

Određivanje pasmina pasa

Iva Sever, Mateja Šarić, Valentina Tucelj i Kristina Vellico

U ovom članku bavit ćemo se problemom određivanja pasmina pasa, odnosno problemom klasifikacije slika pomoću dubokih neuronskih mreža. Koristeći razne metode pokušati ćemo za sliku psa što točnije odrediti koju pasminu predstavlja.

I. UVOD

Danas postoji oko 40 milijuna pasa i oko 800 pasmina, što je više nego bilo koja druga vrsta životinje. Razlikujemo ih prema dimenzijama, fizionomiji, temperamentu, ali isto tako i prema boji ili vrsti dlake. 2017. godine Kaggle je pokrenuo natjecanje Dog Breed Identification, gdje je cilj bio stvoriti klasifikator sposoban za određivanje pasmine psa sa slike. Nekoliko je faktora koji otežavaju problem klasifikacije. Postoji puno različitih pasmina i neke od njih su poprilično slične. Točnije, pasmina na prvu možda nije očita ni ljudima koji imaju puno iskustva na tom polju. Također slike pasa mogu prikazivati pse svih oblika, veličina, boja, pod različitim svjetlima, u raznim pozama i na raznim lokacijama. Slike imaju različite rezolucije, pozadine i dimenzije. Na nekim slikama su po dva ili više psa, na nekima su ljudi s psima, a negdje psi nose i odjeću. Sve navedeno otežava klasifikaciju pasa. Koristeći pretrained modele neuronskih mreža trenirane na ImageNet skupu podataka proučit ćemo koji od modela daje najveću točnost. Koristit ćemo Res-Net50, VGG16, Inception i Xception. VGG16, Inception i Xception ćemo iskoristiti za ekstrakciju značajki na koje ćemo primijeniti višeklasnu logističku regresiju.

II. OPIS PROBLEMA

Za treniranje i evaluaciju dan je skup podataka preuzet sa stranice natjecanja od 20 579 stvarnih slika i 120 pasmina. Podaci su podijeljeni na:

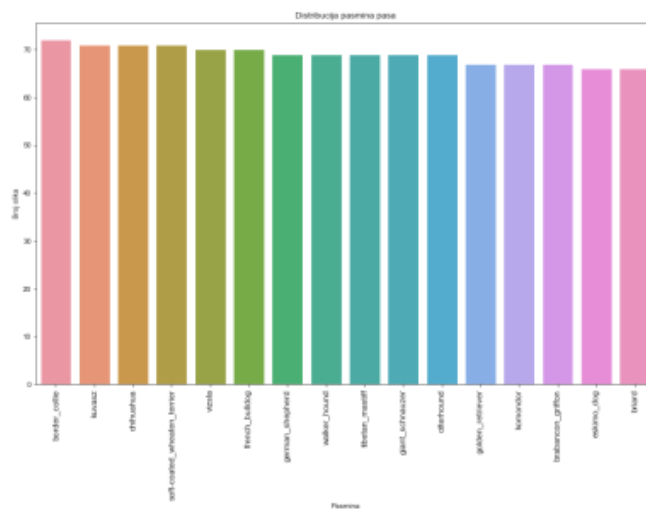
- *train.zip* - mapa s 10 222 slika za treniranje
- *test.zip* - mapa s 10 357 slika za testiranje
- *labels.csv* - tablicu s ide-evima slika i pripadnom pasminom za training skup
- *sample_submission.csv* - tablica primjera pravilnog ispisa vjerojatnosti da je pasmina sa slike jedna od 120 polaznih

Slike su raspoređene u 120 kategorija. U skupu se ne nalaze samo čiste slike pasa, već imamo slike na kojima se pojavljuju ljudi i 2 ili više pasa.

