

Recunoașterea emoțiilor din expresii faciale folosind rețele neuronale convoluționale

Cătălina Beșliu, Cristian Monor

Calculatoare Încorporate, Facultatea de Automatica și Calculatoare Iași

Abstract: Recunoașterea automată a emoțiilor bazată pe expresia facială este un domeniu de cercetare aplicat în domenii precum securitate, sănătate și interfețe om-calculator. Scopul acestei lucrări este de a crea și antrena un model de rețea neuronală, găsirea parametrilor optimi pentru antrenarea acestuia, dorind să obținem o acuratețe mai bună decât în alte lucrări din domeniu.

Index Terms: Analiza imaginilor, expresii faciale, rețele neuronale.

1. Introducere

În cadrul recunoașterii automate a emoțiilor sunt abordate două subiecte: recunoașterea psihologică a emoțiilor umane și inteligența artificială. Starea emoțională a oamenilor poate fi obținută din informații verbale și non-verbale. De exemplu, expresii faciale, tonul vocii și semne fiziologice. În timpul comunicării, expresiile faciale sunt primele semne care transmit starea emoțională a participanților. Acestea pot fi interpretate pentru a oferi informații precum comportamentul, starea mentală, personalitatea și tendința criminală.

Spre deosebire de un observator uman, inteligența artificială are dificultate în clasificarea emoțiilor datorită variațiilor trăsăturilor faciale cauzate de gen, vârstă, rasă, calitatea imaginii.

Folosind o rețea neuronală convoluțională, se extrag trăsăturile faciale pe baza cărora imaginile se încadrează în șapte clase de emoții: enervat, dezgustat, înfricoșat, fericit, neutru, trist, surprins.

2. Structura rețelei neuronale convoluționale

Rețelele neuronale convoluționale sunt o clasă de rețele care pot recunoaște și clasifica anumite caracteristici ale imaginilor. Acestea sunt clasificatori concepuți pentru lucrul cu imagini deoarece permit clasificarea datelor cu un număr mare de parametri și are loc prelucrarea eficientă a acestor categorii de date. Sunt alcătuite din două componente: componenta de extragere a trăsăturilor și componenta complet conectată.

În această lucrare au fost folosite două tipuri de rețele: cu 3 straturi de convoluție și 2 straturi de agregare și 4 straturi de convoluție și 3 straturi de agregare.

2.1. Componenta de extragere a trăsăturilor

Aceasta se compune din mai multe straturi:

- Straturi de convoluție - are loc convoluția dintre imaginea de intrare și o matrice, în urma căreia se obțin una sau mai multe trăsături.
- Straturi de agregare - se realizează o reducere a dimensiunii imaginii primite din pasul anterior, ce se poate realiza prin mai multe metode: Max, Sum, Average Pooling.
- Funcție de activare - aplicată straturilor de convoluție și agregare.

