

Klasifikacija slika: prepoznavanje (pasmine) psa i mačke

Gayatri Čaklović, Branimir Jungić, Lena Kamenjaš, Mislav Kuzmić

U ovom članku bavimo se klasifikacijom slika, korištenjem dubinskih neuronskih mreža. Prvo provodimo binarnu klasifikaciju - svaku sliku svrstamo u jednu od dvije kategorije, pas ili mačka. Zatim isti skup slika pokušavamo svrstati u 37 kategorija, gdje svaka kategorija odgovara jednoj pasmini psa ili mačke.

I. UVOD

Klasifikacija je problem identificiranja kojim kategorijama (od zadanog skupa kategorija) pripada promatrani objekt, na osnovu podataka za koje je poznato kojim kategorijama pripadaju (skup za treniranje). Jedan od najpoznatijih primjera je svrstavanje e-maila u spam ili ne-spam kategoriju. U terminima strojnog učenja, problemu klasifikacije se pristupa preko nadziranog učenja. Problem koji smo mi rješavali je klasifikacija slika na kojima se nalaze mačke i psi. Inspiraciju za problem smo pronašli u Kaggleovom natjecanju *Dogs vs. Cats* iz 2013. godine u kojem je bilo potrebno napraviti klasifikator slika mačaka i pasa. Skup podataka koji smo koristili nalazi se na stranicama Sveučilišta u Oxfordu.¹ Prvi cilj našeg projekta je kategorizirati sliku u kategoriju mačka ili pas. Drugi cilj našeg projekta je izgraditi mrežu koja će kategorizirati slike u 37 kategorija: 25 pasmina pasa i 12 mačaka.

Potaknuti ovim natjecanjem, projekt možemo podijeliti na dva dijela – želimo prvo kategorizirati za zadanu sliku je li na njoj pas ili mačka, a zatim probati odrediti i pasminu zadane životinje.

Ovaj članak se sastoji od 6 poglavlja: prvi je ovaj, uvod, u drugom ćemo opisati skup podataka koji smo koristili, u trećem ćemo opisati model i tehnologije kojima smo rješavali problem, u četvrtom ćemo prikazati rezultate, u petom napraviti osvrt na druge pristupe rješavanju problema i šesti je zaključak. Zbog preglednosti, neke slike su stavljene nakon literature.

II. OPIS PROBLEMA

Korišteni skup podataka sastoji se od dvije cjeline - anotacija i 7390 slika. Anotacije sadrže:

- *trimap* anotaciju za svaku sliku - *trimap* anotacija sadrži tri boje, jedna je za pozadinu slike, jedna za prednji dio slike, jedna za neodređeni dio (ne spada niti u pozadinu niti u prednji dio slike) skup podataka, odakle, koliko se nalazi, opisat skup podataka

- *xml* datoteke - anotacije koje pravokutnikom za svaku sliku označavaju gdje se nalazi glava životinje, u *PASCAL VOC* formatu
- *list.txt* datoteku - listu svih slika
- *trainval.txt* i *test.txt* - datoteke koje opisuju koji dio podataka je odvojen za testiranje, a koji za treniranje

Od priloženih anotacija u ovom projektu korištena je samo datoteka *list.txt*. Slike su raspoređene u 37 kategorija - 12 je kategorija mačaka i 25 kategorija pasa. U svakoj kategoriji ima otprilike 200 slika. Slike koje započinju velikim početnim slovom su slike mačaka, a one koje započinju malim početnim slovom su slike pasa. Na svakoj slici je točno jedna mačka ili točno jedan pas.

III. MODELI

U ovom poglavlju opisujemo modele i tehnologije koje smo koristili za građenje istih.