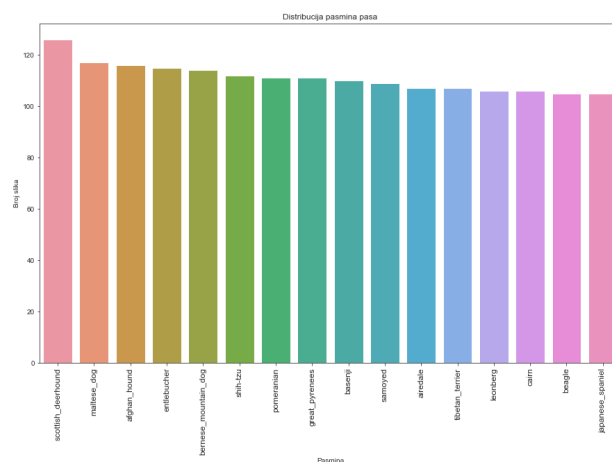
Prvo smo naš skup podataka analizirale eksploratornom analizom te uz pomoć histograma proučile omjer zadanih pasmina. Eksploratorna analiza podataka odnosi se na kritični proces izvođenja početnih istraživanja podataka kako bi se otkrili obrasci, uočile nepravilnosti, testirale hipoteze i provjerile pretpostavke uz pomoć sažetih statistika i grafičkih

prikaza.

Velike razlike u skupu podataka između kategorija (Slika 1 i Slika 2) dovele su nas do toga da se fokusiramo na samo 16 najzastupljenijih kategorija, jer je npr. najzastupljenija pasmina Scottish deerhound prepoznata u 126 slika, a najmanje zastupljene pasmine Briard i Eskimo prepoznate su na samo 66 slika.



Slika 1. 16 kategorija najmanje zastupljenih pasmina u skupu podataka



Slika 2. 16 kategorija najviše zastupljenih pasmina u skupu podataka

III. METODE

U strojnom učenju razlikujemo nadzirano (engl. supervised learning) i nenadzirano učenje (engl. unsupervised

learning). U nadziranom učenju podaci su parovi

$$(ulaz, izlaz) = (x, y)$$

i treba pronaći $\hat{y} = f(x)$. Ako je y diskretna (nebrojčana) vrijednost, tada se radi o klasifikaciji, a ako je y kontinuirana (brojčana) vrijednost, tada se radi o regresiji.

U nenadziranom učenju dani su podaci bez ciljne vrijednosti te treba naći pravilnost u podacima. Kod nenadziranog učenja razlikujemo grupiranje (engl. clustering), procjenu gustoće (engl. density estimation) i smanjenje dimenzionalnosti (engl. dimensionality reduction).

Za rješenje našeg problema koristit ćemo nadzirano učenje, točnije klasifikaciju. Klasifikacija je problem identifikiranja kojim kategorijama iz zadanog skupa kategorija (u ovom slučaju pasmina pasa) pripada promatrani objekt (slika psa) na osnovi podataka (iz skupa za treniranje) za koje je poznato kojim kategorijama pripadaju.

U istraživanju ćemo koristiti pretrained modele temeljene na konvolucijskim neuronskim mrežama (CNN). CNN možemo prikazati kao niz slojeva pri čemu svaki sloj transformira rezultat prethodnog pomoću određenih funkcija i proizvodi izlaz koji se zove mapa značajki. Osnovne komponente CNN-a su ulazni sloj (ulazna slika), konvolucijski sloj (uzima sliku iz prethodnog sloja i konvoluiraju je određenim brojem filtera, izlazna slika je mapa značajki), sloj sažimanja (manji dio uzorka se grupira i obrađuje te rezultira jednom vrijednošću) i potpuno povezani sloj (neuroni potpuno povezani s prethodnima).

A. Logistička regresija

Logistička regresija je klasifikacijski model strojnog učenja, koji se koristi za predviđanje kategorijskih izlaza, bez obzira na tip ulaznih varijabli. Ovisna varijabla kod logističke regresije je diskretna (kategorička) i dihotomna, tj. može poprimiti jednu od dvije vrijednosti od kojih su obje nekakve kategorije. Neovisne varijable mogu biti i kategoričke i metričke. Model binomne logističke regresije koristi se za procjenu vjerojatnosti realizacije jedne od kategorija binomne ovisne varijable na temelju poznatih vrijednosti jedne ili više neovisnih varijabli. Ukoliko postoji više od dvije kategorije ovisne varijable možemo koristiti višeklasnu logističku regresiju.

