

# Klasifikacija muzejskih slika korišćenjem dubokih konvolutivnih neuronskih mreža

Nemanja Mićović

`nemanja_micovic@matf.bg.ac.rs`

`machinelearning.matf.bg.ac.rs`

Matematički fakultet  
Univerzitet u Beogradu

17. novembar 2017

# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije
- 4 Preprocesiranje i trening
- 5 Dalji rad

# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije
- 4 Preprocesiranje i trening
- 5 Dalji rad

# Neki od važnijih rezultata mašinskog učenja

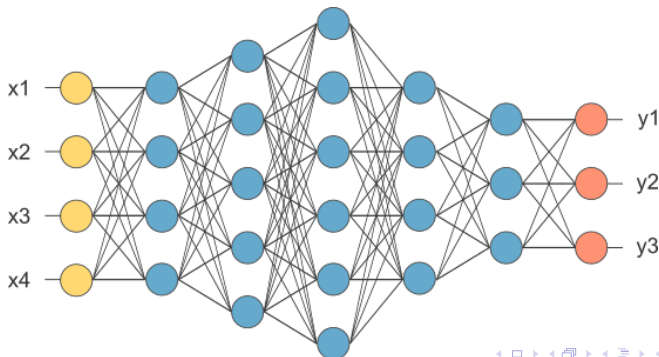
- 1992 - TD-Gammon, sistem koji igra igru Tavla, razvijen od strane Džeralda Tezaura
- 2011 - IBM-om Watson pobeđuje u kvizu *Jeopardy!*
- 2012 - Google X sistem koji prepoznaje mačke na video snimcima
- 2015 - Greška klasifikacije slika na ILSVRC 2015 od 3.6%
- 2016 - Sistem AlphaGo pobeđuje svetskog šampiona sa 4:1
- 2017 - AlphaGo igra protiv sebe iz 2016. godine i 100 partija pobeđuje 100

# Neke od primena mašinskog učenja

- Autonomna vožnja
- Algoritamski portfolio
- Igranje video igara
- Generisanje algoritama optimizacije  
[Andrychowicz et al., 2016]
- Klasifikacija slika
- Prepoznavanje rukopisa
- Generisanje slika
- Obrada prirodnih jezika
- Bioinformatika
- Društvene mreže

# Neuronske mreže

- Univerzalni aproksimatori funkcija
- U osnovi mnogih popularnih algoritama mašinskog učenja
- Više o njima u  
[Murphy, 2012, Bishop, 2006, Goodfellow et al., 2016]



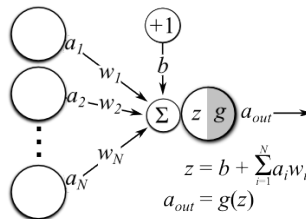
# Neuron neuronske mreže

## Terminologija:

- Ulaz neurona:  $a_i$
- Težine neurona:  $w_i$
- Slobodni član:  $b$
- Nelinearna funkcija:  $g$

Izlaz neurona se izračunava po formuli:

$$a_{out} = g\left(b + \sum_{i=1}^N a_i w_i\right)$$



# Aktivacione funkcije neuronske mreže

- Vrlo je bitno primeniti nelinearnu transformaciju inače će funkcija ostati linearna
- Neke od popularnih aktivacionih funkcija:
  - ReLU:  $g(x) = \max(0, x)$
  - Sigmoidna funkcija:  $g(x) = \frac{1}{1+e^{-x}}$
  - Tangens hiperbolički  $g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$



# Trening neuronskih mreža

Propagacija unazad (eng. backpropagation)

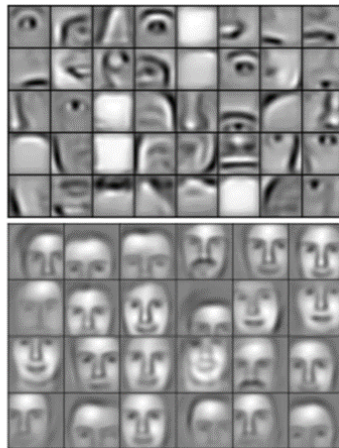
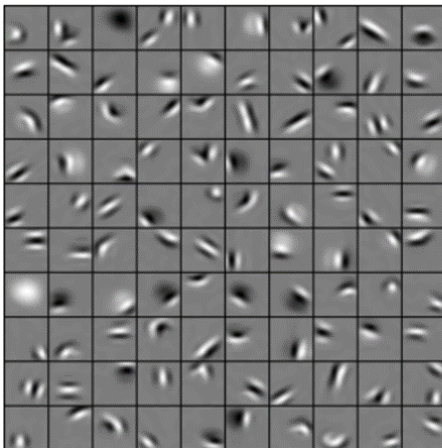
- Izračunava gradijent funkcije greške u odnosu na težine neurona
- Predstavlja osnovu algoritama za trening neuronskih mreža

Grafičke karte omogućavaju da se efikasno paralelizuju mnoge od potrebnih operacija za trening neuronskih mreža.

