

# Recunoasterea Emotiilor si a Genului

Laura Mitrache

## Cuprins

<b>0</b>	<b>Introducere</b>	<b>1</b>
<b>1</b>	<b>Preliminarii</b>	<b>2</b>
1.1	Notiuni utilizate . . . . .	2
	Convolutii separabile in adancime . . . . .	2
	Module reziduale . . . . .	2
1.2	Baze de date . . . . .	3
<b>2</b>	<b>Reteaua Xception</b>	<b>4</b>
2.1	Arhitectura . . . . .	4
2.2	Optimizare . . . . .	4
2.3	Acuratete . . . . .	5
<b>3</b>	<b>Analiza imaginii</b>	<b>5</b>
<b>4</b>	<b>Bibliografie</b>	<b>6</b>

## 0 Introducere

Cele mai bune rezultate in identificarea si clasificarea imaginilor de-a lungul timpului au fost obtinute cu ajutorul retelelor convolutionale [1]. Cu toate aceasta, odata crescand complexitatea sarcinii unei retele, creste si numarul de parametri utilizati de aceasta. Astfel, viteza de procesare a cerintelor scade direct proportional cu complexitatea acestora.

O practica utilizata in mod frecvent in cadrul retelelor convolutionale este reprezentata de adaugarea unui ultim strat complet interconectat. In acest sens, se observa doua aspecte principale. Pe de o parte acest procedeu ajuta la conservarea si utilizarea robusta a tuturor informatiilor extrase de straturile anterioare, insa pe de alta parte numarul de parametri necesari unui astfel de strat este ridicat, atingand procentaje ridicate atunci cand privim in ansamblu retea. Asadar una dintre problemele evidente se reduce la gasirea unei modalitati de scadere a numarului de parametri.

O metoda semnificativa pentru reducerea numarului de parametri este reprezentata de utilizarea unui strat de extragere globala a mediei. Acest procedeu nu face decat sa aleaga o singura componenta din lista trasaturi unui strat (feature map) reprezentand media acestora. Ulterior s-au gasit diferite arhitecturi cobinand proceduri anterior descoperite pentru obtinerea unor rezultate optime. O astfel de arhitectura este Xception [1] care va fi detaliata in sectiunile urmatoare.

In aceasta lucrare se va urmari abordarea unei arhitecturi de dimensiuni reduse pentru a favoriza utilizarea unei retele convolutionale in timp real.

Inclusiv capacitatea umana de identificare a emotiei principale dintr-o imagine a unei fete este limitata. In acest sens, acuratetea umana in clasificarea unei emotii din cele 7 principale este de  $65 \pm 5$  % [4].

# 1 Preliminarii

## 1.1 Notiuni utilizate

### Convolutii separabile in adancime

**Definiție 1.1.** O convolutie separabila in adancime (eng. depthwise separable convolution) este compusa din doua elemente: o convolutie realizata in mod independent pe fiecare canal al datelor de intrare si o convolutie 1x1 (pointwise) care preia rezultatul primei convolutii si il combina intr-un strat de iesire. Astfel, rezultatul este proiectat in adancime intr-un nou spatiu de canale.

Rolul acestui tip de convolutie este de a reduce numarul de parametri utilizati.

### Module reziduale

Modulele reziduale modifica maparea trasaturilor dintre doua straturi alaturate astfel incat trasaturile invatate devin diferenta dintre lista initiala a trasaturilor si trasaturile as-teptate. In consecinta, trasaturile dorite, notate prin functia  $H$  sunt modelate astfel incat sa rezolve o problema de invatare  $F$  de forma:  $H(x) = F(x) + x$ . [5]

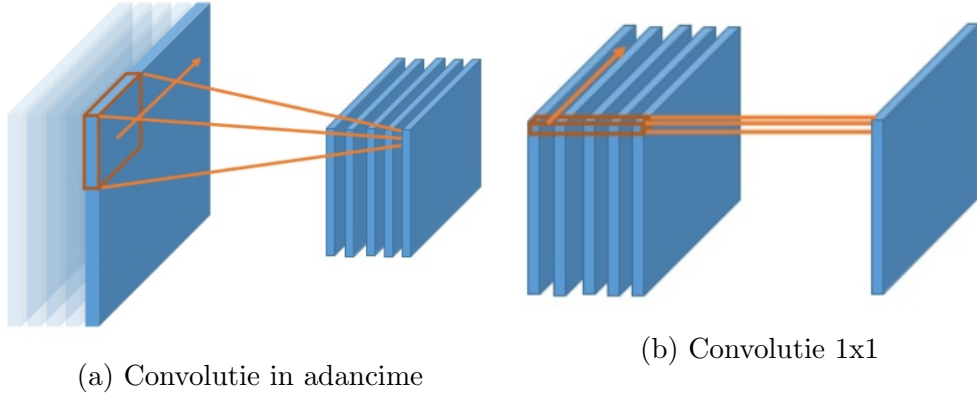


Figura 1: Convolutie separabila in adancime [2]

