Univerzitet u Nišu

Elektronski fakultet

Predmet: Prikupljanje i predobrada podataka za mašinsko učenje

Augmentacija podataka kod slika

Seminarski rad

Smer: Veštačka inteligencija i mašinsko učenje

Profesor:

Doc. dr Aleksandar Stanimirović

Niš, april 2024. godine

Student:

Anđelija Mladenović, br. ind. 1625

Sadržaj

[1. Uvod 3](#_Toc164027365)

[2. Motivacija za korišćenje augmentacije podataka kod slika 4](#_Toc164027366)

[3. Tradicionalne metode augmentacije podataka kod slika i njihova podela 5](#_Toc164027367)

[4. Pregled tradicionalnih metoda augmentacije podataka kod slika 6](#_Toc164027368)

[4.1 Geometrijske transformacije 6](#_Toc164027369)

[4.1.1 Rotacija 6](#_Toc164027370)

[4.1.2 Izvrtanje 7](#_Toc164027371)

[4.1.3 Isecanje i zumiranje 8](#_Toc164027372)

[4.1.4 Pomeranje 8](#_Toc164027373)

[Literatura 9](#_Toc164027374)

# Uvod

Modeli dubokog učenja stekli su tokom godina veliku popularnost i uznapredovali za primenu za rešavanje zadataka vezanih za računarski vid, kao što su klasifikacija slika, detekcija objekata, segmentacija objekata sa slike itd. Ovo je podstaknuto pojavom različitih novih arhitektura neuronskih mreža, dostupnošću računarskih resursa za njihovo pokretanje i dostupnošću sve veće količine podataka [1].

Ipak, algoritmi korišćeni kod dubokog učenja zahtevaju ogromne količine podataka, i često imaju problema sa *overfitting*-om – model radi jako dobro na trening podacima, ali i izuzetno loše na test, odnosno „neviđenim“ podacima. Ovaj problem se obično javlja u slučajevima kada nam nije dostupna dovoljno velika količina podataka za treniranje modela – ovo se može dogoditi zbog zaštite privatnosti ljudi koji se na slikama nalaze (čest problem sa medicinskim slikama), ili u slučajevima kada je labeliranje slika jako zahtevno i skupo [1].

Jedno od potencijalnih rešenja ovog problema predstavlja augmentacija podataka. Augmentacija podataka obuhvata širok spektar tehnika koje se primenjuju na same podatke, a ne na model sa kojim radimo. Ove tehnike mogu poboljšati performanse modela tako što veštački proizvode različite i raznolike uzorke na kojima će model moći da uči. Kada je skup podataka sa kojim radimo dovoljno kvalitetan i poseduje dovoljno uzoraka, performanse modela koji se primenjuje za neki zadatak su znatno bolje. S toga, podaci s kojima radimo moraju ispuniti dva uslova – raznolikost i brojnost, a i jedno i drugo može se postići augmentacijom podataka [2].

Iako postoje razne podele metoda za augmentaciju podataka kod slika, najčešća je ona koja ih deli na tradicionalne metode i metode zasnovane na dubokom učenju. Tradicionalne metode odnose se na one koje izvršavaju neku vrstu transformacije na slici, bile to geometrijske, ili fotometrijske transformacije, dodavanje šuma, primena filtera i slično. Sa druge strane, medote zasnovane na dubokom učenju najčešće koriste *GAN-*ove (*Generative Adversarial Network*) da na osnovu trening slika generišu potpuno nove, sintetičke slike koje će se koristiti za treniranje našeg modela. U ovom radu, fokus će biti na tradicionalnim metodama, i one će biti obrađene u nastavku [2].

# Motivacija za korišćenje augmentacije podataka kod slika

Modeli dubokog učenja uspešno su primenjeni za rešavanje zadataka iz oblasti računarskog vida kao što su klasifikacija slika i segmentacija slika, detekcija objekata i slično, a sve to zahvaljujući razvoju konvolucionih neuronskih mreža (engl. *Convolutional Neural Network* - CNN). Ove neuronske mreže koriste parametrizovane, retko povezane kernele koji očuvavaju prostorne karakteristike slike. Konvolucioni slojevi sekvencijalno *downsample*-uju prostornu rezoluciju slike dok u isto vreme povećavaju dubinu mapa karakteristika (engl. *feature map*). Niz konvolucionih transformacija može napraviti reprezentaciju slike koja je znatno manjih dimenzija od originala, ali koja predstavlja znatno korisniji oblik slike [3].

