

**AUTOMATSKO GENERISANJE SKUPA PODATAKA ZA TRENIRANJE MODELA ZA
AUTOMATSKO PREPOZNAVANJE OSOBE NA SLICI****AUTOMATED DATASET GENERATION FOR PERSON-IDENTIFICATION MODEL
TRAINING**

Noemi Sabadoš, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – Kvalitetni skupovi podataka su retko dostupni, što otežava treniranje kvalitetnih modela mašinskog učenja. Cilj ovog rada jeste generisanje skupa podataka koji bi mogao biti korišćen za obuku modela mašinskog učenja čiji bi cilj bio da uporedi slike ljudi sa slikama poznatih ličnosti i prepozna sličnosti u izgledu. Generisanje skupa podataka se vrši automatski, na osnovu unapred zadatih kriterijuma. Svrha kriterijuma je da učine skup podataka raznovrsnim, sa ciljem poboljšanja generalizacije nad njime obučениh modela. Podržani kriterijumi uključivanja slika u skup podataka su sledeći: (1) da li se osoba smeši ili ne, (2) specifikacija smera pogleda, (3) činjenica da li osoba ima: bradu, šiške, kosu, (4) da li osoba nosi naočare ili kapu i (5) da li osoba rukom zaklanja lice. Prikupljeni skup podataka sadrži slike o 100 glumca, koje su skinute sa interneta prema zadatim kriterijumima.

Ključne reči: skup podataka, generisanje skupa podataka, skup podataka o licu, skup podataka o slavim ličnostima, detekcija atributa lica

Abstract – The basis for successful training of quality machine learning models are the datasets, but quality ones are hard to find in most cases. This paper aims to generate a dataset that could be used in machine learning as a training set for comparing celebrities with ordinary people. The generation process is automatic and based on predetermined criteria, aimed to make the dataset diverse. Dataset diversity should enhance the generalization of the trained models. Among the supported criteria are: (1) whether the person is smiling, (2) the direction of the person's gaze, (3) whether the person has a beard, bang, or hair, (4) whether the person wears glasses or a hat, or (5) is the person's hand in front of the face. The generated dataset contains publicly available images of 100 actors.

Keywords: dataset, generating dataset, face dataset, celebrity dataset, face attribute detection

1. UVOD

U oblasti informacionih tehnologija, mašinsko učenje je oblast koji se najbrže razvija. Korektan rad sistema zasnovanih na mašinskom učenju u velikoj meri zavisi od

skupa podataka korišćenog za treniranje modela. Loš kvalitet skupova podataka uzrokuje loš rad sistema.

Postoje javno dostupni skupovi podataka za prepoznavanje ličnosti na slici. Međutim, ni jedan se ne fokusira na anotaciju smera pogleda osobe na slici. Pri tome, ovakve anotacije bi bile od velike pomoći za aplikacije čiji je cilj određivanje sličnosti lica prikazanog na dve različite slike. Cilj ovog rada je napraviti skup podataka koji će ublažiti ovaj i slične nedostatke. Pored smera pogleda, razmatrane su još neke osobine lica koje mogu isto da doprinesu detekciji sličnosti:

- prepoznavanje osmeha
- prepoznavanje brade
- prepoznavanje šiški
- prepoznavanje ćelavosti
- prepoznavanje naočara
- prepoznavanje kape i/ili kačketa
- prepoznavanje ruku koje zaklanjaju lice.

U drugom poglavlju će biti predstavljeni radovi koje su imali uticaj na rešenje predstavljeno u ovom radu. U trećem poglavlju će biti opisane faze generisanja skupa podataka. O rezultatima će biti reč u četvrtom poglavlju. U petom poglavlju prikazan je rezime rada i pravci budućeg razvoja.