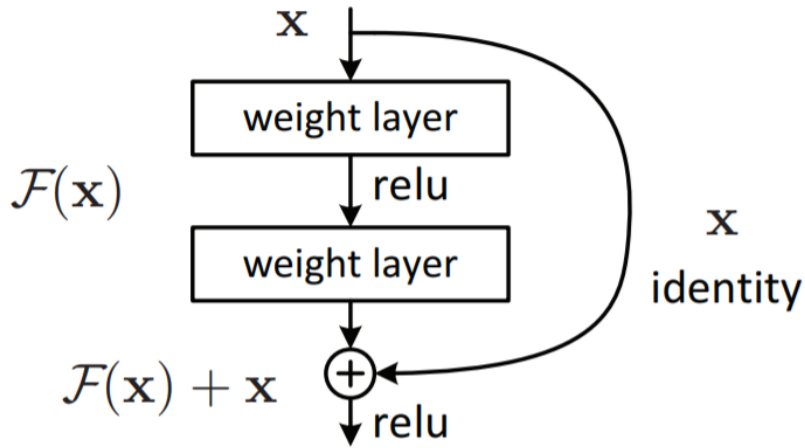


Figura 2: Modul rezidual

## 1.2 Baze de date

Pentru antrenarea rețelei s-au utilizat doua baze de date cunoscute si folosite anterior in diverse competitii. Pentru clasificarea emotiilor a fost utilizata baza de date FER-2013, iar pentru clasificarea genului baza de date IMDB.

Baza de date IMDB contine un total de 460,723 imagini color impartite in doua clase principale: "barbat" sau "femeie". Baza de date FER-2013 contine 35,887 imagini alb-negru, iar fiecare dintre imagini este etichetata cu una dintre cele 7 clase principale de emotii: "furie", "feicire", "tristete", "dezgust", "surprindere", "teama" sau "stare neutra".

Cea mai mare acuratete obtinuta in cadrul competitiei de catre o retea, pe baza de date FER-2013 este de 71 %, iar aceasta utilizeaza un total de aproximativ 5 milioane de parametri avand o pondere foarte mare a parametrilor in ultimul strat interconectat, de peste 90 % [4].

## 2 Reteaua Xception

### 2.1 Arhitectura

Arhitectura de retea Xception este realizata prin imbinarea a doua arhitecturi revolutionare: utilizarea modulelor reziduale si utilizarea convolutiilor separabile in adancime.

Pentru reducerea parametrilor utilizati, am utilizat in locul unui ultim strat complet interconectat, un strat de extragere globala a mediei (Global Average Pooling). Acest scop a fost sustinut de utilizarea ulterioara a convolutiilor separabile in adancime. In acest sens, obiectivul acestor convolutii este de a separa legaturile spatiale de cele dintre canale, aflate in adancime.

Arhitectura retelei este una convolutionala, continand 4 convolutii separabile in adancime, fiecare dintre acestea fiind urmata de cate o normalizare batch si functia de activare ReLU. Ultimul strat utilizeaza extragerea globala a mediei si o functie de activate soft-max pentru a obtine predictia claselor. Arhitectura va contine in final un numar de aproximativ 60,000 de parametri, ceea ce reprezinta cu 80 % mai putin decat in cazul unei retele convolutionale cu structura identica dar fara convolutii separabile in adancime si strat de extragere a mediei. [5]

Figura 3 prezinta structura fiecarui strat al retelei.

### 2.2 Optimizare

Optimizarea retelei a fost realizata cu ajutorul optimizatorului predefinit de catre tensorflow, Adam.

**Definiție 2.1.** Adam reprezinta un algoritm de optimizare folosit pentru actualizarea iterativa a ponderilor in cadrul antrenarii. Algoritmul este utilizat drept inlocuitor al gradientului descendent stocastic.