Povećanje sposobnosti generalizacije ovih modela pokazalo se kao značajan problem. Generalizacija predstavlja sposobnost modela da radi podjednako dobro na već viđenim primerima (trening podaci) i na podacima sa kojima se nije susreo u prošlosti (test podaci). Ukoliko model ne generalizuje dobro, odnosno radi lošije na test podacima, to obično znači da je došlo do *overfitting*-a – model je naučio da radi sa trening slikama, ali nije imao dovoljno podataka da podesi svoje parametre tako da može da radi i sa drugim podacima [3].

Razvijen je veliki broj metoda za suočavanje s ovim problemom, a jedna od njih je i augmentacija podataka. Augmentacija podataka predstavlja veoma moćan metod za poboljšanje performansi modela. Njenom primenom proširujemo trening skup podataka modifikacijom već postojećih slika, čime dobijamo opširniji i temeljniji skup mogućih podataka, čime se smanjuje moguća razlika između trening podataka i podataka na kojima će se model primenjivati u budućnosti [3].

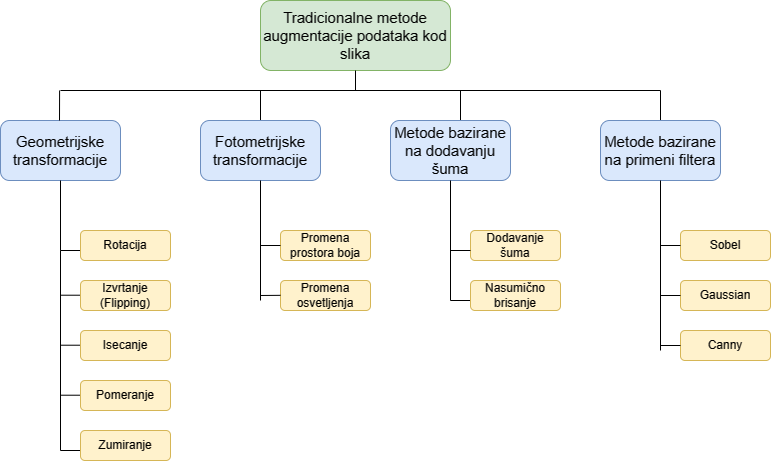
Za razliku od većine drugih tehnika, augmentacija podataka prisupa problemu *overfitting*-a od njegovog korena, odnosno od samog trening skupa podataka. Ovo se radi pod pretpostavkom da se dodatne informacije mogu dobiti iz originalnog skupa podataka primenom različitih augmentacija. Ove augmentacije veštački povećavaju veličinu trening skupa podataka. Augmentacije mogu predstavljati različite stvari – geometrijske transformacije slike, njihovo zamućenje, promenu njihove reprezentacije, odnosno prostora boja, zatim brisanje određenih delova slike, i druge [3].

Činjenica da augmentacija podataka povećava broj trening primera, jeste najočiglednija, ali ne i jedina prednost ove metode. Cilj pri pravljenju modela za rešavanje zadataka iz oblasti računarskog vida jeste da ti modeli budu što bliži čoveku. Transformacije koje se najčešće primenjuju prilikom augmentacije su percepciono moguće – ljudsko oko radi gotovo podjednako dobro bez obzira na osvetljenje, vid može da se zamuti, odnosno – augmentacijom podataka zapravo simuliramo realne uslove pod kojima ljudsko oko može da radi. Augmentacijom podataka sposobni smo da generišemo veliki broj novih primera kojima ćemo „predviđati“ potencijalne realne primere na kojima će model biti korišćen [4].

# Tradicionalne metode augmentacije podataka kod slika i njihova podela

Kao što je ranije pomenuto, metode augmentacije podataka mogu se podeliti na tradicionalne metode, koje se baziraju na modifikaciji slika iz trening seta, i na metode zasnovane na dubokom učenju, koje koriste određene modele dubokog učenja da na osnovu trening primera generišu potpuno nove primere [2]. Fokus u ovom radu biće stavljen na tradicionalne metode augmentacije podataka.

Tradicionalne metode mogu se podeliti u četiri grupe: geometrijske transfomacije, fotometrijske transformacije, metode bazirane na dodavanju šuma i metode bazirane na primeni nekog filtera [2]. Taksonomija tradicionalnih metoda augmentacije podataka na slikama data je na slici 3.1. Važno je napomenuti da ovo nije jedina podela metoda augmentacije podataka kod slika, ali jeste ona koju ćemo usvojiti u ovom radu. Svaka od metoda biće obrađena pojedinačno u nastavku rada.