2. POSTOJEĆA REŠENJA

Postoji mnogo javno dostupnih skupova podataka sa slikama javnih ličnosti, ali je fokus ovog rada na skupovima podataka koji sadrži slične karakteristike razmatrane u ovom rešenju. Odnosno, skup podataka bi trebao da sadrži podatke o smeru glave, da osobe na slikama budu poznate ličnosti i da, osim toga, sadrže više osobina koje mogu pomoći u definisanju sličnosti lica na dve različite slike. Po znanju autora, trenutno ne postoji ni jedan skup podataka koji zadovoljava sve ove kriterijume.

Jedan od najpoznatijih javno dostupnih skupova podataka je *Labeled Faces in the Wild* (LFW) [1]. Skup podataka je napravljen za prepoznavanje lica. Verifikacija lica ili drugačije *pair matching*, je metoda u kome upoređujemo dve slike i određujemo da li je ista osoba prisutna na obe slike. Skup sadrži 13.233 slika o 5.749 osoba. Za većinu osoba prisutna je samo jedna slika, ali je za 1.680 imena prisutno 2 ili više slika. Dodatna važna osobina jeste da su slike snimane u prirodnoj okolini, dakle, ne u veštački pripremljenim uslovima. Cilj da su prikupljene slike iz svakodnevnice, bez uticaja okoline ili poboljšanja.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bila dr Jelena Slivka, vanr. prof.

Rad „Names and faces in the news“ [2] je dodao slike iz online novina. Tako slike iz LFW sadrže slike o osobama, koje su se pojavile u ovim člancima, ali uglavnom se radi o slikama poznatih ljudi koje ne sadrže nikakve dodatne informacije poput smera pogleda, osmeha, itd.

Skup podataka *CelebFaces Attributes Dataset* (CelebA) [3] sadrži slike poznatih ličnosti na kojima je označeno mnogo različitih osobina. Ovaj skup podataka je najbliži onom predstavljenom u ovom radu, sa izuzetkom smera pogleda, koji nije uključen u [3]. Skup podataka sadrži 10.177 osoba na više od 200 hiljada slika. Pored toga, na slike su dodate anotacije sa 5 lokacijskih obeležja i 40 binarnih atributa.

Među anotiranim osobinama nalazi se postojanje šiški, osmeha, naočara, i sl. U [3] kreiran je skup podataka na osnovu slika iz *CelebFaces+* [4], pri čemu su kod svake manuelno dodate anotacije za 40 atributa. Nažalost, smer glave nije bio između ovih osobina. Cilj rada [5] jeste da se kreira skup podataka koji će služiti za treniranje softvera za prepoznavanje lica, procenu smera pogleda ili lokalizaciju tačaka lica. Konačan skup sadrži 367.888 slika od 8.277 osoba. Slike nisu iz kontrolisane okoline, već iz svakodnevnice. Skup podataka sadrži i anotaciju graničnog okvira lica, koja je verifikovana manuelno. Za svaku sliku su dodate i lokacije ključnih tačaka, pozicija (rotacije oko z, y i x ose) i informacije o polu.

Ovaj skup već sadrži smer pogleda, međutim, u drugoj formi od očekivane, to jest, nisu zabeleženi uglovi rotacija i definisani smerovi. Ne sadrži nikakve dodatne informacije koje bi mogli biti iskorišćene u definiciji sličnosti dve osobe na slici.

3. GENERISANJE SKUPA PODATAKA

Prvi korak za generisanje skupa podataka prikazanog u ovom radu je kolekcija liste poznatih osoba. Za ovo je korišćena lista¹ od 1000 poznatijih glumaca. Za prvih 100 glumaca sačuvano je oko 200 slika sa *Bing Image Downloader*². Sledeći korak je bio filtriranje slika (više detalji u poglavlju 3.1). Nakon kolekcije slika, slede faze generisanje skupa podataka. Na kraju su manuelno uklonjene pogrešne anotacije.