Na osnovu ovih podataka gradimo i treniramo naše modele. Prije opisivanja modela, objašnjavamo često korištene pojmove vezane za neuronske mreže:

- konvolucijski sloj - sloj u neuronskoj mreži korišten za otkrivanje značajki u slici. Svaki konvolucijski sloj sadrži određen broj kanala.
- kanal - svaki kanal je zadužen za otkrivanje specifične značajke
- filter - matrica kojom se otkriva značajka u svakom kanalu
- *ReLU* sloj - standardna aktivacijska funkcija

$$ReLU(x) = \max\{0, x\}$$

- *MaxPool2d* sloj - koristi se za reduciranje dimenzionalnosti, umanjuje visinu i širinu slike za faktor 2
- linearni sloj - potpuno povezani sloj koji računa vrijednosti za svaku od klasa (u našem slučaju broj klasa je 2 ili 37)
- *BatchNorm2d* sloj - normalizira podatke za veću točnost modela
- *Dropout2d* sloj - na slučajan način eliminira neke dijelove ulaznog podatka
- *Sigmoid* sloj - aktivacijska sigmoidna funkcija

$$f(x) = \frac{1}{1 + e^{-x}}$$

A. Model za binarno klasificiranje

Za implementaciju modela korišten je *PyTorch*, *framework* za duboko učenje.

Model se sastoji od 43 sloja. Označimo s **baza** sekvencijalnu kombinaciju od 3 sloja:

$$\text{baza} = \text{Conv2d} - \text{BatchNorm2d} - \text{ReLU} \quad (1)$$

Tada konvolucijska neuronska mreža nastala od ovog modela izgleda:

3 * baza - MaxPool2d - 4 * baza - MaxPool2d - 3 * baza - MaxPool2d - 2 * (Linear - BatchNorm1d - ReLU - Dropout2d) - Linear - Sigmoid.

B. Modeli za klasificiranje u kategorije

U nastavku su navedena dva modela kojim smo pokušali klasificirati slike u 37 kategorija.

1. Model zasnovan na konvolucijskoj neuronskoj mreži (CNN Model)

Za implementaciju ovog modela također je korišten *PyTorch*.

Model se sastoji od 43 sloja. Uz prethodnu oznaku 1, neuronska mreža izgleda:

2 * baza - MaxPool2d - 3 * baza - MaxPool2d - 5 * baza - MaxPool2d - Linear - BatchNorm1d - ReLU - Dropout - Linear - BatchNorm1d - ReLU - Dropout - Linear - LogSoftMax.

Za ovaj, i prethodni (binarni) klasifikator, skup podataka je na slučajan način podijeljen 70%(podaci za treniranje) - 30%(podaci za testiranje). Na podacima za treniranje izvršeno je preprocesiranje - za svaku sliku kreirano je dodatnih 16 sa različitim efektima (npr. izoštravanje, *blur*, rezanje slike, linearne transformacije) u svrhu poboljšanja točnosti.

2. Model zasnovan na ResNet-50 modelu (ResNetModel)

Za implementaciju ovog modela korišten je *Keras*. Model je baziran na ResNet-50 modelu s izmijenjenim zadnjim slojem. Dodali smo 4 sloja: *ReLU* (potpuno povezani sloj), *AveragePooling2D*, *Dropout* (0.4) i *SoftMax* (potpuno povezni sloj).

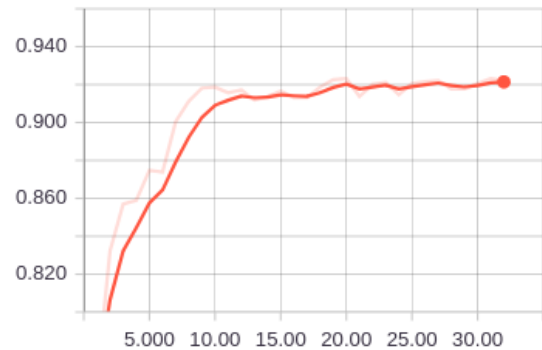
Skup podataka je podijeljen 70% podaci za treniranje i 30% podaci za testiranje.