U ovisnosti o kakvim se kategorijama radi, odnosno kakva je slika ovisne varijable, razlikujemo dvije vrste višeklasne logističke regresije: logistička regresija s nominalnom ovisnom varijablom i logistička regresija s ordinalnom ovisnom varijablom.

Nominalna varijabla ima sliku u kojoj ne postoji skala ili redoslijed po kojoj je neka kategorija više ili manje vrijedna od druge, kao što je npr. plava, zelena, crvena, žuta boja, ili, odgovor na pitanje u nekom upitniku: da, ne, ne znam, nema odgovora.

Ordinalna varijabla je ona koja poprima vrijednosti iz skupa u kojem je prisutna skala kvalitete ili prirodni slijed, kao npr. dobar, srednji ili loš klijent.

Logistička regresija se koristi kada zavisna varijabla uzima samo jedan konačni skup vrijednosti. Mi ćemo koristiti višeklasnu logističku regresiju.

B. ResNet50

ResNet (Residual Network) je skraćenica za rezidualne mreže. Rezidualno učenje pokušava riješiti probleme klasifikacije i točnosti prepoznavanja. S obzirom na ulazni podatak x cilj neuronske mreže je pronaći korisnu funkciju mapiranja $H(x)$. Funkcija $H(x)$ (u sklopu klasifikacije) pokušava mapirati ulaz na izlaz, odnosno klasificirati dani podatak u jednu od specifičnih klasa. Neka je:

$$F(x) = H(x) - x,$$

gdje je $F(x)$ rezidualna funkcija. Hipoteza arhitekture ResNet je da je lakše optimizirati rezidualnu funkciju $F(x)$ nego originalnu funkciju mapiranja $H(x)$. Izvorna funkcija mapiranja $H(x)$ je ono što je potrebno pa se u rezidualnim mrežama koristi:

$$H(x) = F(x) + x.$$

Osnovna gradivna jedinica ResNet arhitekture je ResNet blok gdje su veze preko slojeva težina zvane preskočne veze. ResNet50 je 50 slojna rezidualna mreža. Postoje još inačice sa 18, 34, 101, 152 slojeva.

C. Inception

Neuronska mreža Inception bila je važna prekretnica u razvoju CNN klasifikatora. Prije svog početka, većina popularnih CNN-ova samo su slagali slojeve konvolucija dublje i dublje, nadajući se da će dobiti bolje performanse.

Neuronska mreža Inception, s druge strane, bila je složena (jako projektirana). Koristila je mnogo trikova za poboljšanje performansi u pogledu brzine i točnosti. Njegov stalni razvoj vodi do stvaranja nekoliko verzija mreže. Popularne verzije su: Inception v1, Inception v2 i Inception v3, Inception v4 i Inception-ResNet.

Svaka verzija je iterativno poboljšanje u odnosu na prethodnu. Razumijevanje nadogradnji može nam pomoći u izgradnji prilagođenih klasifikatora koji su optimizirani u brzini i točnosti. Isto tako, ovisno o podacima, niža verzija može bolje funkcionirati.

Inception v2 i Inception v3 predstavljeni su u istom radu. Autori su predložili brojne nadogradnje koje su povećale točnost i smanjile složenost računanja. Intuicija je bila da neuronske mreže bolje funkcioniraju kada konvolucije ne mijenjaju drastično dimenzije ulaza. Preveliko smanjenje dimenzija može uzrokovati gubitak informacija, poznato kao "representational bottleneck". Koristeći pametne metode faktorizacije, konvolucije se mogu učiniti učinkovitijima u smislu složenosti računanja.

U našem projektu koristile smo Inception v3.

Autori su primijetili da pomoćni klasifikatori nisu mnogo doprinijeli sve do gotovo kraja trening procesa, kada su se točnosti približavale zasićenju. Trebalo je istražiti mogućnosti poboljšanja na Inception v2 bez drastičnog mijenjanja modula. Inception Net v3 sadržavao je sve nadogradnje navedene za Inception v2, a osim toga koristio je i

komponente za regulaciju koje se dodaju kako bi spriječile preklapanje.

