**Elektronski Fakultet**

**Univerzitet u Nišu**

****

**Augmentacija slika, osnovne tehnike i tehnike dubokog ucenja**

**- Seminarski rad -**

**Predmet:**

**Prikupljanje i predobrana podataka**

**Student: Mentor:**

**Dimitrije Petrović (1682) Prof. dr Aleksandar Stanimirović**

**Niš, decembar 2024. Godine**

**Sadržaj**

[**1.** **Uvod** 3](#_Toc184589987)

[**2.** **Geometrijske transformacije** 4](#_Toc184589988)

[**Translacija** 4](#_Toc184589989)

[**Rotacija** 5](#_Toc184589990)

[**Okretanje** 5](#_Toc184589991)

[**Isecanje/Kropovanje** 5](#_Toc184589992)

[**Padding** 5](#_Toc184589993)

[**Skaliranje** 5](#_Toc184589994)

[**Ubacivanje šuma** 5](#_Toc184589995)

[**3.** **Transformacije boja !!!REFERENCA** 6](#_Toc184589996)

[**Osvetljenost** 6](#_Toc184589997)

[**Saturacija/Zasićenost** 7](#_Toc184589998)

[**Kontrast** 7](#_Toc184589999)

[**4.** **Tehnike brisanja i maskiranja** 7](#_Toc184590000)

[**Cutout** 7](#_Toc184590001)

[**Nasumično brisanje (Random Erasing)** 8](#_Toc184590002)

[**Hide-and-Seek** 9](#_Toc184590003)

[**5.** **Tehnike dubokog učenja** 10](#_Toc184590004)

[**5.1 Autoenkoderi** 10](#_Toc184590005)

[**5.2 Varijacioni autoenkoderi** 11](#_Toc184590006)

[**5.2 Generativne protivničke mreže (GAN – Generative Adverserial Networks)** 12](#_Toc184590007)

[**6.** **Zaključak** 12](#_Toc184590008)

[**7.** **Literatura** 12](#_Toc184590009)

# **Uvod**

U problemima računarskog vida, tehnike augmentacije se konstantno koriste zbog toga što omogućavaju povećanje samih ulaznih podataka, bolju generalizaciju, i omogućavaju mreži da se ne fokusira na najupečatljivije delove slike, već da bolje razume sam objekat koji se na njoj nalazi.

Pod ovim tehnikama se najčešće misli na

* Geometrijske tehnike – menjaju prostorne dimenzije slike, skoro uvek primenljive
* Fotometrijske tehnike – modifikacije boje, osvetljenja

Pri augmentaciji se takođe koriste i neke arhitekture čiji je cilj generisanje novih podataka sličnim ulaznim.

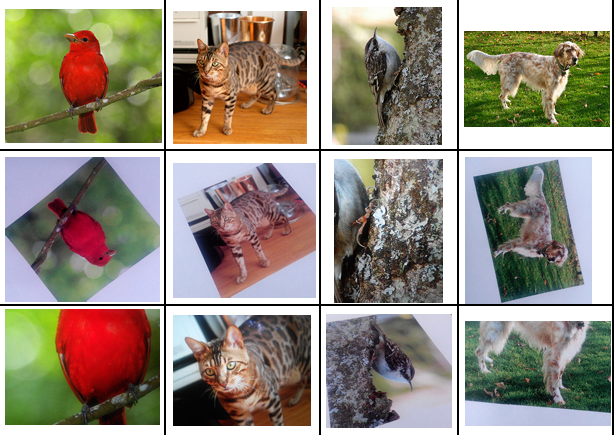
Cilj ovog rada je upoznavanje sa osnovnim tehnikama, kao i sa tehnikama brisanja i arhitekturama koje takođe mogu da se iskoriste za augmentaciju slika.

# **Geometrijske transformacije**

Geometrijske transformacije menjaju prostorne dimenzije slike. Ove transformacije se veoma lako implementiraju i mogu da se kombinuju.

Pri upotrebi ovih transformacija moramo imati na umu da mogu negativno da utiču na samu sliku – ako radimo klasifikaciju cifara moramo paziti na rotaciju zbog sličnosti cifra 6 i 9. Takođe, ako rotiramo sliku neki delovi mogu da izađu van same dimenzije slike. Zbog ovih problema, koji ponekad traže od nas i posmatranje samog rezultata nakon transformacije, njihova primena ponekad može biti veoma ograničena.

Ove transformacije uključuju translaciju, rotaciju, skaliranje, okretanje, ubacivanje padding-a, dodavanje šuma...



