



SISTEM ZA PREPOZNAVANJE I PREPORUČIVANJE ODEVNIH PREDMETA SA VIDEO ZAPISA

SYSTEM FOR CLOTHES RECOGNITION AND RECOMMENDATION FROM A VIDEO

Nebojša Basarić, *Fakultet tehničkih nauka, Novi Sad*

Oblast – RAČUNARSTVO I AUTOMATIKA

Kratak sadržaj – Zadatak ovog rada je specifikacija, implementacija i verifikacija sistema za detekciju, prepoznavanje i preporučivanje odevnih predmeta upotrebom složenih arhitektura konvolucionih neuronskih mreža i algoritama u oblasti kompjuterske vizije. Za obučavanje modela korišćen je javno dostupan skup podataka objavljen od strane univerziteta u Hong Kongu.

Ključne reči: sistemi za preporučivanje proizvoda, analiza slike, kompjuterska vizija, prepoznavanje i detekcija, E-commerce

Abstract – This paper describes the specification, implementation, and verification of systems for clothes detection, recognition, and recommendation using the complex convolutional neural network architectures and algorithms in the field of computer vision. DeepFashion dataset, published by the University of Hong Kong, was used to train the model.

Keywords: recommender systems, image analysis, computer vision, recognition and detection, E-commerce

1. UVOD

Online marketing se konceptualno razlikuje od ostalih vidova marketinga time što omogućava direktnu komunikaciju između prodavca i krajnjeg korisnika. Danas je online marketing najbrže rastući segment online prodaje. Najveća razlika između online i tradicionalnog marketinga je obim interakcije između kupca i prodavca. Ovakav vid reklamiranja omogućava kupcima da sami utiču na sadržaj reklama koji će im se prikazivati, kao i prodavcima da svoj proizvod reklamiraju samo potencijalnim kupcima sa određenom sferom interesovanja.

U prethodnoj deceniji, nastala je dramatična izmena u načinu na koji kupci realizuju kupovinu. Iako kupci i dalje fizički posećuju prodavnice, postali su mnogo sigurniji u kupovinu preko interneta. E-commerce omogućava ljudima iz malih mesta pristup svim proizvodima u velikim gradovima koji su pre bili nedostupni.

Studija [1] prikazuje najprodavanije tipove proizvoda preko interneta. U ovoj studiji, od 100 kupaca - 33% kupilo je knjige, 23% kupaca koristilo je internet za dobavljanje elektronskih komponenti, 18% kupovalo je odeću, 12% obuću, i 14% ostale tipove proizvoda.

Na osnovu priložene studije može se zaključiti da je kategorija odeće zajedno sa obućom skoro najzastupljenija kategorija prodaje preko interneta. To nas dovodi do motivacije za ovaj rad, odnosno kako dodatno poboljšati interakciju između kupaca i prodavaca u pogledu prodaje odeće. Često nam se dešava da na filmu ili spotu ugledamo osobu koja nosi odeću koja nam se dopada. Prvo što poželimo je da saznamo gde da pronađemo istu takvu ili bar sličnu. Namena sistema predstavljenog u ovom radu jeste da krajnjem korisniku olakša pronalazjenje i kupovinu najpogodnijeg odevnog predmeta, sličnog onom koga zapazi na filmu ili u spotu.

U sistemu predstavljenom u ovom radu je kao početni korak vršena detekcija i prepoznavanje tipa odevnog predmeta iz video zapisa. Prepoznavanje se vrši na osnovu jednog frejma videa koristeći konvolucionu neuronsku mrežu (*Convolutional Neural Network, CNN*) *Faster-R CNN* arhitekture. Sistem prepoznaje sledeće klase: majica, haljina, sako, farmerke, sportska majica, venčanica, jakna, suknja i šorc. Nakon prepoznavanja tipa odevnog predmeta proizvodi koji su u ponudi iz te klase se rangiraju po različitim kriterijumima koji definišu sličnost proizvoda: prosečna vrednost boje reprezentovane u dva različita modela i histogram koji označava teksturu. Izračunavanje kriterijuma rangiranja zasnovano je na primeni tehnika kompjuterske vizije.

