

Učenje izražanja čustev s pomočjo detekcije obraza

Jernej Vrhunc¹[63150316] and Žan Pečovnik¹[63160406]

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, Slovenija

Povzetek. Izkazovanje in prepoznavanje čustev je za ljudi precej enostavna naloga, medtem ko je za računalnike to izjemno zahtevna naloga. Ena izmed bolj uspešnih metod je uporaba konvolucijskih nevronske mreže na sivinskih slikah obrazov ljudi. V programskem jeziku Python s pomočjo knjižnic OpenCV, MediaPipe in implementacije konvolucijske nevronske mreže Real-Time Multimodal Emotion Recognition je bila narejena aplikacija, ki je zmožna zaznati tako človeški obraz kot človeške roke, hkrati pa je zmožna še iz obraznih potez zaznati čustva. Igrifikacija aplikacije omogoča zbiranje točk, če uporabnik izrazi in izbere pravilno čustvo, kar omogoča učenje boljšega izražanja in prepoznavanja čustev.

Ključne besede: Detekcija obraza · Detekcija rok · Detekcija čustev

1 Uvod in motivacija

Obrazne poteze igrajo ogromno vlogo pri izkazovanju in prepoznavanju čustev, študije pa so pokazale, da lahko pravilno prepoznana čustva bistveno spremenijo potek pogovora in razumevanje izgovorjenih besed. Izkazovanje čustev se razlikuje od posameznika do posameznika, čustva na sploh pa so zelo subjektivne in spontane narave, torej jih je zelo težko načrtovati, težko je vnaprej napovedati kaj bo posameznik čutil v določeni situaciji. To pomeni, da jih je tudi zelo težko pravilno prepoznati v tistem trenutku, ljudje so v tem zelo dobri, ker lahko prepoznajo še kontekst, zgodovino in druge dejavnike, ki bi lahko vplivali na čutenje, medtem ko je za računalnike to izjemno zahtevna naloga.

V tej seminarski nalogi sva si zadala cilj, da bi ustvarila aplikacijo, ki bi ljudem pomagala pri izražanju čustev in bi jih učila, kako jih izraziti na način, da jih bo čim lažje prepoznati. Zaradi situacije z virusom Covid-19 ogromno časa preživimo z masko, ki nam prekrije skoraj pol obraza, tako da je še težje prepoznati čustva in smo vsi že malce pozabili, kako izgledajo čustva na obrazu drugih ljudi.

Aplikacija lahko v realnem času zazna posameznikov obraz preko kamere na računalniku ter prepozna čustvo, ki ga v tistem trenutku izraža. Na zgornjem delu zaslona se nahajajo slike, ki simbolizirajo 7 različnih čustev, in če uporabnik izbere sliko, ki se ujema z njegovim zaznanim čustvom, si prisluži točko. Na takšen način sva poskusila z igrifikacijo aplikacijo še dodatno narediti bolj zanimivo in jo približati tako otrokom kot odraslim.

2 Podobni projekti

Ker so čustva subjektivne narave in se tako njihovo doživljanje kot izražanje razlikuje od človeka do človeka jih je Prudhvi Raj Dachapally [4] poskušal zaznati iz slik na dva različna načina. Prvi način je z reprezentativnimi avtokodirnimi enotami (RAU), s pomočjo katerih zgradi unikatno predstavitev vsakega čustva, drugi način pa je s konvolucijskimi nevronskimi mrežami z 8 nivoji. Obe metodi sta bili naučeni na podatkovni bazi JAFFE, ki vsebuje 215 slik ženskih obrazov, ki izražajo 7 različnih čustev in dodatno stestirani še na 100 naključnih slikah podatkovne baze LFW. Izkazalo se je, da bi konvolucijska nevronska mreža s še malo dodatnega učenja in fine-tuninga dosegla boljše rezultate kot takratni state-of-the-art način za prepoznavo čustev.