Principalele beneficii ale utilizarii acestui algoritm sunt reprezentate de:

- Rapiditatea implementarii
- Eficienta din punct de vedere computational
- Cantitatea redusa de memorie necesara
- Eficienta in cadrul problemelor cu dimensiune ridicata a datelor si parametrilor
- Eficienta in cazul datelor cu mult zgomot

Gradientul descendent stocastic mentine constanta rata de invatare pe tot parcursul antrenarii. Spre deosebire de acesta, Adam genereaza rate de invatare diferite in functie de ponderi pe baza estimarilor momentului I si II ale gradientilor. Adam combina avantajele altor doi algoritmi derivati din gradientul descendent stocastic: Algoritmul de gradient

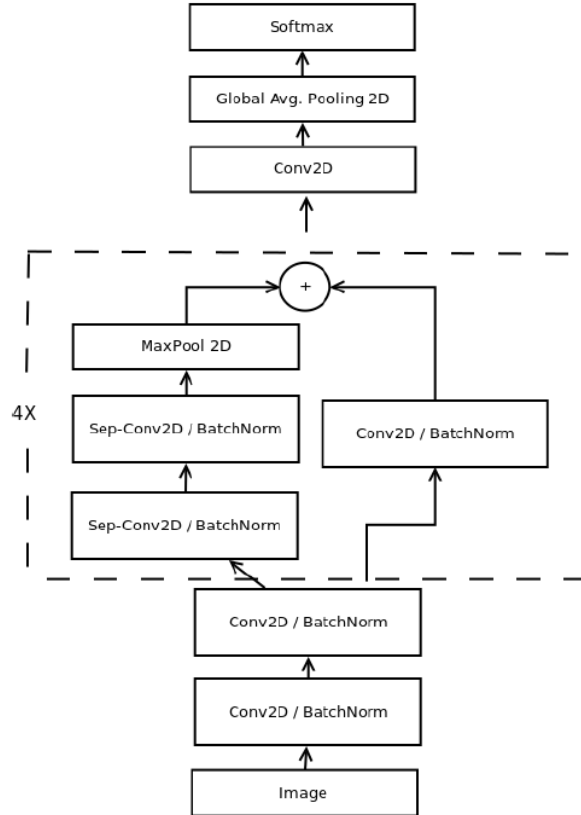


Figura 3: Arhitectura rețelei Xception

adapabil (AdaGrad) care mentine o rata de invatare care imbunatateste performanta in problemele cu gradienti redusi si Propagarea radacinii mediei patrate (RMSProp).

## 2.3 Acuratete

Aceasi retea din punct de vedere arhitectural a fost antrenata pe cele doua baze de date utilizate. Acuratetea obtinuta in clasificarea genului a fost de 95%, iar cea obtinuta in clasificarea celor 7 emotii, de 67%.

## 3 Analiza imaginii

Intregul ciclu de analiza a unei imagini, de identificare a emotiilor si a genului, consta intr-o imbinare de trei module principale. Initial imaginea este procesata cu ajutorul modelului de recunoastere faciala in cascada, oferit de catre biblioteca OpenCV. Astfel, acesta extrage coordonatele chenarelor ce incadreaza fiecare dintre fetele existente intr-o imagine. Ulterior, fiecare set de coordonate este utilizat pentru a selecta regiunea specificata din cadrul imaginii

initiale, aceasta fiind trecuta separat prin rețeaua de identificare a genului și prin rețeaua de identificare a emoției.



Figura 4: Identificarea emoțiilor

## 4 Bibliografie

[1] Francois Chollet, *Xception: Deep learning with depthwise separable convolution*, CoRR, abs/1610.02357, 2016.

[2] Dong-Won Shin *Depthwise separable convolution*, <https://www.slideshare.net/DongWonShin4/depthwise-separable-convolution>

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun *Deep Residual Learning for Image Recognition*, <https://arxiv.org/abs/1512.03385>

[4] Ian Goodfellow et al. *Challenges in Representation Learning: A report on three machine learning contests*, 2013.

[5] Matias Valdenegro, Octavio Arriaga, Paul G. Ploger *Real-time Convolutional Neural Networks for Emotion and Gender Classification*

[6] Diederik P. Kingma, Jimmy Ba, *Adam: A Method for Stochastic Optimization*, <https://arxiv.org/abs/1412.6980>