IV. REZULTATI

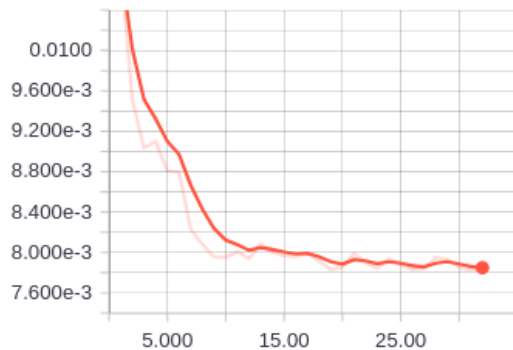
Unutar ovog poglavlja iznosimo dobivene rezultate za svaki od prethodno 3 definirana modela.

A. Model za binarno klasificiranje

Treniranje i testiranje modela je izvedeno na Intel Core i7-7700HQ procesoru, s 8GB RAM memorije i Nvidia Geforce GTX - 1050 TI (4GB) grafičkoj kartici (*CUDA Computability* je 6.1). Treniranje je trajalo 475 minuta. Model je treniran kroz 32 epohe. Dobivena točnost modela na skupu podataka za treniranje je iznosila 94.5%, a na skupu podataka za testiranje 92.3%.



Slika 1: Graf prikazuje točnost kroz epohe učenja modela za binarnu klasifikaciju.



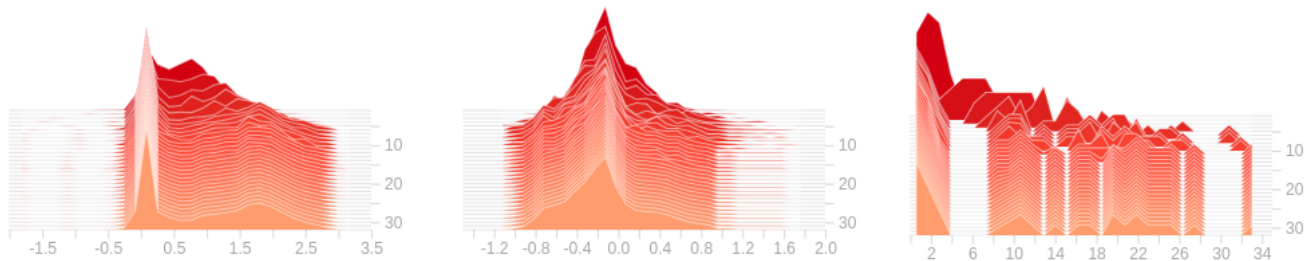
Slika 2: Graf prikazuje točnost kroz epohe učenja modela za binarnu klasifikaciju.

Na grafovima na slikama 1 i 2 vidimo kako točnost raste kroz epohe. No, vidi se i da se točnost stabilizirala nakon 20. epohe (što bi značilo da nije bilo potrebe za provođenjem 32 epohe).

Matrica konfuzije za testne podatke je:

	Predviđene mačke	Predviđeni psi
Stvarne mačke	924	61
Stvarni psi	95	948

Na slici 3 vidimo kako se model poboljšava kroz epohe. Slika lijevo pokazuje kako prvi sloj računa težine na temelju ulaznih podataka. Te težine reprezentiraju značajke



Slika 3: Lijevo: naučene težine za prvi sloj mreže kroz epohe. Sredina: pristranost prvog sloja mreže kroz epohe. Desno: pristranost 34. sloja mreže kroz epohe.

koje će mreža koristiti pri klasifikaciji slika. Slika u sredini za isti sloj prikazuje njegovu pristranost, i kako se ona kroz epohe smanjuje. Slika desno prikazuje 34. sloj mreže, gdje vidimo da je njegov bias već na početku manji od početnog sloja.

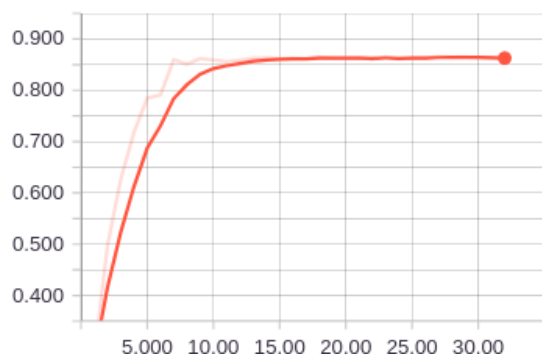
Na slici 8 nalazi se slika psa, a na slici 9 mačka s raznim transformacijama koje radi prvi sloj mreže. Možemo vidjeti da prvi sloj radi s crno-bijelim slikama te da je izvukao 32 značajke.