D. Xception

Xception je proširenje Inception arhitekture koja zamjenjuje standardne Inception module s duboko razdvojitim konvolucijama. Xception je konvolucijska neuronska mreža ("ekstremna verzija" neuronske mreže Inception) koja je trenirana na više od milijun slika iz baze podataka ImageNet. Mreža je duboka 71 sloj i može klasificirati slike u 1000 kategorija objekata, kao što su tipkovnica, miš, olovka i mnoge životinje. Kao rezultat toga, mreža je naučila bogate prikaze značajki za širok raspon slika. Mreža ima veličinu ulaza slike od 299×299 .

Pomoću modela Xception možemo klasificirati nove slike i utrenirati mrežu na novom zadatku klasifikacije. Koristi se za duboko učenje.

E. VGG16

VGG je duboka konvolucijska neuronska mreža za prepoznavanje objekata razvijena i trenirana od strane Oxfordove Visual Geometry Group (VGG) koja je postigla jako dobru performansu na Image Net Large Scale Visual Recognition Challenge-u 2014. godine. Zauzeli su prvo mjesto zadatka lokalizacije i drugo mjesto zadatka klasifikacije. Glavni doprinos VGG mreže je pokazati da točnost klasifikacije ili lokalizacije može biti poboljšana povećanjem dubine konvolucijske neuronske mreže usprkos korištenju malih filtera u slojevima.

VGG16 mreža je mreža sa 16 slojeva i karakterizira ju jednostavnost. Slika prolazi kroz niz konvolucijskih slojeva gdje se koriste 3×3 filteri, slojevi sažimanja s 2×2 filtrom i potpuno povezani slojevi na kraju. Prostorna dimenzija volumena smanjuje se prolaskom kroz dublje slojeve, ali se zbog većeg broja konvolucijskih slojeva dubina volumena povećava.

Na slici 3 vidimo pojednostavljeni prikaz modela VGG16 dok na slici 4 vidimo potpuni pregled modela na razini slike.

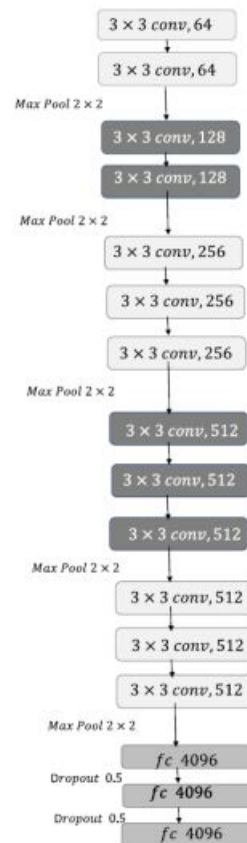
VGG16 je duboka mreža sa 138 milijuna parametara, to je poboljšanje koje se dobilo korištenjem tri 3×3 konvolucijska sloja umjesto naprimjer jednog 7×7 sloja. Broj parametara se smanjio sa $7^2 C^2$ na $3 \cdot (3^2 C^2)$, gdje je C broj kanala sloja.

IV. PRIKAZ REZULTATA

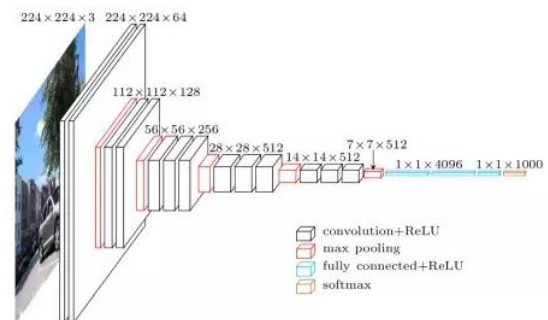
Originalni ImageNet skup ima dosta različitih pasmina pasa pa u konvolucijskim neuronskim mrežama koristimo predtreinirane ImageNet težine. Iako su resursi Kernala nedavno povećani, još uvijek ne možemo trenirati konvolucijske neuronske mreže (CNN) bez grafičke kartice (GPU). Originalni skup ImageNet ima dosta različitih klasa pasa tako da možemo prenamijeniti konvolucijske neuronske mreže s pretreiniranim težinama ImageNet-a. Predviđanje je tako mnogo brže (manje od 1s po slici), što omogućuje izvođenje značajnih eksperimenata s Kaggle kernelima.