Za potrebe demonstracije rada ovog sistema implementirana je veb aplikacija koja prikazuje različite video klipove sa interneta i omogućava korisniku da u bilo kom momentu pritiskom na dugme pokrene pretraživanje prodavnica za proizvodom koje se u tom trenutku prikazuje na video klipu.

Implementacija sistema podeljena je na tri modula:

- Modul za Detekciju i prepoznavanje odevnog predmeta sa frejma video zapisa
- Modul za Preporučivanje proizvoda
- Korisnički interfejs

2. METODOLOGIJA

Modeli za detekciju i prepoznavanje odevnih predmeta definisani su, podešavani, trenirani i testirani koristeći *Python* [2] programski jezik i *TensorFlow* [3] biblioteku koja predstavlja platformu otvorenog koda *state-of-the-art* algoritama mašinskog učenja.

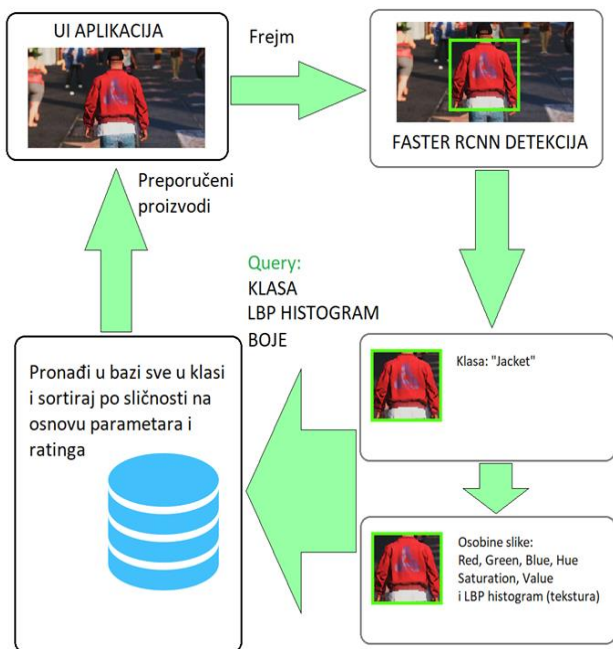
Za implementaciju detaljne analize i segmentacije slike korišten je C# programski jezik na .NET platformi sa *EmguCV* [4] wrapper-om koji predstavlja biblioteku koja omogućuje pozivanje *OpenCV* [5] funkcija iz .NET

NAPOMENA:

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

okruženja. *OpenCV* (*Open Source Computer Vision Library*) predstavlja biblioteku otvorenog koda za kompjutersku viziju i mašinsko učenje.

Za potrebe prikaza rada sistema implementirana je *web* aplikacija, i to koristeći *C#* programski jezik na *.NET* platformi za implementaciju back-end aplikacije, a *JavaScript* i *jQuery* [6] su korišćeni za implementaciju klijentske aplikacije.



Slika 2.1. Arhitektura rešenja

Na slici 2.1. je predstavljena skica arhitekture rešenja predloženog u ovom radu gde se mogu prepoznati sva tri modula koja čine sistem: Modul za detekciju i prepoznavanje odevnih predmeta, modul za preporučivanje n najbližnjih proizvoda i modul koji predstavlja korisničku aplikaciju koja prikazuje rad sistema.

2.1. Modul za prepoznavanje i detekciju odevnih predmeta

Zadatak modula za prepoznavanje i detekciju odevnih predmeta jeste da prihvati ulaznu sliku sa korisničkog interfejsa i na njoj pronađe odevne predmete, odnosno *bounding box* objekte (*bounding box* objekat sadrži koordinate gornje leve tačke i donje desne tačke pravougaonika) u okviru kojih se nalazi odevni predmet koji pripada jednoj od devet razmatranih klasa.