Na slici 10 vidimo predikcije klasifikatora na testnim podacima, a na slici 11 vidimo predikcije na novim, dosad neviđenim podacima.

B. CNN model za više kategorija

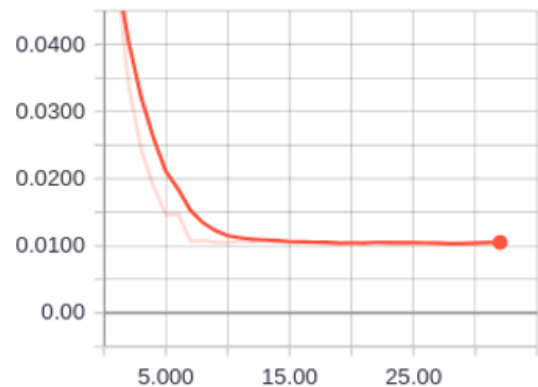
Treniranje i testiranje modela je izvedeno na Intel Core i7-7700HQ procesoru, s 8GB RAM memorije i Nvidia Geforce GTX - 1050 TI (4GB) grafičkoj kartici (*CUDA Computability* je 6.1).

Treniranje je trajalo 342 minute. Model je treniran kroz 32 epohe. Točnost podataka na skupu za testiranje iznosila je 86.6%.

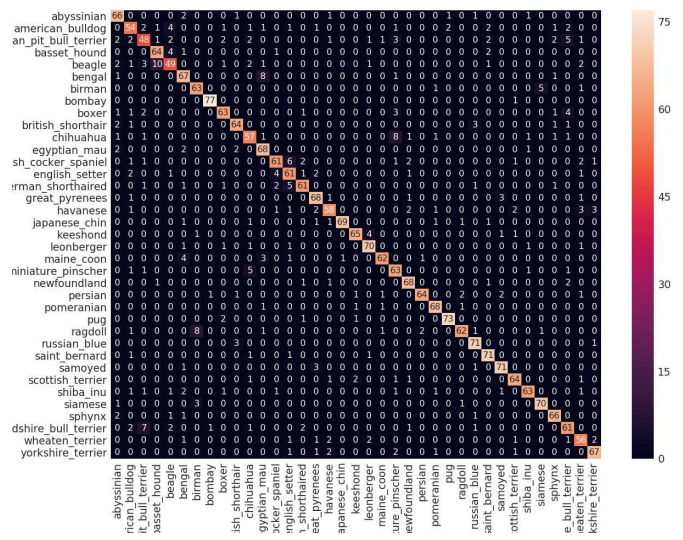


Slika 4: Graf prikazuje točnost kroz epohe učenja CNNModela.

Analogno kao i kod binarnog klasifikatora, točnost raste kroz epohe, no stabilizira se oko 15. epohe. Slike 4 i 5 prikazuju grafove točnosti i gubitka. Matrica konfuzije je prikazana na slici 6.