Slika 3.1 Taksonomija tradicionalnih metoda augmentacije podataka [2]

# Pregled tradicionalnih metoda augmentacije podataka kod slika

U nastavku će biti dat pregled tradicionalnih metoda za augmentaciju podataka kod slika.

## Geometrijske transformacije

Osnovne geometrijske transformacije, kao što su izvrtanje (engl. *flipping*), isecanje ili rotacija su i dalje popularni metodi za augmentaciju podataka. Ove tehnike generalno povećavaju količinu dostupnih podataka za treniranje, i doprinose diverzitetu podataka. Takođe, njihova primena je jako jednostavna [2].

Ovaj odeljak baviće se različitim geometrijskim transformacijama slike koje se mogu iskoristiti za augmentaciju podataka. Takođe, biće razmotrena i „bezbednost“ primene ovih transformacija. Bezbednost augmentacije podataka odnosi se na verovatnoću očuvanja labele slike nakon transformacije. Na primer, rotacija i izvrtanje su generalno bezbedne kada je u pitanju clasifikacija slika na slike sa psima i slike sa mačkama, ali neće biti bezbedne ukoliko je neohodno da razvrstamo slike koje sadrže broj 9 ili broj 6. Transformacija koja ne očuvava labele u potpunosti može potencijalno pojačati sposobnost modela da da odgovor čak i onda kada nije potpuno siguran u svoju predikciju. Međutim, da bi se ovo postiglo neophodne su dodatne modifikacije modela, što može biti skup proces. Zbog toga, važno je razmotriti „bezbednost“ augmentacije podataka. Ona je donekle zavisna od domena primene, zbog čega je teško razviti neka generalna pravila o augmentaciji podataka. Ne postoji ni jedna metoda obrade slike koja nikada ne može rezultovati izmenom labele. Sve ovo neophodno je razmotriti pre primene metoda augmentacije navedenih u nastavku [3].

### Rotacija

Rotacija je jednostavna geometrijska transformacija korišćena za augmentaciju podataka. Njenom primenom, postojeće slike se rotiraju za određeni ugao, i zatim se novodobijene slike koriste kao trening primeri [2].

Rotacija se može vršiti u levo ili u desno za ugao od 1° do 359°. Bezbednost rotacije je određena vrednošću ugla rotacije. Blaga rotacija za ugao između 1° i 20° (odnosno -1 do -20°) može biti korisna za treniranje modela koji prepoznaje cifre, ali što se ugao više povećava, sve su veće šanse da labela podataka neće biti očuvana nakon transformacije [3]. Primer slike na kojoj je primenjena rotacija može se videti na slici 4.1.

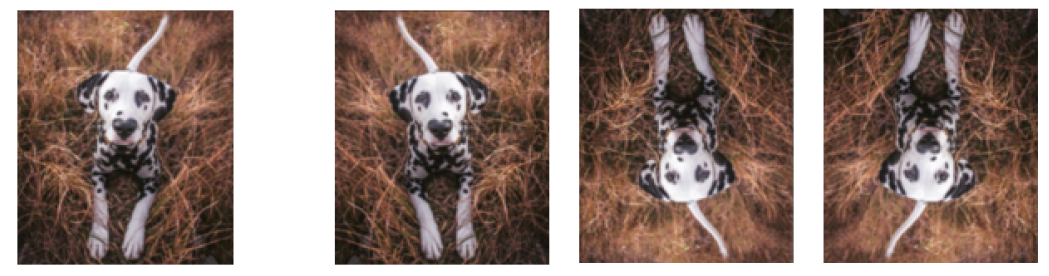
1. b)

Slika 4.1 a) Slika pre rotacije; b) Slika nakon rotacije [2]

### Izvrtanje

Termin izvrtanje (engl. flipping) odnosi se na proces izvrtanja slike horizontalno, vertikalno ili u oba pravca, kao što je prikazano na slici 4.2 [2]. U praksi, horizontalno izvrtanje je mnogo češće korišćeno od vertikalnog, zbog toga što je obično realističnija[2, 3]. Ova augmentacija jedna je od najlakših za implementaciju, i pokazala se kao izuzetno korisna na brojnim skupovima podataka [3].