Izražanje čustev je večinoma spontane narave, ne pa posledica premišljenega truda in dejanj, kar pa se najbolj opazi v spremembi obraznih mišic oz. izrazov, ki jih le-te povzročijo, in imajo vidno vlogo v neverbalni komunikaciji. Ravno zaradi tega in tudi zaradi zgoraj opisanega razloga je zelo težko enolično razbrati čustva iz slike. S pomočjo konvolucijskih nevronskih mrež, modelov za globoko učenje in še drugih modelov za strojno učenje so se Forhad Ali, Khatun in Turzo [5] lotili problema prepoznav 7 različnih čustev ter bili pri tem zelo uspešni, skoraj tako zelo kot je uspešen pri tem človek.

3 Uporabljene tehnologije

3.1 Python

Za programski jezik sva si izbrala Python, verzijo 3.8.0, ker je učinkovit in hkrati enostaven jezik, ki vsebuje vse knjižnice, ki so potrebne za zaznavo rok, obraza in strojno učenje oz. za uporabo nevronskih mrež.

3.2 OpenCV

OpenCV [1] (Open Source Computer Vision Library) je odprtokodna knjižnica, ki vključuje več sto algoritmov računalniškega vida in je v osnovi napisana v programskem jeziku C++, vendar je s pomočjo API-ja na voljo tudi za druge programske jezike, med drugim tudi za Python. Ima modularno strukturo, kar pomeni, da moduli vključujejo več deljenih oz. statičnih knjižnic. Vključuje naslednje module:

- Core functionality - definira vse osnovne podatkovne strukture in osnovne funkcije, ki jih uporabljajo še drugi moduli
- Image processing - vključuje vse linearne in nelinearne filtre za filtriranje slik, geometrične transformacije slik, konverzijo med barvnimi prostori, histograme, itd.
- Video analysis - vključuje ocenjevanje poti gibanja, odstranjevanje ozadja, sledenje objektom
- Object detection - detekcija objektov in instanc definiranih razredov (oči, obraz, ljudje, avtomobili, ipd.)
- obstaja še nekaj drugih bolj in manj pomembnih modulov

3.3 MediaPipe

MediaPipe [2] ponuja odprtokodno, prilagodljivo ogrodje z algoritmi strojnega učenja za uporabo pri živih in pretočnih vsebinah. Na voljo je za več platform (Android, iOS, Python, Javascript). Knjižnica ponuja ogromno rešitev, npr.

- detekcija 33 ključnih točk, potrebnih za sledenje človeški silhueti
- razpoznavo kar 468 ključnih točk na človeškem obrazu
- razpoznavo 21 ključnih točk, potrebnih za sledenje rokam
- detekcija in sledenje objektom
- detekcija obraza
- detekcija las

3.4 Zaznavanje in detekcija čustev - Real-Time Multimodal Emotion Recognition

Za prepoznavo in detekcijo emocij smo uporabili implementacijo Real-Time Multimodal Emotion Recognition [3], ki z uporabo številnih knjižnic in orodij kot so Python, Keras, Tensorflow in drugi, preučuje prepoznavanje in obdelavo človeških vplivov. S pomočjo FER2013 Kaggle podatkovne množice [6], katere podatki so sestavljeni iz 48 x 48 pikslov sivinskih slik obrazov, izdelajo model, ki je sposoben v realnem času izvesti in uporabniku sporočiti prepoznano čustva na človeškem obrazu.

Knjižnica omogoča detekcijo čustev preko zvočnih posnetkov, vpisovanja besedila in prepoznavo na obrazu preko video posnetka oz. vhoda spletne kamere.

Čustva, ki jih je knjižnjic sposobna prepoznati so naslednje:

- veselje,
- jeza,
- žalost,
- strah,
- gnus,
- prenesečenost in
- nevtrarno počutje.

4 Opis delovanja aplikacije

Delovanje naše aplikacije je sestavljeno iz treh faz:

- prva faza, ki vsebuje zajemanje in prepoznavanje čustva, ter uporabnikovo izbiro, katero čustvo je želel doseči 4.1,
- druga faza, ki predstavlja fazo točkovanja in prikazovanja informacij uporabniku o njegovi uspešnosti 4.2, ter
- tretja faza, kjer lahko uporabnik z uporabo detekcije roke odpira pridobljene informacije skrite pod kvadratki na zaslonu, ali le-te kvadratke predstavlja 4.3.