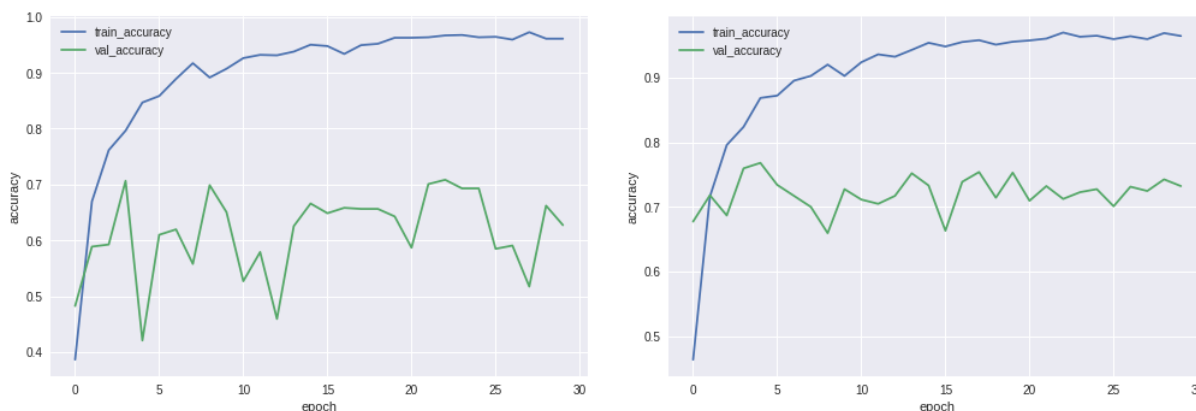


Slika 5: Graf prikazuje gubitak kroz epohe učenja CNNModela.



Slika 6: Matrica konfuzije za CNN Model za 37 kategorija.

Na slici 12 vidimo kako klasifikator (ne) radi - na slikama gdje je u prvom fokusu pas i gdje se točno vide specifičnosti za pojedinu pasminu, klasifikator ispravno pogađa pasminu. Na istoj slici se vidi kako je istog psa



Slika 7: Lijevo: graf točnosti kroz epohe učenja modela za mačke.
Desno: graf točnosti kroz epohe učenja modela za pse.

smjestio u 3 različite kategorije.

Na slici 13 vidimo kako klasifikator pokušava smjestiti pse (čije pasmine nisu u skupu podataka). Istog psa će smjestiti u različite kategorije, ovisno o tome što je vidljivo na slici i što klasifikator prepoznaje - npr. kombinaciju boja krzna, položaj tijela.

C. Model zasnovan na ResNet-50 modelu (ResNetModel)

Trening i testiranje modela izvedeno je na servisu Google Colab, a trajalo je 20 minuta. Zbog ograničenja sustava nismo odjednom mogli trenirati model na svim slikama pa smo prvo trenirali model na slikama mačaka (12 kategorija), potom na slikama pasa (25 kategorija).

Modeli su trenirani na 40 epoha. Dobivena točnost na skupu podataka za treniranje je iznosila 95.67% za mačke i 96.97% za pse, a točnost na skupu za testiranje 65.76% za mačke i 76.29% za pse.

Na slici 7 se vide grafovi točnosti za modele za mačke i pse. Na slici 14 se vidi histogram predikcija modela za mačke. S histograma se može primijetiti da su tri problematične kategorije koje imaju više netočnih nego točnih predikcija modela: *russian*, *bengal* i *abyssinian*. Na slici 15 se vidi histogram predikcija modela za pse. S histograma se može primijetiti da su tri problematične kategorije koje imaju više netočnih nego točnih predikcija modela: *beagle*, *miniature*, *american*.

V. OSVRT NA DRUGE PRISTUPE

Obzirom da je 2013. godine objavljen natječaj na stranici Kaggle, za binarno klasificiranje, ovaj je problem rješavan na dosta različitih pristupa. Jedan od primjera naveden je u literaturi², gdje se korištenjem *transfer learning* metode postigla točnost od 98% (koristi se već definirani *ResNet50* model). Godine 2012. su autori

korištenog skupa podataka objavili članak⁴, gdje su pristupili problemu klasificiranja u 37 kategorija i dobili prosječnu točnost od 59%.

VI. ZAKLJUČAK

Vidimo da je ResNetModel postigao slabiju točnost od CNNModela, iako ima više slojeva. Možemo zaključiti da preprocesiranje slika bitno utječe na točnost modela. Preprocesiranje nije bilo moguće izvršiti za ResNetModel zbog ograničenja memorije na Google Colab servisu.

Jedno od poboljšanja (evaluacijskog i optimizacijskog) za klasifikatore bi moglo biti treniranje težina preko SVM-a (*Support Vector Machine*), u svrhu otkrivanja bitnih značajki slike, i ubacivanje tih naučenih težina u neuronsku mrežu, koja tada ne bi morala sama detektirati bitne značajke.