Izvrtanje je jedna od najintuitivnijih metoda za povećanje diverziteta i broja podataka. Ipak, nije pogodna za primenu na podacima koji imaju jedinstvena svojstva. Na primer, izvrtanje nije pogodno onda kada su objekti koje segmentiramo asimetrični ili su osetljivi na smer, kao što su to slova ili brojevi. U ovakvim slučajevima, primena izvrtanja rezultovaće promenom labele podataka, što će prouzrokovati brojne probleme [2].



d)

c)

b)

a)

Slika 4.2 a) Originalna slika; b) Horizontalno izvrtanje; c) Vertikalno izvrtanje; d) Oba izvrtanja [2]

### Isecanje i zumiranje

Ove dve metode najčešće idu ruku pod ruku. Isecacnje je često korišćena metoda augmentacije koja može biti primenjena na nasumični deo slike ili na sam centar slike. Ova metoda podrazumeva isecanje dela slike i zatim njegovo reskaliranje (zumiranje) na način da veličina novodobijene slike odgovara veličini inicijalne slike [1, 2]. Zbog toga što trening skup podataka može imati slike različite veličine, isecanje se često koristi za preprocesiranje podataka kako bi se sve slike dovele na iste dimenzije [2].

Bitno je pomenuti da isecanje može proizvesti primere sa pogrešnom labelom. Na primer, slika koja sadrži više od jednog objekta, koja je klasifikovana na osnovu objekta koji je na slici dominantan po veličini, može biti pogrešno labelirana ukoliko se taj objekat većinski iseče sa slike. U tom slučaju, moguće je iseći deo slike koji sadrži više detalja tih pratećih objekata, umesto onog koji sadrži dominantni objekat, kao što je prikazano na slici 4.3. Ova slika inicijalno je bila označena labelom „Pas“, ali se nakon isecanja pas uopšte ne nalazi na slici, što znači da je labela postala netačna. Zbog toga, neophodno je obratiti pažnju na primenu ove metode.

b)

a)

Slika 4.3 a) Slika pre isecanja, označena labelom „Pas“; b) Slika nakon isecanja i reskaliranja (zumiranja), i dalje, ali sada netačno, označena labelom „Pas“ [2]

### Pomeranje

Pomeranje, odnosno translacija, je metoda augmentacija podataka na slici koja podrazumeva pomeranje slike gore, dole, levo ili desno kako bi se dobila raznolika reprezentacija podataka [1]. Ovakvo pomeranje može biti značajno za sprečavanje tendencije modela (engl. *bias*) da povezuje objekte sa njihovom prostornom lokacijom. Na primer, ukoliko su sve slike u skupu podataka centrirane, što je često kod skupova podataka vezanih za prepoznavanje lica, to znači da će i sve test slike morati da budu centrirane da bi model radio kako treba, što gotovo nikad nije slučaj. Zbog toga bi pomeranje bila pogodna transformacija za augmentaciju ovakvog skupa podataka [3].

Nakon pomeranja, delovi slike koji ostanu „prazni“, odnosno iz kojih je originalna slika izmeštena, mogu se dopuniti nekom konstantnom vrednošću kao što su 0 ili 255, ili se pak mogu popuniti nekim Gausovim šumom. Ova dopuna služi da sačuva prostorne dimenzije slike i nakon augmentacije [3].

Magnituda pomeranja mora biti pažljivo izabrana u zavisnosti od slučaja primene. Veliko pomeranje slike može rezultovati značajnom promenom izgleda slike. Na primer, pomeranje broja 8 u levo za polovinu širine slike može rezultovati slikom koja je približnija broju 3 nego originalnoj labeli. Zbog toga, jako je bitno razmotriti prirodu skupa podataka sa kojima radimo pri određivanju magnitude traslacije kako bi se osigurala njena efikasnost [1].

Primer pomeranja dat je na slici 4.4.

Slika 4.4 a) Originalna slika; b) Slika nakon pomeranja na dole i u desno [1]

## Fotometrijske transformacije

# Literatura

[1] T. Kumar, A. Mileo, R. Brennan, M. Bendechache, „Image Data Augmentation Approaches: A Comprehensive Survey and Future directions“, *arXiv preprint*, 2023.

[2] K. Alomar, H. I. Aysel, X. Cai, „Data Augmentation in Classification and Segmentation: A Survey and New Strategies“, *Journal of Imaging*, vol. 9, 2023.

[3] C. Shorten, T. M. Khoshgoftaar, „A survey on Image Data Augmentation for Deep Learning“ *Journal of Big Data*, vol. 6:60, 2019.

[4] A. Hernández-García, „Data augmentation and image understanding“, PhD thesis, Institute of Cognitive Science, University of Osnabrück, 2020.