Za realizaciju ovog modula korištena je *Faster R-CNN* konvoluciona neuronska mreža, sa arhitekturom *Faster RCNN Inception v2* [7]. Korišćena arhitektura je integrisana u *Tensorflow* biblioteku na detection model zoo repozitorijumu koji predstavlja kolekciju prethodno treniranih modela nad COCO skupom podataka. Ova arhitektura je odabrana na osnovu rada [8] koji se bavio sličnom temom detekcije odevnih predmeta. Ovaj rad je imao veoma uspešne rezultate i, po ugledu na njega, inicijalna arhitektura neuronske mreže koja je isprobana je *Faster RCNN Inception Resnet v2*.

Ova arhitektura je davala precizne rezultate na sistemu prezentovanom u ovom radu, međutim, zbog veoma male

brzine bila je skoro neupotrebljiva. Zbog toga, odlučeno je da se koristi arhitektura *Faster RCNN Inception v2* koja predstavlja model sa najboljim odnosom brzine i preciznosti.

Arhitekturu izabranog modela čine dva konvoluciona sloja, i dva sloja sažimanja koja primenjuju max pooling funkciju. Za normalizaciju se koristi ReLu funkcija [9]. Model je prethodno treniran na COCO skupu podataka koji sadrži opšte slike, i dotreniran je nad skupom podataka koji sadrži okvirno 400-500 slika odeće po klasi, što je ukupno oko 4000 slika (veći broj slika bi bio poželjan, ali je ovaj broj izabran kao kompromis vremena i kvaliteta treniranja – slike su ručno uređivane, i vremenski je zahtevno prolaziti kroz ceo *DeepFashion* [10] skup). Dotreniranje je vršeno u 18 epoha (veći broj epoha bi dao bolje rezultate, međutim, zbog vremenskog ograničenja treniranje je zaustavljeno, i pošto su rezultati bili zadovoljavajući, nije bilo neophodno nastaviti treniranje).

Prilikom dotreniranja ostavljeni su zamrznuti konvolucioni slojevi koji sadrže već podešene težine na osnovu slika iz COCO skupa, a odmrznuti su poslednji potpuno povezani slojevi namenjeni za klasifikaciju. Time je postignuto da konvoluciona mreža koja je već naučena da prepoznaje šablone sa slike, zna da klasifikuje sliku u devet klasa za potrebe ovog sistema.

Izlaz iz ovog modula su koordinate odevnog predmeta sa slike, odnosno, koordinate pravougaonika u kom se nalazi, kao i klasa kojoj predmet pripada.

2.2. Modul za preporučivanje

Ulaz u modul za preporučivanje predstavlja izlaz iz prethodnog modula, odnosno, deo slike koji predstavlja odevni predmet. U ovom delu sistema se bavimo tehnikama kompjuterske vizije, i oslanjamo se na *OpenCV* biblioteku.

Kako bismo izvukli relevantne parametre sa odevnog predmeta, moramo eliminisati pozadinu. Za to koristimo *Canny Edge detector* [11] i tražimo najveću konturu na slici. S obzirom na to da smo izvršili selekciju regiona od interesa u prethodnom koraku, pretpostavka je da je najveća kontura sam odevni predmet. Nakon što smo izolovali odevni predmet sa slike možemo da pristupimo procesu izdvajanja sledećih parametara:

- Prosečna vrednost *Red*, *Green* i *Blue* komponente izdvojenog regiona
- Prosečna vrednost *Hue*, *Saturation* i *Value* komponente izdvojenog regiona
- Koristeći lokalni binarni obrazac, dobijamo histogram koji predstavlja teksturu odevnog predmeta.

Do preporuke prvih n proizvoda dolazi se na jednostavan način. Vršiti se upit u bazu gde se zahtevaju proizvodi iz iste klase kao detektovani predmet, sortirani po sličnosti prosleđenih parametara (odnosno, gde je apsolutna vrednost razlike parametara najmanja). Mora se uzeti u obzir da nemaju svi parametri isti uticaj, pa tako *HSV* boja mnogo više utiče na računanje sličnosti nego *RGB*. To je regulisano tako što se apsolutna vrednost razlike *HSV* parametara množi određenom vrednošću kako bi imala veću težinu u procesu preporučivanja. Do konačne

vrednosti tog fakotra došlo se empirijski posmatranjem ponašanja sistema.