Koraci generisanja skupa podataka se mogu videti na slici 1. Prvo je detektovano lice na slici pomoću kombinacije *Histogram of Oriented Gradients* (HOG) obeležja i SVM (*Support Vector Machine*) modela [6]. Nakon toga su filtrirane slike gde oči nisu otvorene. Nakon toga sledi detekcija: osmeha, smera glave, šiški, brade, ćelavosti, naočara, kape i ruke ispred lica. Za svaku od navedenih osobina su detektovane slike iz filtriranog skupa. Nakon što su za datu osobu pronađene 3 slike po gore navedenim kriterijumima, potraga za slikama ove osobe se završava.



Slika 1. Koraci kreiranja skupa podataka

3.1. Filtriranje slika

Zbog velike količine prikupljenih podataka, filtriranje je trebalo biti automatsko. Ovaj korak je neophodan jer su prikupljene slike pomoću *Bing* pretraživača koji ne garantuje da je tražena osoba na slici.

Prvi korak za filtriranje je detekcija lica. Ako broj pronađenih lica nije jedan, slika je uklonjena. Sledeći korak je upoređivanje lica na slici sa referentnom slikom (za koju znamo da sadrži traženu osobu). Cilj upoređenja lica je da se filtriraju slike na kojima je lice osobe jako različito, odnosno, da se uklone direktorijumi koji verovatno sadrže slike različitih osoba. Za referentnu sliku uzeta je profilna slika od *TMDb*³ stranice. Upoređivanje slika sa referentnom je vršeno pomoću *Openface-a* [7], pri čemu su filtrirane slike koje su bile jako različite. Poslednji korak je bio brisanje duplikata. Ovo je izvršeno *scalable template matching* metodom.

3.2 Prepoznavanje zatvorenih očiju

Inspiraciju za prepoznavanje zatvorenih očiju je dao rad [8] u kome je opisana metoda za prepoznavanje treptanja na video snimcima. Otvorenost očiju se vrši merenjem *EAR* (*eye aspect ratio*). U ovom cilju izvršena je klasifikacija sa sledećim ulaznim podacima: 3 normalizovane udaljenosti tačaka očiju, EAR, pol i rasa. Najbolji rezultat je postignut primenom *Gradient Boosting* modela.

3.3 Prepoznavanje osmeha

Konvolucione neuronske mreže (*Convolutional Neural Networks*, CNN) su najefektivnije za klasifikaciju slike i stoga su često korišćene za prepoznavanje osmeha [9, 10]. Ovo je bio razlog za odabir CNN modela za prepoznavanje osmeha, na osnovu *Keras mnist_cnn.py*⁴ primera. Ova arhitektura trenirana je na slikama iz *SMILEsmileD*⁵ skupa, i korišćena je za prepoznavanje osmeha.

3.4 Prepoznavanje smera glave

Za prepoznavanje smera glave isprobano je više načina, ali je na kraju kreirana posebna neuronska mreža. Mreža ima 15 ulaznih neurona, koji kao ulaz primaju udaljenosti ključnih tačaka lica. Mreža ima jedan skriveni sloj sa 60 neurona. Aktivaciona funkcija je *softmax*. Za treniranje je korišćen *Head Pose Image Database* [11] skup podataka. Skup sadrži 2 790 slika o 15 osoba. Smer pogleda je određen sa 2 ugla. Vertikalni ugao označava levi i desni smer, a horizontalni gore i dole. Mnogo slika ovog skupa podataka je moralo biti odbačeno, jer ovde predstavljena aplikacija nije detektovala lica na slikama uslikanim iz profila. Na kraju je za treniranje korišćeno 792 slika.