Koristimo Tensorflow backend i Keras Pretrained Models skup podataka. Moramo kopirati pretreinirane modele u direktorij cache ($\sim /.keras/models$) gdje ih keras traži.

Slika 3. Pojednostavljeni prikaz

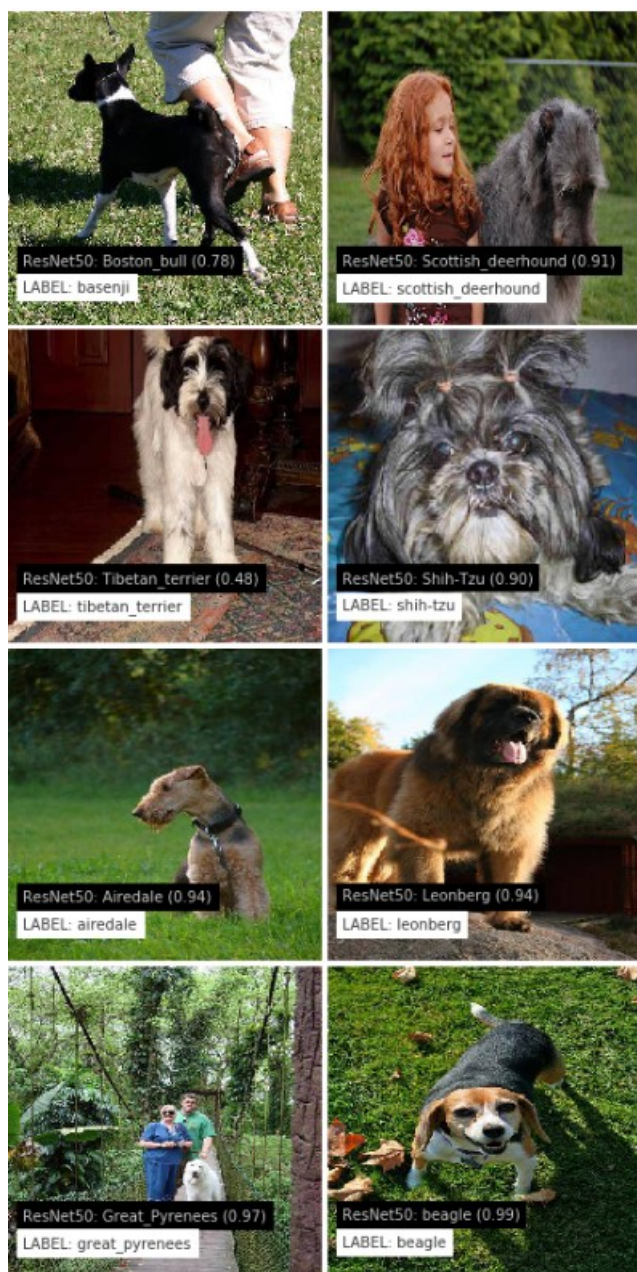


modela VGG16

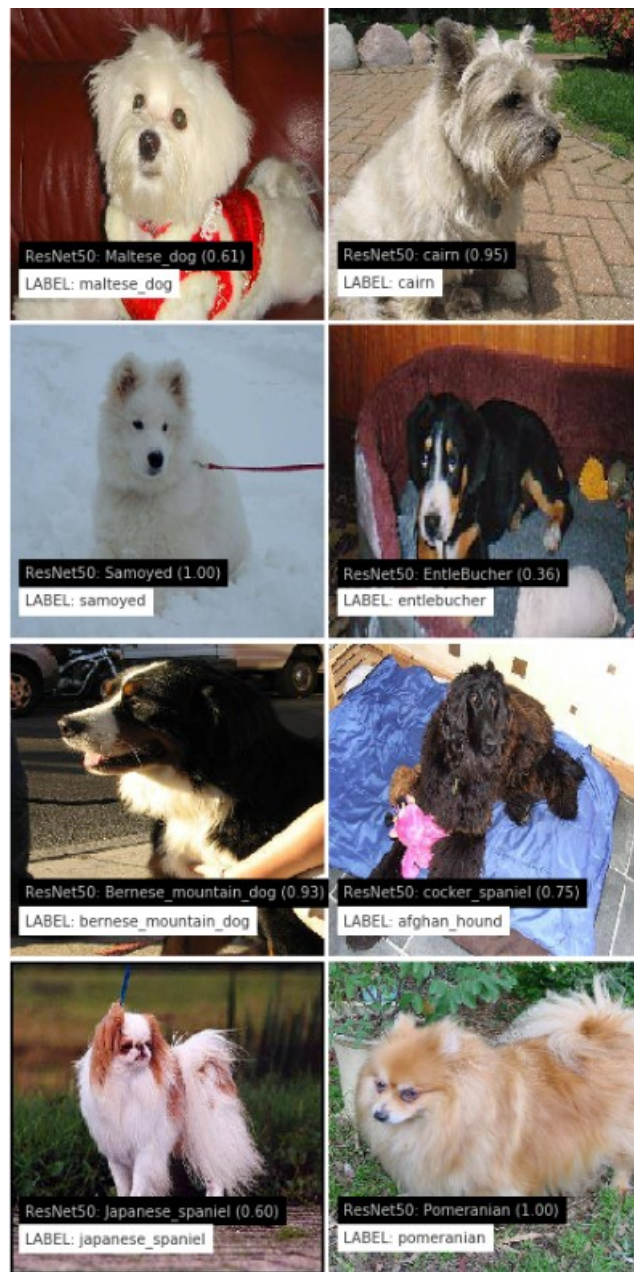


Slika 4. Potpuni pregled modela VGG16 na razini slike

Korištenje svih slika trajalo bi više od jednog sata pa smo se usredotočili na 16 najčešćih pasmina.



Slika 5. Ispis modela ResNet50



Slika 6. Ispis modela ResNet50

Prilikom primjene ResNet50 modela dobivamo predviđanje od 81% točnosti na prethodno prikazanih 16 slika. Koristeći VGG16 točnost raste na 92%, a gubitak je 0.35.

Primjenom Inception modela dobivamo točnost od 97% za prvih 16 klasa i gubitak od 0.08, dok za Xception model dobivamo 98% točnosti i 0.07 gubitka.

Za svaki od modela najprije smo izvukle bottleneck značajke na koje je zatim primijenjen pojedini model. Kombiniranje predikcija više modela koji su napravljeni s istim ili različitim algoritmom na istim ili različitim podacima s ciljem poboljšavanja predikcije u odnosu na jedan model nazivamo ansambl. Kombiniranje modela dobro funkcionira jer optimalna funkcija cilja ne mora biti ni jedan individualni

klasifikator, no može se bolje aproksimirati usrednjavanjem većeg broja individualnih modela.