2.3. Grafički korisnički interfejs

Korisnički interfejs aplikacije je implementiran kao *web* aplikacija koja omogućava korisnicima sledeće osnovne funkcionalnosti:

- Pregled video zapisa, pretragu i navigaciju različitih video zapisa
- U bilo kom momentu trajanja video zapisa, pritiskom na dugme korisnici mogu da pretraže bazu proizvoda u potrazi za odevnim predmetom koji se u tom trenutku prikazuje na videu
- Pregled različitih proizvoda i ocenjivanje istih.

2.4. Skup podataka

Za obučavanje i testiranje modela konvolucione neuronske mreže čiji zadatak je detekcija i klasifikacija odevnih predmeta u devet kategorija, korišten je *DeepFashion dataset* [8] koji je sačinjen od preko 800.000 kvalitetnih slika (različitih dimenzija) na kojima su odevni predmeti dobro vidljivi jer se radi o slikama iz *online* prodavnica, gde su poze pogodne tako da odeća dolazi do izražaja.

Zbog hardverskih ograničenja, za potrebe ovog projekta nije korišten ceo skup podataka, već samo oko 4 000 slika, raspoređenih u devet klasa.

Neke slike sadrže opis teksture koji je omogućio evaluaciju podsistema za preporučivanje (npr., *Striped_Maxi_Dress*, *Floral_Side-Slit_Maxi_Dress*, *Leopard_Print_Maxi_Dress*, *Geo_Striped_Maxi_Dress*).

Problem koji postoji kod ovog skupa podataka jeste da ima previše klasa koje sadrže veoma slične podatke. Iz tog razloga je rađeno ručno prolaženje kroz klase skupa podataka i više klasa je spajano u jednu (na primer, postoji nekoliko klasa farmerki, ali su razlike neprimetne čak i čoveku kao posmatraču, tako da su ove klase spojene u jedinstvenu klasu - farmerke).

Spajanjem sličnih klasa i izbacivanjem klasa koje imaju previše šuma (postoje klase koje imaju toliko različitih slika da im čak ni čovek ne može naći sličnost, npr. klasa *Peplum_top*) došlo se do devet klasa razmatranih u ovom radu.

3. EVALUACIJA

Dva osnovna modula koja čine ovaj sistem (modul za detekciju i prepoznavanje i modul za preporučivanje proizvoda) su evaluirani kao potpuno nezavisne komponente nad različitim delovima *DeepFashion* skupa podataka.

Za evaluaciju modula za prepoznavanje i detekciju korišćene su slike različitih klasa, a za modul za preporučivanje korišćene su slike predmeta iste klase, ali različite teksture i boje.

3.1. Evaluacija modula za detekciju i prepoznavanje

Za potrebe evaluacije modela *Faster RCNN* konvolucione neuronske mreže, skup podataka je podeljen na trening skup (85%) i test skup (15%). Kao mera evaluacije korišćena je F-mera (*F1-score* metoda) i postignuti su zadovoljavajući rezultati, s obzirom na kvalitet skupa podataka i hardverska ograničenja.

Kategorija	F-mera
Majica	82%
Haljina	80%
Sako	89%
Farmerke	96%
Sportska majica	92%
Venčanica	86%
Jakna	92%
Suknja	70%
Šorc	87%

Tabela 3.1. Rezultati evaluacije modela za detekciju i prepoznavanje

Kada posmatramo tabelu 3.1, iako nema velikih odstupanja u zavisnosti od klase, možemo izdvojiti sledeća zapažanja:

- Najveća preciznost postignuta je kod farmerki - razlog za to je najverovatnije što su farmerke jedini odevni predmet koji predstavlja dugački donji deo odeće. Kada bismo uneli klasu pantalone pored farmerki, verovatno bismo značajno narušili preciznost.
- Najlošije rangirana klasa je suknja sa 70% tačnosti. Posmatranjem rezultata testa, zaključeno je da se suknja veoma često pogrešno klasifikuje kao haljina ili venčanica.
- Primećeno je često pogrešno prepoznavanje šorca na određenim slikama gde se ne vidi ceo donji deo tela.