Drugo moguće poboljšanje bi bilo kombinacija binarnog i 37-arnog klasifikatora - klasifikator bi učio i informaciju da je na slici pas/mačka i informaciju kojoj pasmini pripada. Predikcija bi se sastojala tad od 2 koraka - određivanje je li na slici pas ili mačka te ovisno o tom odgovoru predikcija kojoj pasmini pripada.

LITERATURA

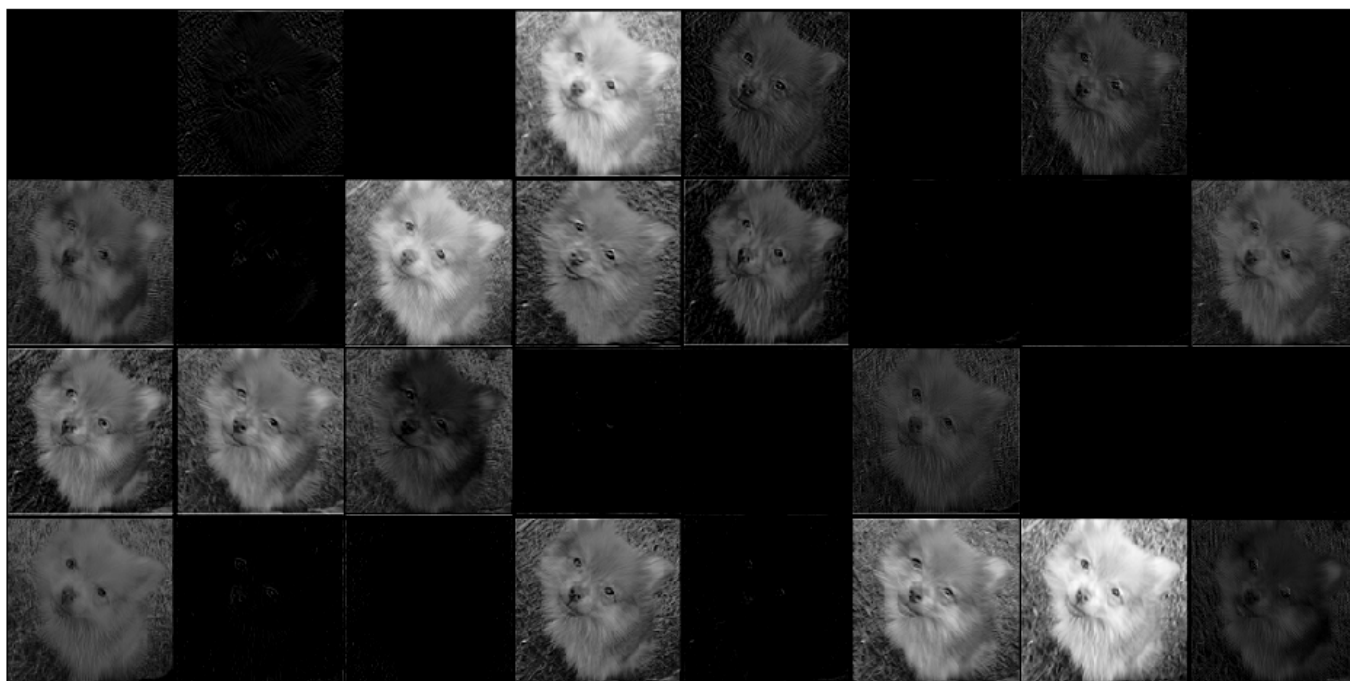
¹The Oxford-IIIT Pet Dataset, <http://www.robots.ox.ac.uk/~vgg/data/pets/>

²Abhik Sarkar, Transfer Learning Cats And Dogs, <https://www.kaggle.com/abhiksarkar/transfer-learning-cats-and-dogs/notebook>

³Basics of Image Classification with PyTorch, <https://heartbeat.fritz.ai/basics-of-image-classification-with-pytorch-2f8973c51864>

⁴Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, C. V. Jawahar, Cats and Dogs, <http://www.robots.ox.ac.uk/~vgg/publications/2012/parkhi12a/parkhi12a.pdf>

⁵J. Donini, Cats and Dogs Classification, <https://github.com/JDonini/Cats-and-Dogs-Classification>



Slika 8: Slika psa nakon prvog sloja mreže binarnog klasifikatora.