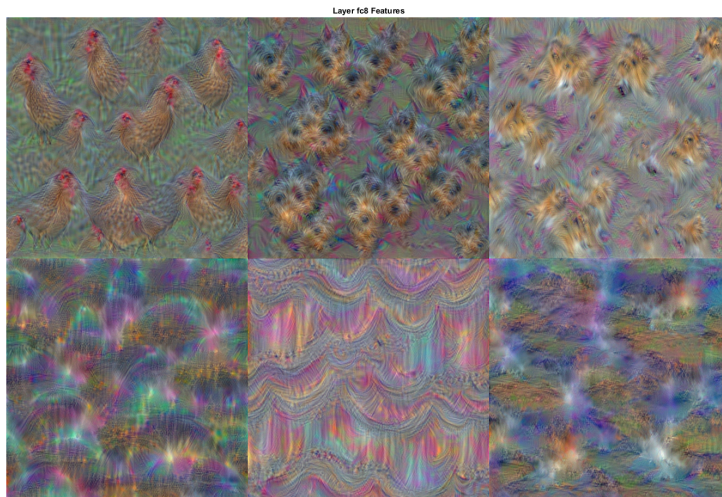
# Konvolutivne mreže

- Vrsta neuronskih mreža
- Prilagođena obradi signala u kojima postoji prostorna lociranost (slika, zvuk, video)
- Mogu vršiti konstrukciju relevantnih atributa
- Kompleksnost atributa koji se prepoznaje raste sa dubinom mreže

# Konvolutivne mreže



# Konvolutivne mreže



# Konvolutivne mreže - arhitektura

Arhitektura se sastoji iz kombinacije sledećih elemenata:

- Sloj konvolucije
- Sloj agregacije
- Standardna neuronska mreža

# Konvolutivne mreže - arhitektura

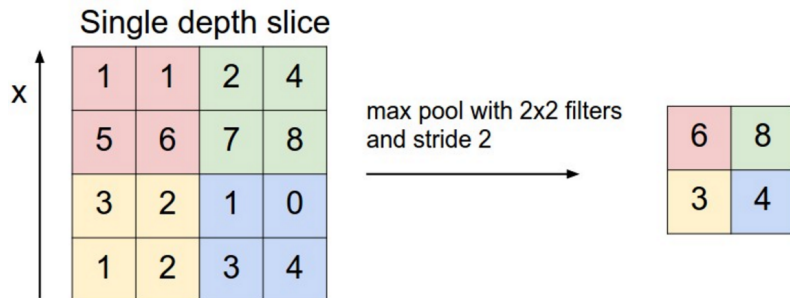
Konvolutivni sloj:

- Služi da detektuje određenu pravilnost u podacima
- Na primer, da detektuje horizontalne, vertikalne i kose linije (niži sloj) ili oči, uši i usne (viši sloj)

# Konvolutivne mreže - arhitektura

Agregatni sloj (eng. pooling):

- Uključuje informaciju iz prethodnog sloja (uglavnom je to konvolutivni sloj)
- Kao funkcija uključivanja se koristi maksimum ili prosek



# Konvolutivne mreže - arhitektura

Standardna neuronska mreža:

- Vrší klasifikaciju nad atributima koji su konstruisani u prethodnim slojevima





# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije
- 4 Preprocesiranje i trening
- 5 Dalji rad

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije
- da rezolucija fotografije ne utiče na preciznost klasifikacije

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije
- da rezolucija fotografije ne utiče na preciznost klasifikacije
- da ugao slike minimalno utiče na preciznost klasifikacije

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije
- da rezolucija fotografije ne utiče na preciznost klasifikacije
- da ugao slike minimalno utiče na preciznost klasifikacije
- da klasifikator skoro nikada ne greši



# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije
- da rezolucija fotografije ne utiče na preciznost klasifikacije
- da ugao slike minimalno utiče na preciznost klasifikacije
- da klasifikator skoro nikada ne greši
- to sve radi na mobilnim uređajima (i njihovim različitim platformama)

# Zahtevi

Želimo da korisnicima omogućimo da:

- fotografišu sliku na zidu i dobiju odgovor šta/ko je na njoj
- ne čekaju duže od 5 sekundi na odgovor
- kvalitet fotografije ne utiče drastično na preciznost klasifikacije
- da rezolucija fotografije ne utiče na preciznost klasifikacije
- da ugao slike minimalno utiče na preciznost klasifikacije
- da klasifikator skoro nikada ne greši
- to sve radi na mobilnim uređajima (i njihovim različitim platformama)
- mobilni uređaj ne pregori dok se vrši klasifikacija.