Prilikom testiranja detekcije, nije primećena nijedna slika gde je osoba gledala na dole, ali je bilo mnogo slučajeva da je za lice sa centralim smerom pogleda prediktovano da osoba gleda dole. Zbog ovoga, svaki dole prediktovan smer je nadalje računat kao centralni. Smer pogleda na gore je takođe redak, i preciznost detekcije je svega 0.49,

¹ www.imdb.com/list/ls058011111/

² github.com/gurugaurav/bing_image_downloader

³ www.themoviedb.org

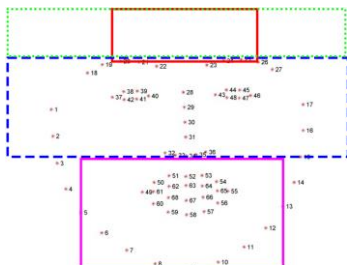
⁴ github.com/keras-team/keras/blob/

⁵ github.com/hromi/SMILEsmileD

tako da i on odbačen. Na kraju, aplikacija traži smerove levo, desno i centriran.

3.5 Prepoznavanje brade

Za prepoznavanje korišćena je ista arhitekturu kao kod prepoznavanje osmeha (poglavljja 3.2). Za treniranje su korišćeni iseći slika iz *Celeba* [3]. Iseći su predstavljali područja usana (slika 2), i model je treniran na ovim slikama. Detekcia brade je izvršena samo u slučaju da je osoba na slici muškarac.



Slika 2. Isecanje za različite detekcije – čelavosti: crvena puna linija, naočare: plava isprekidana linija, kapa: zelena tačkasta linija, brada: ljubičasta puna linija

3.6 Prepoznavanje šiški

Prepoznavanje šiški je izvršeno na sličan način kao prepoznavanje brade (poglavljje 3.5). Jedina razlika je da se ovde za treniranje koriste slike sa tamnom kosom, jer su svetle kose zbunjivale detekciju. Najvažniji deo je bio izbeći *false positive* (FP) u trening skupu, tako da su slike koje su bile rizičnije izbačene iz trening skupa. Rizičnije slike su one na kojima aplikacija prepoznaje i čelavost (poglavljje 3.7) i šiške na istoj slici. Takođe su odbačene slike gde je prepoznata kapa (poglavljje 3.9).

3.7 Prepoznavanje čelavosti

Problem prepoznavanja čelavosti je sličan problem kao detekcija brade ili šiški, tako da je opet korišćena ista CNN arhitektura. I u ovom slučaju su korišćene slike i labela iz *Celeba* [3] skupa. Međutim, sada je umesto da se koristi celo lice za treniranje korišćen samo deo lica između spoljnih uglova oka i od obrva do vrha okvirne glave (slika 2). Slično kao i kod prepoznavanja šiški, i ovde su izbačene slike na kojima su detektovane šiške ili kape. Aplikacija je detektovala čelavost samo ako je osoba na slici bio muškarac.

3.8 Prepoznavanje naočara

Za prepoznavanje naočara isprobane su dve metode. Obe metode treba da prepoznaju naočare u ciju dodavanja takvih slika u skup podataka.

Prvi pristup je koristio istu CNN arhitekturu za prepoznavanje drugih aspekata lica (poglavljje 3.2, 3.5-3.7). Takođe je korišćen *Celeba* [3] skup podataka, gde je isećeno područje koja počinje od nosa i ide do najviše tačke obrva (slika 2).

Drugi pristup je detekcija objekta pomoću *Mask R-CNN* [12]. Za ovaj pristup sačuvane su slike za treniranje iz *Open Images Dataset* [13] sa naočarima za sunce, gde je dodat granični okvir za svake naočare. Nakon toga je

istreniran *Mask R-CNN* model. Dodat je kriterijum da naočare trebaju biti prepoznate sa minimalnom verovatnoćom 0,9 u da se pronađeno područje nalazi u području lica.

3.9 Prepoznavanje kape i/ili kačketa

Mask R-CNN [12] pristup je isproban i na problemu prepoznavanja kape, ali nije bio uspešan. Ovde je zato ipak korišćena jednostavna CNN arhitektura (poglavljje 3.2, 3.5-3.8). Prvi korak je bila anotacija slika iz *Open Images Dataset* [13], gde je isećena regija od obrva do vrha lica (slika 2). Nad takvim slikama je trenirana CNN.