Slika 9: Slika mačke nakon prvog sloja mreže binarnog klasifikatora.

predicted: dog



predicted: dog



predicted: dog



predicted: dog



predicted: dog



predicted: cat



predicted: cat



predicted: cat



predicted: cat



predicted: cat



predicted: dog



predicted: dog



Slika 10: Predikcije binarnog klasifikatora na testnim podacima.

Slika 11: Predikcije binarnog klasifikatora na novim podacima.

predicted: english_setter



predicted: miniature_pinscher



predicted: german_shorthaired



predicted: yorkshire_terrier



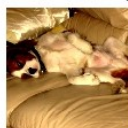
predicted: ragdoll



predicted: scottish_terrier



predicted: bengal



predicted: miniature_pinscher



predicted: saint_bernard



predicted: leonberger



predicted: russian_blue

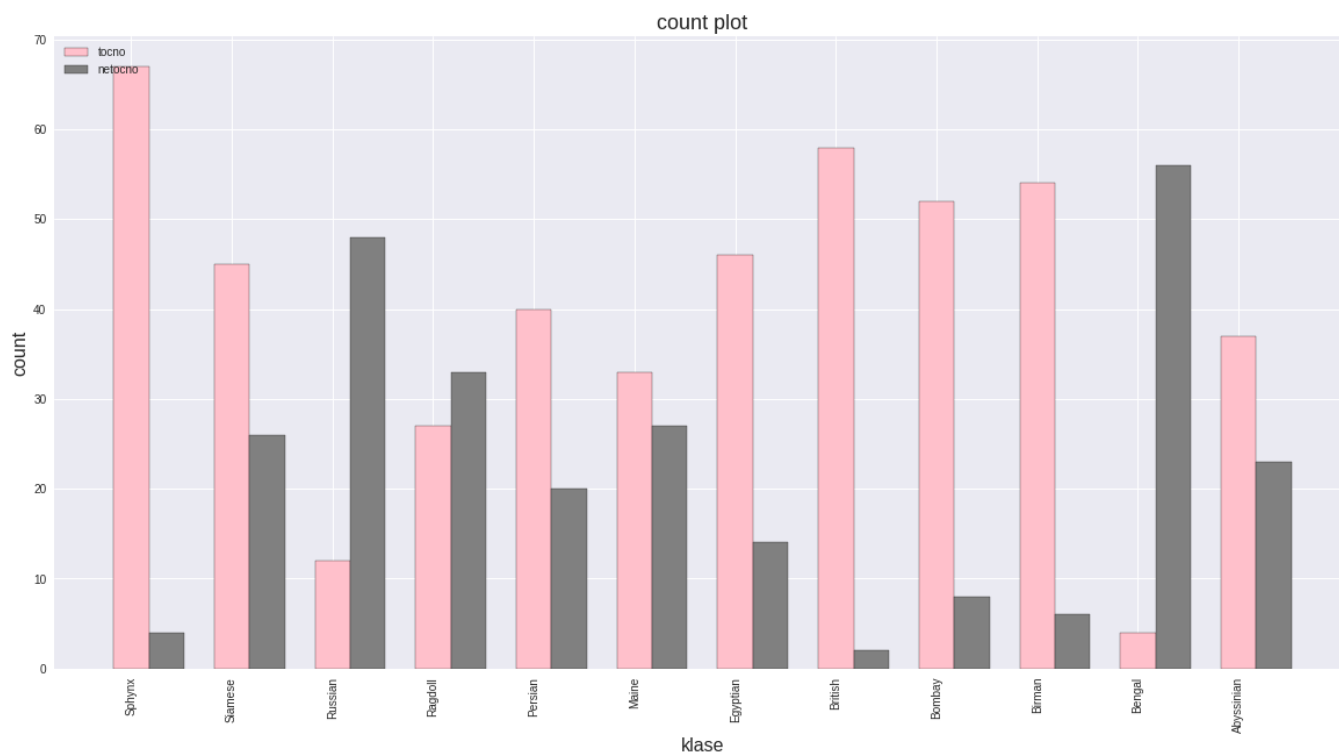


predicted: german_shorthaired

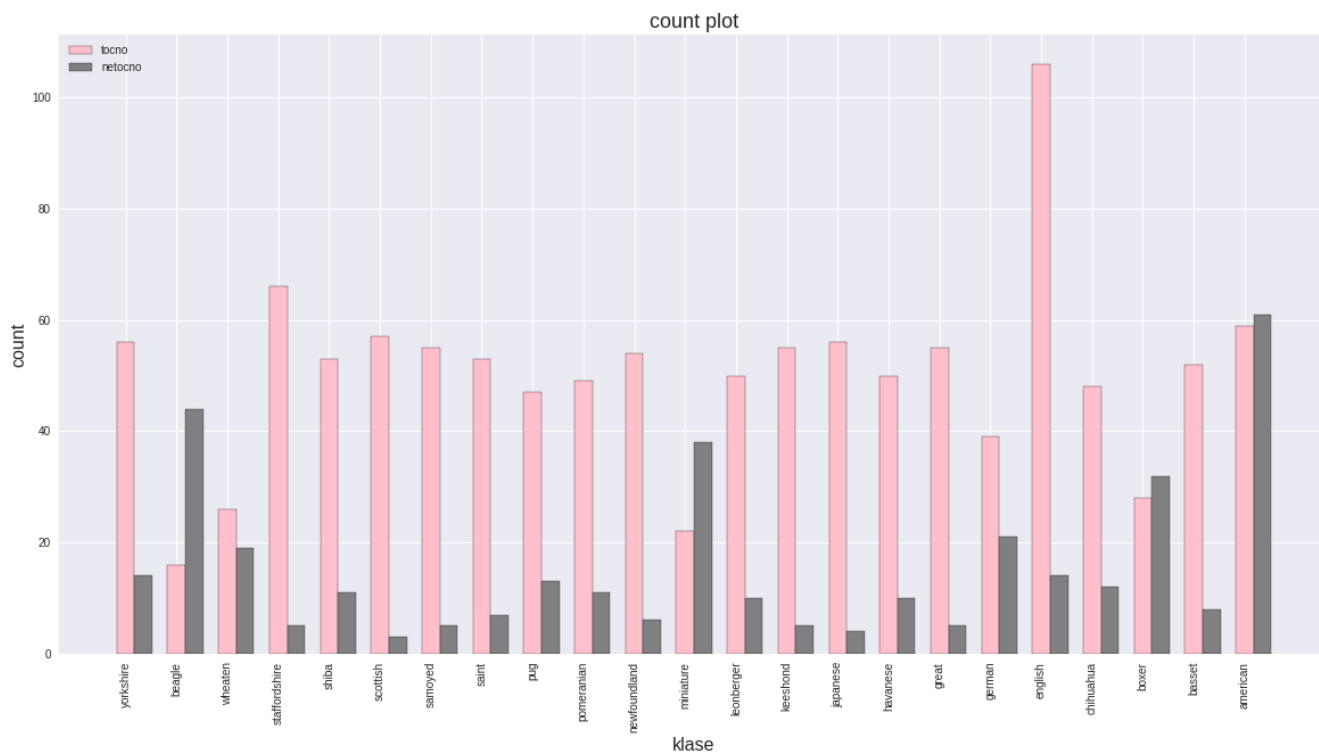


Slika 12: Predikcije *CNNModela* na novim podacima.

Slika 13: Predikcije *CNNModela* na novim podacima.



Slika 14: Histogram predikcija kategorija mačaka. Rozi stupci su točno klasificirane slike, sivo netočno.



Slika 15: Histogram predikcija kategorija pasa. Rozi stupci su točno klasificirane slike, sivo netočno.