https://www.deeplearningbook.org/

convolutional layer – first, main building block, weights

clustering layer – no weights, a lot of information is lost but it helps reduce complexity, improves efficiency and limits risk of overfitting

fully connected layer – last, ???

increasing in complexity by each layer (more abstract). First layers focuses on simple features, each next can recognize more complex patterns

filter?

Feature map?

Subsampling? – layer clustering, reducing number of input parameters, maximum and average clustering

### 1.4.2 Rekurencyjne Sieci Neuronowe (RNN)

### 1.4.3 Uczenie Transferowe

### 1.4.4 Modele hybrydowe

1. Plutchik, Robert. Emotion: A Psychoevolutionary Synthesis. Harper & Row, 1980. [↑](#endnote-ref-1)
2. Ekman, Paul. Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. 2nd ed., Henry Holt and Co., 2007. [↑](#endnote-ref-2)
3. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. [↑](#endnote-ref-3)
4. Ekman, Paul. "An Argument for Basic Emotions." Cognition & Emotion, vol. 6, no. 3-4, 1992, pp. 169-200. [↑](#endnote-ref-4)
5. Cootes, Timothy F., et al. "Active Shape Models—Their Training and Application." Computer Vision and Image Understanding, vol. 61, no. 1, 1995, pp. 38-59, doi:10.1006/cviu.1995.1004 [↑](#endnote-ref-5)
6. Cootes, Timothy F., et al. "Active Appearance Models." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, 2001, pp. 681-685, doi:10.1109/34.927467. [↑](#endnote-ref-6)
7. Bartlett, Marian Stewart, et al. "Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 568-573, doi:10.1109/CVPR.2005.297. [↑](#endnote-ref-7)