# Alati za klasifikaciju slika

- Izabrana je konvolutivna mreža usled zahteva o preciznosti i robusnosti klasifikatora
- Konvolutivne mreže imaju višegodišnji uspeh u klasifikaciji slika [Krizhevsky et al., 2012, He et al., 2015, Girshick et al., 2013, Karpathy and Li, 2014]
- Time je teže ispuniti ograničenje o podržanim platformama i problemu rada na mobilnom uređaju

# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije**
- 4 Preprocesiranje i trening
- 5 Dalji rad

# Rešavanje problema podržavanja platformi - rešenje 1

Iako postoje biblioteke koje rade na sistemima Android i iOS:

- zahtevaju pisanje odvojenih kodova (duplira se posao)
- nisu dovoljno popularne i korišćene, teško je dobiti podršku ukoliko se javi neki problem
- mobilni uređaji potencijalno nisu dovoljno hardverski snažni
- ažuriranje modela klasifikacije nije trivijalno (treba svaki korisnik da ažurira aplikaciju)
- kako proširiti skup podataka od korisničkih fotografija?

# Rešavanje problema podržavanja platformi - rešenje 2

Potencijalno rešenje:

- Dobijeni klasifikator (treningom konvolutivne mreže) postavljamo na javno dostupan server
- Mobilne aplikacije koriste uslugu servera (server pruža REST [Fielding, 2000] API<sup>1</sup>)
  - Za poslatu sliku, dobijaju klasu kojoj slika pripada

---

<sup>1</sup>Application programming interface

# Dobre strane

- Ne opterećuje se više mobilni uređaj
- Za klasifikator možemo koristiti proizvoljnu biblioteku za mašinsko učenje
- Model se može trenirati bilo gde
- Na serveru možemo čuvati nove korisničke fotografije i proširiti skup podataka
- Ažuriranje klasifikatora istovremeno ažurira svim korisnicima preciznost na mobilnim uređajima



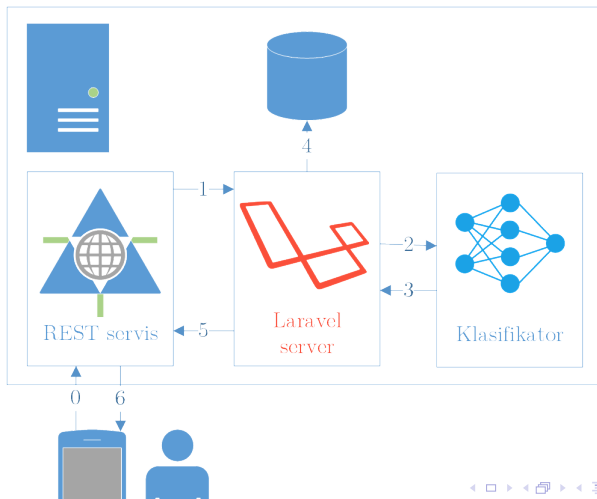
# Loše strane

- Neophodna je internet konekcija
- Velika količina istovremenih zahteva može stvoriti problem pri klasifikaciji
- Potreban server koji omogućava korišćenje biblioteka za mašinsko učenje
- Potencijalni bezbednosni propusti
- Održavanje servera





# Arhitektura



# Primer



# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije
- 4 Preprocesiranje i trening**
- 5 Dalji rad

# Skup podataka

- Potrebno napraviti skup podataka
- U muzeju postoji 9 slika na zidu (dakle 9 klasa)
- Za početak napravljen skup podataka od 1800 slika (200 slika po klasi)
  - Trening konvolutivne mreže prošao (očekivano) katastrofalno
  - Preciznost oko 0.3

# Proširivanje skupa podataka

Da li se skup podataka veštački može proširiti bez preprilagođavanja modela?

- Nije naivno pitanje a nije jednostavno detektovati preprilagođavanje u tom slučaju
- Neuronske mreže su ozloglašene i za preprilagođavanje i za potreban veliki broj podataka

Ispostavlja se da može.