3.2. Evaluacija modula za preporučivanje

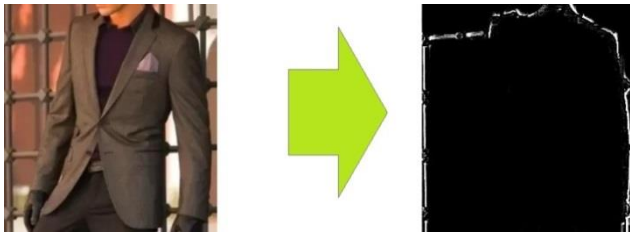
Celokupan modul za preporučivanje koji bi uključivao ocene pojedinačnih korisnika, te time pružio mogućnost razvoja personalizovanog sistema za preporuku za sada nije moguće precizno evaluirati. Međutim, evaluiran je deo sistema koji je najkompleksniji u ovom modulu, a to je *LBP* operator zadužen za ekstrakciju teksture sa slike. Ostala obeležja koja označavaju boju nisu uzeta u obzir zato što ne postoje odgovarajuće anotacije u skupu podataka koje bi omogućile evaluaciju preporuke koja uzima u obzir i za obeležja.

Kako bi se uspešno izvršila evaluacija *LBP* operatora, problem preporučivanja je predstavljen kao problem klasifikacije. *DeepFashion dataset* sadrži attribute koji opisuju teksturu, tako da je to omogućilo evaluaciju *LBP* operatora.

U bazi proizvoda su sačuvana četiri tipa haljina različitih tekstura: leopard tekstura, tačkasta tekstura, tekstura na horizontalne linije i tekstura na vertikalne linije. Testiranje je vršeno tako što se za prosledeni tip teksture od sistema za preporuku očekuje da preporučuje haljinu sa istim dezenom. Ovo je moguće izvršiti i na ostalim klasama, ali je odlučeno da se koristi klasa haljine jer ima najviše različitih tipova tekstura. U procesu evaluacije ovog podsistema uspešnom preporukom se smatra slučaj kada istu oznaku teksture ima slika na ulazu kao i slika u bazi proizvoda koja je preporučena kao najpogodnija. Na tačnost evaluacije ovog podsistema ne utiče prethodni korak, odnosno klasifikacija. Svi proizvodi koji ulaze u evaluaciju priparadju istoj klasi. Kao mera evaluacije korišćena je F-mera i postignut je rezultat od 76%.

Analizom grešaka modela utvrđeno je da je u ovom najslabija karika pronalaženje najveće konture. Problem se najčešće javlja kada je iza detektovanog objekta previše šarena pozadina, te se deo pozadine uzme kao površina za ekstrakciju atributa boje i teksture (slika 3.1).

Može se desiti da je kontrast između odevnog predmeta i pozadine previše mali tako da se ivice neuspešno detektuju i u tom slučaju se uzima ceo *bounding box* kao površina od interesa što vodi ka nepreciznom računanju prosečne boje i histograma teksture.



Slika 3.1. Pogrješno detektovana najveća kontura na slici

4. ZAKLJUČAK

U ovom radu je predstavljen sistem za prepoznavanje i preporučivanje odevnih predmeta sa video zapisa. Motivacija za takav sistem potekla je iz sfere *online* prodaje i marketinga. Osnovna ideja je bila da se napravi sistem koji pomaže potencijalnim kupcima odeće preko interneta da pronadu kombinaciju koju vide na filmu, a samim tim i prodavcima da lakše prodaju svoje proizvode. Sistem je sačinjen iz tri modula: modul za detekciju i prepoznavanje (ulaz u modul predstavlja slika dobijena sa korisničkog interfejsa, a izlaz je *bounding box* koji govori o tome gde se na slici nalazi detektovani proizvod), modul za preporučivanje proizvoda (ulaz predstavlja *bounding box* iz prethodnog modula, a izlaz najslabijih n proizvoda), i grafički korisnički interfejs kao odvojena aplikacija.