Budući da nijedan pristup od prije ne trenira model na skupu za treniranje koji nam je dan nego samo koristi već pretrained modele trenirane na ImageNet datasetu provele smo VGG16 pretrained model i trenirale ga na danim slikama za trening. Sve slike, trening i test skupa, smo najprije prebacile u odgovarajuću dimenziju 224×224 . Pomoću VGG16 modela izvlače se značajke na trening skupu. Skup smo podijelile na 90% slika za treniranje i ostalih 10% za validaciju. Treniramo u 10 epoha. U svakoj od epoha validacija gubitka pada, a točnost raste. No, u zadnjoj desetoj epohi dođemo samo do otprilike 50% točnosti pri čemu je gubitak jednak 3.5895. Također smo izvukle značajke testnog



Slika 7. Primjer krive predikcije

skupa te na njih primijenile predikciju treniranog modela. Rezultati se zapisuju u excel tablicu. Točnost je poprilično niska pa smo izabrale nasumično 8 slika iz testnog skupa i za svaku od njih po rezultatima gledale za koju pasminu model daje najveću vjerojatnost. Od tih 8 slika nijedna nije bila točno klasificirana. Npr. za sliku 8 model je 100% siguran da nije german shepherd, irish setter ni irish terrier dok je 100% siguran da je eskimo dog. Za sliku 9 je 87.42% siguran da je bull mastiff i 0.018% da je groenendael.