# Proširivanje skupa podataka

- Primenjujemo nekoliko transformacija nad svakom slikom
  - Rotacije za  $\pm 5$  stepeni
  - Horizontalna translacije slike za najviše 10% njene visine
  - Vertikalna translacije slike za najviše 10% njene širine
  - Smicanje
  - Zumiranje
- Za svaku sliku generišemo 10 novih slika

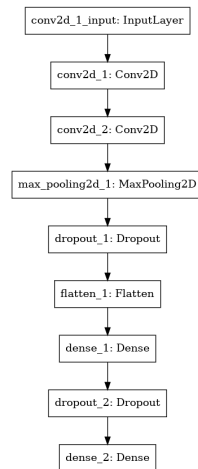


# Trening mreže

- Dobijen skup podataka od 18000 instanci rezolucije  $100 \times 178$
- Izvršen trening sa trening skupom od 12600 instanci (70% celog skupa)
- Korišćen algoritam optimizacije Adam [Kingma and Ba, 2014]
- Izabrano 40 epoha, i serija veličine 32 (eng. batch)
- Trenirano na grafičkoj karti Nvidia GeForce 1060 GTX
- Trening trajao oko 1h
- Dobijena preciznost od 99.83%
- Korišćene biblioteke TensorFlow [Abadi et al., 2015] i keras [Chollet et al., 2015]

# Arhitektura konvolutivne mreže

- conv2d\_1 - veličina  $3 \times 3$ , 32 filtera
- conv2d\_2 - veličina  $3 \times 3$ , 64 filtera
- maxpooling2d\_1 - agregacija veličine  $2 \times 2$
- conv2d\_1 - kovolutivni sloj, veličina  $3 \times 3$
- dropout\_1 - stepen 0.25  
[Srivastava et al., 2014]
- flatten\_1 - serijalizuje vektor
- dense\_1 - neuronska mreža sa 128 neurona
- dropout\_2 - stepen 0.25
- dense\_2 - neuronska mreža sa 9 neurona i funkcijom softmax





# Sadržaj

- 1 Uvod
- 2 Problem klasifikacije muzejskih slika
- 3 Arhitektura aplikacije
- 4 Preprocesiranje i trening
- 5 Dalji rad**

# Dalji rad

- Sagledati kako se dobijeni model pokazuje u praksi nakon nekoliko meseci
- Analizirati pogrešno klasifikovane slike
- Probati nekoliko drugih arhitektura mreže
- Pokušati klasifikaciju slika koristeći druge metode mašinskog učenja
- Evaluirati opterećenje servera pri nekoliko istovremenih klasifikacija (i poboljšati)
- Povećati skup podataka

# Pitanja?

# Hvala na pažnji!

# Bibliografija I



Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015).

TensorFlow: Large-scale machine learning on heterogeneous systems.

Software available from [tensorflow.org](https://www.tensorflow.org).

# Bibliografija II



Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., Shillingford, B., and de Freitas, N. (2016).

Learning to learn by gradient descent by gradient descent.  
*ArXiv e-prints*.



Bishop, C. M. (2006).

*Pattern Recognition and Machine Learning (Information Science and Statistics)*.

Springer-Verlag New York, Inc., Secaucus, NJ, USA.



Chollet, F. et al. (2015).

Keras.

<https://github.com/fchollet/keras>.

# Bibliografija III



Fielding, R. T. (2000).

*Architectural Styles and the Design of Network-based Software Architectures.*

PhD thesis.

AAI9980887.



Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2013).

Rich feature hierarchies for accurate object detection and semantic segmentation.

*ArXiv e-prints.*

# Bibliografija IV



Goodfellow, I., Bengio, Y., and Courville, A. (2016).

*Deep Learning*.

MIT Press.

<http://www.deeplearningbook.org>.



He, K., Zhang, X., Ren, S., and Sun, J. (2015).

Deep residual learning for image recognition.

*CoRR*, abs/1512.03385.



Karpathy, A. and Li, F. (2014).

Deep visual-semantic alignments for generating image descriptions.

*CoRR*, abs/1412.2306.



# Bibliografija V



Kingma, D. P. and Ba, J. (2014).

Adam: A method for stochastic optimization.

*CoRR*, abs/1412.6980.



Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).

Imagenet classification with deep convolutional neural networks.

In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.



Murphy, K. P. (2012).

*Machine Learning: A Probabilistic Perspective*.

The MIT Press.

# Bibliografija VI



Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014).

Dropout: A simple way to prevent neural networks from overfitting.

*J. Mach. Learn. Res.*, 15(1):1929–1958.