Prvi modul se oslanja na *Faster RCNN* konvolucionu neuronsku mrežu kreiranu sa *Faster RCNN inception v2* arhitekturom. Za evaluaciju prvog modula korištena je *F-mera* koja je za ovaj modul iznosila 86% tačnosti. Za obučavanje i testiranje veštačke neuronske mreže korišten je *DeepFashion* skup podataka. Osnovni problem korišćenog skupa podataka je bio preveliki broj kategorija, tako da je skup morao biti ručno prerađen za potrebe ovog sistema.

Drugi modul se oslanja na različite metode kompjuterske vizije (*Canny detektor ivica*, *LBP* operator, itd.) i zasniva se na ekstrakciji boje i teksture. Ovaj podsistem nije pogodan za dubinsku evaluaciju usled nedostataka odgovarajućih anotacija skupa podataka, ali je svakako evaluiran *LBP* operator koristeći *F-meru* i postignuta *F-mera* iznosi 76%.

Prilikom evaluacije, uočeni su i neki nedostaci sistema: Zbog hardverskih ograničenja nije iskorišten ceo obučavajući skup, i rađen je mali broj epoha pri obučavanju. Duže obučavanje bi verovatno unapredilo performanse sistema.

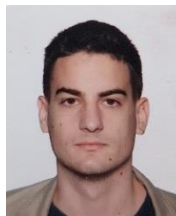
Drugi problem sistema je što vrši detekciju samo jednog proizvoda sa video zapisa i ukoliko bi više odevnih predmeta bilo prikazano u isto vreme, sistem bi sam odabrao predmet koji će se preporučiti, odnosno, korisnik ne može da utiče na to. *R-CNN* mreža svakako pronalazi više predmeta, međutim samo jedan se uzima u obzir (onaj koji je prepoznat sa najvećom sigurnošću) kako bi se u aplikaciji prikazali različiti proizvodi iste vrste i korisnik imao mogućnost upoređivanja.

Problem je neodstatak opcije da korisnik selektuje konkretan odevni predmet.

5. LITERATURA

- [1] Dr.R.Shanthi I Dr. Desti Kannaiiah, Vol.13, 2015, Consumers' Perception on Online Shopping, Assistant Professor, Department of Commerce, University of Madras, Chennai -600005, Tamil Nadu,India
- [2] Python.org. (2019). Welcome to Python.org. Available at: <https://www.python.org/> [15 Oct. 2019].
- [3] TensorFlow. (2019). TensorFlow. Available at: <https://www.tensorflow.org/> [15 Oct. 2019].
- [4] Emgu.com. (2019). Emgu CV: OpenCV in .NET (C#, VB, C++ and more). Available at: http://www.emgu.com/wiki/index.php/Main_Page [15 Oct. 2019].
- [5] Opencv.org. (2019). OpenCV. Available at: <https://opencv.org/> [15 Oct. 2019].
- [6] js.foundation, J. (2019). jQuery. Jquery.com. Available at: <https://jquery.com/> [15 Oct. 2019].
- [7] Tensorflow detection model zoo https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md
- [8] Clothing Detection for Fashion Recommendation <https://blog.valohai.com/clothes-detection-for-fashion-recommendation>
- [9] Yun Ren,Changren and Shunping Xiao, Object Detection Based on Fast/Faster RCNN Employing Fully, Convolutional Architectures, Hindawi Mathematical Problems in Engineering Volume 2018
- [10] Large-scale Fashion (DeepFashion) Database, Multimedia Laboratory, The Chinese University of Hong Kong <http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html>
- [11] Bill Green, Canny Edge Detection Tutorial (2002), Drexel Autonomous Systems Lab., 2002. <http://masters.donntu.org/2010/fknt/chudovskaja/library/article5.htm>

Kratka biografija:



Nebojša Basarić rođen je 18.05.1994. godine u Kninu, Republici Hrvatskoj. 2009. godine upisuje Srednju tehničku školu u Somboru, smer elektrotehničar računara. Srednju školu završava 2013. godine kada i upisuje osnovne studije na Fakultetu tehničkih nauka u Novom Sadu, smer računarstvo i automatika. Osnovne studije završava u roku 2017. godine i upisuje master studije na istom fakultetu. Položio je sve ispite predviđene planom i programom. kontakt: nebojsabasarić94@gmail.com