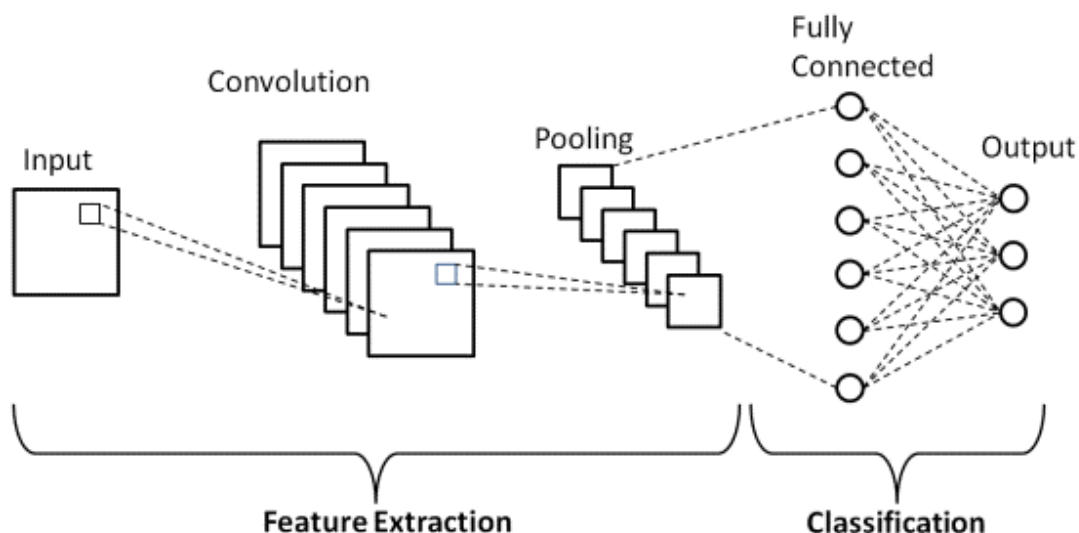


Fig. 1. Structura unei rețele neuronale convoluționale

2.1.1. Componenta complet conectată

În cadrul acesteia se realizează clasificarea propriu-zisă. Ieșirea componentei este un vector de probabilități, ce reprezintă apartenența procentuală a imaginii de intrare în clase.

3. Lucrări existente

Pentru îmbunătățirea acurateței rețelei antrenate, un factor important îl reprezintă antrenarea rețelei cu cât mai multe exemple de imagini având variații de poziționare a feței și luminozitate. În cadrul lucrării a fost folosit setul de date FER2013 ce conține 35887 imagini grayscale colectate din Google Image Search. Acest set de date a mai fost folosit în cadrul antrenării mai multor proiecte ce aveau drept scop obținerea unei detecții cât mai precisă în urma automatizării procesului de recunoaștere a emoțiilor.

În 2019, Agrawal and Mittal (2020) au încercat să obțină o acuratețe cât mai mare a modelului antrenat folosind setul de date FER2013, prin variația parametrilor rețelei neuronale convoluționale. Au schimbat dimensiunile imaginilor, numărul de filtre aplicate precum și optimizatorul folosit pe o rețea neuronală convoluțională simplă ce constă din două straturi de convoluție succesive, un strat de agregare și o funcție softmax pentru clasificare. În urma încercărilor, acuratețea obținută a modelelor a fost de 65.23% și 65.77%. Particularitatea acestor modele este că nu conțin straturi complet conectate.

Mollahosseini et al. (2016) au dezvoltat un model de rețea neuronală ce l-au propus pentru a fi folosit pe mai multe seturi de date. După extragerea trăsăturilor faciale, imaginile erau reduse la dimensiunea de 48 x 48 pixeli și erau supuse tehnicilor de augmentare a datelor. Arhitectura modelului constă din două straturi de convoluție-agregare urmat de două module "Inception" ce permite îmbunătățirea locală a performanței prin reducerea problemelor de "over-fitting". Acuratețea modelului a fost de 61.1% pe setul de date FER2013.

4. Rezultate obținute

Cele mai bune rezultate obținute au fost trecute în tabelul de mai jos. Pentru obținerea rezultatelor menționate au fost antrenate mai multe modele de rețele neuronale cu mici variații a parametrilor rețelei.

În cadrul modelului 1 s-au folosit 4 nivele de straturi convoluționale-agregare, 3 nivele complet

Nr model	Nr epoci	Acuratețe obținută	Eroare
1	30	0.5838	1.0593
1	100	0.6269	1.0294
2	30	0.5911	1.0226
3	100	0.6722	0.8722
4	100	0.7230	0.7484

TABLE I
REZULTATELE OBTINUTE ÎN URMA ANTRENĂRII

conectate și 3 etape de "dropout" pentru a reduce problema de "overfitting". Spre deosebire de modelul 1, în cadrul modelului 2 au fost folosite doar 2 nivele complet conectate. Modelul 3 conține 4 straturi convoluționale-agregare cu o singură etapă de dropout. Modelul 4 constă din 3 straturi convoluționale-agregare cu o singură etapă de dropout.



Fig. 2. Rezultate obținute în urma aplicării modelului 1

5. Concluzii

În concluzie, adăugarea straturilor convoluționale adiționale și complet conectate a îmbunătățit acuratețea rețelei neuronale cu aproximativ 2%. Această îmbunătățire are costul timpului crescut de antrenare.

References

- Agrawal, A. and Mittal, N. (2020). Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2):405–412.
- Mollahosseini, A., Chan, D., and Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. In *2016 IEEE Winter conference on applications of computer vision (WACV)*, pages 1–10. IEEE.