4.1 Izbiranje in prepoznavanje čustva uporabnika

Ko se aplikacija zažene, lahko uporabnik na zaslonu vidi 5 identičnih kvadratov, pod katerimi se skrivajo navodila za uporabo aplikacije. Ko se uporabnik z odprto dlanjo premakne nad vsakega izmed kvadratov, se mu pod njim prikaže eno od navodil za uporabo aplikacije. Ob končanem pregledu navodil, uporabnik začne uporabljati/igrati igro čustev.

V zgornji vrstici ekrana ima na voljo 7 različnih čustev: veselje, jeza, žalost, presenečanje, strah, gnus in nevtralen obraz. Uporabnik si izbere, katerega izmed čustev želi izraziti, se z miško prestavi na slikovno ikono, ki predstavlja to čustvo, pogleda v kamero in s klikom miške na ikono zajame trenutno čustvo, ki ga izraža.

Na sliki 4.1 vidimo, kako se lahko uporabnik z dlanjo sprehodi preko kvadratkov, pod katerimi se skrivajo navodila za uporabo aplikacije.

Sledi faza točkovanja uporabnikove natančnosti.

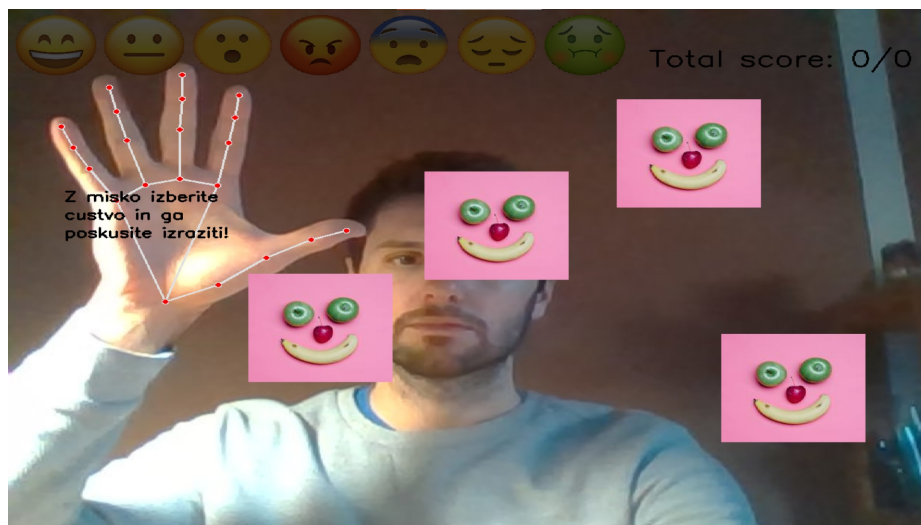


Fig. 1. Prikazovanje navodil

4.2 Točkovanje

V fazi točkovanja, se uporabniku, v kolikor je izbrano čustvo uspešno izrazil prišteje ena točka oz. se mu število pridobljenih točk ne spremeni, če čustva ni izrazil uspešno. V skrajno desnem kotu zaslona, lahko uporabnik vidi število prejetih točk v primerjavi z vsemi poskusi, ki jih je opravil.

Da uporabnik sistem lažje razume, se mu ikona čustva, ki jo je izbral v prejšnji fazi osvetli, prav tako pa se mu nad skupnim rezultatom v zgornjem desnem kotu prikaže dejanska vrednost čustva, ki je bila prebrana iz njegovega obraza.

Kot vidimo na slikah 2 in 3 se v desnem zgornjem kotu glede na zadetek ali pogršek pri izražanju čustva označi število pravih oz. napačnih odgovorov.

Sledi faza prikazovanja informacij, glede na rezultat izražanja čustva.