3.10 Prepoznavanje ruku ispred lica

Prvi korak za detekciju ruku koje zaklanjaju lice je bio detekcija ruke. Za procenu tačke ruke je korišćen postojeći model *OpenPose* [14]. Model procenjuje 21 tačku za ruku. Aplikacija je trebala samo da odluči da li je ta ruka ispred lica ili ne. Prvo je isećeno lice iz slike, nakon čega se vrši predikcija tačkaka ruke. Ako je aplikacija na isećenoj slici lica našla bar 5 tačkaka, slika je anotirana da sadrži ruku ispred lica.

4. REZULTATI I SKUP PODATAKA

U ovom poglavlju prvo će biti predstavljeni rezultati za svako prepoznavanje na jednom velikom skupu. Nakon toga će biti predstavljen izgenerisani skup podataka.

4.1. Rezultati prepoznavanja

Da testiram svoje detekcije, želela sam da kreiram jedan test skup podataka koji će biti sličan onome koji je cilj bio generisati. Za filtriranje su korišćene profilne slike iz *TMDb* stranice. S obzirom na to da se tamo se može naći više od jedne slike, sve su sačuvane. Ukupno sam uspela da sakupim 1 921 sliku, od čega su 1 313 slika žena, a 608 slika muškaraca. Za pojedinačne slike sistem nije mogao da pronađe lice, neke tačke na licu i slično. Te slike su preskočene za određenu detekciju. Anotacije skupa podataka sam uradila manualno.

U tabeli 1 mogu se videti rezultati za evaluaciju. Budući da kreirani test skup sadrži jako mnogo negativnih primera, mikro f1 mera nije toliko značajna. Umesto toga, cilj je maksimizovati preciznost. Prikazani rezultati ne sadrže prepoznavanje zatvorenih očiju jer u ovom skupu nije postojao ni jedna slika gde bi oči bile zatvorene. Od 1 778 slika, samo sam 5 puta dobila rezultate da oči nisu otvorene.

U rezultatima prikazanim u tabeli 1 može se videti da prepoznavanje osmeha, smer glave i brade nema gotovo nijedan FP rezultat. Prepoznavanje šiški je nešto lošije, ali broj FP nije značajan. Najveći problem sa ovom detekcijom jeste da sistem nije prepoznao mnogo slika sa plavim šiškama.

Za prepoznavanje čelavosti i naočara gotovo za svaku treću sliku nije bio dobar rezultat. Detekcija čelavosti ima problem sa sedam kosom i visokim čelom. Rezultati za prepoznavanje kape i ruke ispred lica su takođe nezadovoljavajući zbog velikog broja FP.

Tabela 1: Rezultati prepoznavanja različitih osobina slike

Prepoznavanje	Mikro f1 mera	Preciznost
prepoznavanje osmeha	0.77	0.97
prepoznavanje smer glave	0.95	0.95
prepoznavanje brade	0.83	0.98
prepoznavanje šiški	0.84	0.83
prepoznavanje čelavosti	0.92	0.59
prepoznavanje naočara	0.98	0.62
prepoznavanje kape i/ili kačketa	0.98	0.35
prepoznavanje ruku ispred lica	0.95	0.29

4.2. Izgenerisan skup podataka

Ceo proces generisanje skupa podataka je automatski, pomoću implementirane aplikacije. Kao izvor slika, korišćen je *Bing*, što nije bila najbolja opcija. Od 200 sačuvanih slika, nakon filtriranja je preostalo svega oko 20-30 slika po osobi. Dobavljene slike sadržale su mnogo duplikata, slika više ljudi ili slike bez ljudi, slike slabog kvaliteta, itd. Nakon generisanja skupa, manuelno su eliminisane pogrešno prediktovane slike. Zbog lošeg izvora, konačan skup je sadržao samo 716 slika. Za svaku osobu, za svaku detekciju možemo da imamo maksimum tri slike. U tabeli 2. mogu se videti statistički podaci o generisanom skupu podataka.