Slika 8. Primjer krive predikcije modela VGG16



Slika 9. Primjer krive predikcije modela VGG16

V. OSVRT NA DRUGE PRISTUPE

Na Kaggle-ovom natjecanju sudjelovalo je 1286 timova te na stranicama natjecanja postoji velik broj gotovih rješenja zadanog problema s raznim točnostima. Također, postoji

još nekoliko dodatnih istraživanja na istu temu s drugim skupovima podataka. Metode kojima se pokušalo riješiti problem su logistička regresija, metoda potpornih vektora (SVM), K-najbližih susjeda, konvolucijske neuronske mreže (CNN). Y. Aussat u svom istraživačkom radu koristi nekoliko različitih metoda i kombinacije istih. Prva od njih je CNN metoda s kojom dobiva točnost od otprilike 35%, nakon toga kombinira CNN s data augmentation metodom čime se točnost povećava na 55%. Koristeći pretrained modele VGG16, VGG19, ResNet50, Inception-V3 i Xception točnost poraste na 85%.

LaRow, Mittle, Singh koriste u svom istraživanju SVM s linearnim kernelom. Točnost predviđanja pasmina iz prvog pokušaja je 52%, te 90% vremena pogođena pasmina je u 10 najboljih predikcija.

VI. ZAKLJUČAK

Promatrajući rezultate velikog natjecanja u strojnom učenju, zaključujemo da se vrhunski rezultati postižu skupom modela, a ne jednim modelom. Budući da svaki model ima svoje nedostatke, slaganjem različitih modela koji predstavljaju različite hipoteze o podacima možemo pronaći optimalnu hipotezu koja nije u hipoteznom prostoru modela iz kojih je izgrađena cjelina. Time se poboljšava točnost u odnosu na individualne modele, jer reduciraju ili varijancu ili bias (ili čak oboje). Nedostatak ansambla je zahtjevnost (vremenski i memorijski) koja proizlazi iz učenja većeg broja modela. Tako se, dobrim odabirom skupa podataka, dobrom redukcijom dimenzionalnosti i kombinacijom više modela, naizgled težak zadatak prepoznavanja pasmine pasa može riješiti dosta uspješno, čak sa uspješnosti predikcije većom od 90%.

LITERATURA

- [1] Kaggle, Dog Breed Identification, <https://www.kaggle.com/c/dog-breed-identification/>
- [2] LaRow W., Mittl B., Singh V., Dog Breed Identification [https://web.stanford.edu/class/cs231a/prev_projects_2016/output%20\(1\).pdf](https://web.stanford.edu/class/cs231a/prev_projects_2016/output%20(1).pdf)
- [3] Wikipedia, Multiclass classification, https://en.wikipedia.org/wiki/Multiclass_classification
- [4] Y. Aussat, Dog Breed Identification, http://blizzard.cs.uwaterloo.ca/iss4e/wp-content/uploads/2017/10/yerbol.aussat_cs698_project_report.pdf
- [5] <https://www.kaggle.com/gaborfodor/dog-breed-pretrained-keras-models-lb-0-3>
- [6] <https://www.kaggle.com/bkkaggle/xception-inceptionv3-ensemble-methods>
- [7] <https://becominghuman.ai/classifying-images-with-vggnet-resnet-inception-and-xception-usingpython-and-keras-c0bf77af923a>

- [8] <https://github.com/girishsg24/Dog-Breed-Identification>
- [9] <https://www.learnopencv.com/keras-tutorial-using-pre-trained-imagenet-models/>
- [10] <https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202>
- [11] <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>