Fig. 2. Primer zadetka čustva

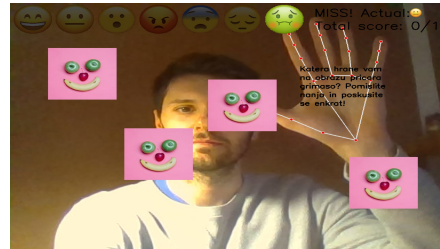


Fig. 3. Primer pogrška čustva

4.3 Prikazovanje informacij uporabniku

Kot smo omenili že v poglavju 4.1, so uporabniku na zaslonu prikazani kvadrati pod katerimi se skrivajo sprva navodila, a se le-ta po izvedbi izbire čustva in njegovem točkovanju spremenijo. V kolikor je bil uporabnik pri izražanju čustva uspešen, se mu glede na izbrano in izraženo čustvo (tj. strah, jeza, veselje, nevratnost, gnus, presenečenost ali žalost) pod kvadrati izpišejo dobre misli, pohvale in drugi napisi, ki mu vlivajo samozavest in ga spodbudijo k nadaljnjemu poskušanju. V kolikor pa je bil uporabnik pri izražanju določenega čustva neuspešen, mu napisi pod kvadratki pomagajo, kako naj do danega čustva pride, ali pa ga spodbudijo k pravemu razmišljanju za doseganje določenega čustva. Uporabnik lahko kvadratke ponovno (kot pri koraku izbire čustva - branje navodil) pregleda tako, da se z odprto dlanjo premika preko kvadratov in prebere misel, ki ga čaka pod njim.

Na slikah 2 in 3 vidimo, kako uporabnik z odprto dlanjo prikazuje informacije, ki se prikažejo glede na uspešnost njegovega izražanja čustva.

Da je uporabniku lažje, si lahko kvadratke po zaslonu razvrsti kakor to njemu ustreza. To naredi tako, da se z odprto dlanjo prestavi nad kvadrat, nad njim naredi pest in začne zaprto pest premikati. Kvadrat se nato dokler ima uporabnik dlan zaprto v pest premika z njim. Ko uporabnik dlan razpre, se kvadrat preneha premikati in ostane na mestu, kamor si ga je uporabnik želel postaviti.

4.4 Ponovitev celotnega procesa

Ko uporabnik skrbno prebere napotke, ki se skrivajo pod kvadrati se s pritiskom tipke 'R' ponovno vrne na korak izbiranja čustva 4.1 in tako se cikel ponovi od začetka.

V kolikor želi uporabnik program zapustiti, mu je ves čas na voljo tipka 'Q', katera ga popelje na izhod programa.

5 Rezultati

Preizkus delovanja aplikacije oz. natančnosti zaznave čustev sva testirala vsak z 5 x 10 poskusi zaznave. Rezultati so zadovoljivi, ampak niso popolnoma realni, saj je zaznava čustev odvisna od veliko drugih dejavnikov, kot je osvetlitev prostora, zmožnost posameznika za igranje čustev, kvaliteta kamere, itd.

poskus	Žan	Jernej
1.	6	7
2.	5	6
3.	6	6
4.	7	8
5.	7	7

Table 1. Število zadetkov v posameznem poskusu

Kot je razvidno iz zgornje tabele, sva opravila 100 poskusov zaznave čustev, uspešno pa jih je aplikacija prepoznala 65.

Literatura

1. <https://github.com/opencv/opencv-python>, dostopano dne 9.1.2022
2. <https://github.com/google/mediapipe>, dostopano dne 9.1.2022
3. https://github.com/maelfabien/Multimodal-Emotion-Recognition/tree/master/03-Video?fbclid=IwAR18adsg7RhrTbguNLdmUqT6_fNgH0J0yi4adUOXuHQSt0HkcuZFqQPm058, dostopano dne 9.1.2022
4. Dachapally, P. R. (2017). Facial emotion detection using convolutional neural networks and representational autoencoder units. arXiv preprint arXiv:1706.01509.
5. Ali, M. F., Khatun, M., Turzo, N. A. (2020). Facial Emotion Detection Using Neural Network. the international journal of scientific and engineering research.
6. <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge>, dostopano dne 9.1.2022