Tabela 2: Statistički podaci i broj slike za različite kriterijume

	Žena	Muškarac	Ukupno
Broj osoba	60	40	100
Broj slika	450	266	716
Osmeh	179	113	292
Smer centar	177	118	295
Smer levo	47	21	68
Smer desno	45	13	58
Šiške	150	28	178
Brada	-	80	80
Čelavost	-	45	45
Naočare	5	19	24
Kapa i/ili kačket	8	3	11
Ruka ispred lica	30	6	36

5. ZAKLJUČAK

U ovom radu predstavljena je aplikacija za generisanje skupa podataka na osnovu više kriterijuma kao što su smer glave, postojanje osmeha, brade, šiški, čelavosti, naočara, kapa i ruku ispred lica. Prepoznavanje ovih kriterijuma vršeno je automatski. Najbolji rezultati su dobijeni na prepoznavanju osmeha, smer glave i brade (preciznost preko 0.95). Najgore rezultate imaju prepoznavanje kape i ruku ispred lica (preciznost oko 0.3).

Pomoću implementirane aplikacije generisan je novi skup podataka. Izvor slika je bio *Bing*, koji nije predstavljao najbolju opciju. Konačan skup ima 716 slika od 100 glumaca.

Budući plan je ispraviti prepoznavanja sa lošijim rezultatima, i generisati skup podataka korišćenjem drugog izvora.

6. LITERATURA

[1] Huang, Gary B., et al. "Labeled faces in the wild: A database for studying face recognition in unconstrained environments." 2008.

[2] Berg, Tamara L., et al. "Names and faces in the news." Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.. Vol. 2. IEEE, 2004.

[3] Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE international conference on computer vision. 2015.

[4] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[5] Bansal, Ankan, et al. "Umdfaces: An annotated face dataset for training deep networks." 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017.

[6] Déniz, Oscar, et al. "Face recognition using histograms of oriented gradients." Pattern recognition letters 32.12 (2011): 1598-1603.

[7] Amos, Brandon, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. "Openface: A general-purpose face recognition library with mobile applications." CMU School of Computer Science 6.2 (2016).

[8] Soukupová, Tereza, and Jan Cech. "Real-time eye blink detection using facial landmarks." 21st Computer Vision Winter Workshop. 2016.

[9] Chen, Junkai, et al. "Smile detection in the wild with deep convolutional neural networks." Machine vision and applications 28.1-2 (2017): 173-183.

[10] Zhang, Kaihao, et al. "Facial smile detection based on deep learning features." 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, 2015.

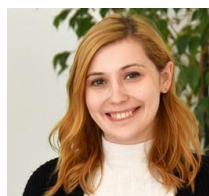
[11] Gourier, Nicolas, Daniela Hall, and James L. Crowley. "Estimating face orientation from robust detection of salient facial features." ICPR International Workshop on Visual Observation of Deictic Gestures. 2004.

[12] He, Kaiming, et al. "Mask r-cnn." Proc. of the IEEE international conference on computer vision. 2017.

[13] Kuznetsova, Alina, et al. "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale." arXiv preprint arXiv:1811.00982 (2018).

[14] Simon, Tomas, et al. "Hand keypoint detection in single images using multiview bootstrapping." Proc. of the IEEE conference on Computer Vision and Pattern Recognition. 2017.

Kratka biografija:



Noemi Sabadoš rođena je 1995. godine u Vrbasu. Osnovne akademske studije završila 2018. godine na Fakultetu Tehničkih Nauka Novi Sad, na kom brani i master rad 2020. godine iz oblasti Elektrotehnike i računarstva – Softversko inženjerstvo i informacione tehnologije.
kontakt: nomisabi@gmail.com