*Slika 1. geometrijske transformacije* [1]

## **Translacija**

Translacija podrazumeva pomeranje same slike horizontalno ili vertikalno (ili po obe ose). Nakon ove transformacije ako želimo da zadržimo prostorne dimenzije možemo da ubacimo piksele (padding).

## **Rotacija**

Ovde okrećemo sliku u smeru kazaljke na satu (ili obrnuto) za 1 do 359 stepeni (rotiranjem za 0 ili 360 stepeni krajnji rezultat je slika identična početnoj pre rotacije). Ovde moramo paziti da može doći do praznina u slici, kao i da je moguće u pojedinim slučajevima promeniti kategoriju same slike (npr. brojevi 6 i 9, kao i 8 i beskonačnost).

## **Okretanje**

Okretanje se vrši po horizontalnoj ili vertikalnoj osi (ili po obe), u praksi je mnogo češće okretanje po vertikalnoj osi. Može se uočiti povezanost ove transformacije i rotacije, tj. ovo predstavlja poseban slučaj rotacije, za konkretan broj stepeni.

## **Isecanje/Kropovanje**

Ovde možemo (nasumično) iseći region slike i na taj način smanjujemo prostorne dimenzije same slike, zbog čega ovu transformaciju ponekad prati i skaliranje (resize). Isecanje se najčešće radi u odnosu na sam centar slike, ali je takođe od značaja isecanje u odnosu na neku drugu koordinatu.

## **Padding**

Ova transformacija se najčešće primenjuje zajedno sa drugim transformacijama, tačnije nakon transformacija ako se promene njene prostorne dimenzije. Tada se ubacuju pikseli na ivicama kako bi se povećala dimenzije slike. Najčešće se dodaju “ivice” crne ili bele boje, ali takođe možemo nasumično izabrati boju ili uzimajući u obzir normalnu raspodelu.

## **Skaliranje**

Ovde se najčešće misli na zumiranje slike, koje može i da se posmatra kao poseban primer isecanja gde odnos stranica ostaje isti. Ove transformacije takođe menjaju dimenziju slike.

## 

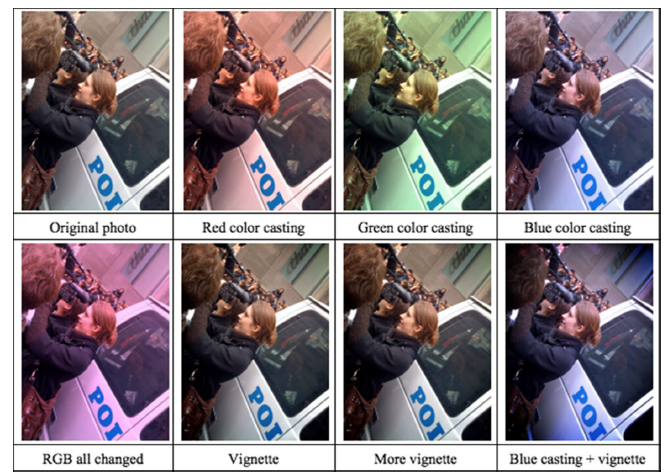
## **Ubacivanje šuma**

Ovde ubacujemo piksele nasumičnih vrednosti koje prate normalnu raspodelu. Takođe ovi pikseli mogu biti i potpuno crni ili beli, tj. sivi. Ova transformacija utiče na sam kvalitet slike, slika izgleda “zrnasto” i gubi jasnoću.

# **Transformacije boja**

Ove transformacije, takođe poznate pod nazivom fotometrijske transformacije, se koriste za modifikovanje samih boja slike, kao i osvetljenja. Za razliku od geometrijskih transformacija, ove transformacije ne utiču na prostorne dimenzije slike. Glavna namena ovih transformacija je promena osvetljenosti, kontrasta, izoštravanje i zamućivanje...

Kao i kod geometrijskih transformacija moramo biti pažljivi kod ovih transformacija, posebno ako su poje od velikog značaja za naš problem ili ako promenom boja moće doći i do promene same klase/labele slike. Takođe, slike su najčešće u RGB formatu, dok je ponekad zgodnije koristiti HSV format.



*Slika 2. fotometrijske transformacije*[2]

## **Osvetljenost**

Primenom ove transformacije slika može postati svetlija/tamnija. Ova transformacija može biti veoma korisna ako želimo da simuliramo različite periode dana u slici, ali takođe moramo paziti jer velikim izmenama osvetljenosti može samom modelu biti teško da prepozna šta se nalazi na slici.

## **Saturacija/Zasićenost**

Ovde povećavamo ili smanjujemo sam intenzitet boja, pri čemu možemo doći i do crno-bele slike nakon transformacije. Ova transformacija je poželjna ako same informacije o boji nisu bitne za naš problem, ako jesu, primenom ove transformacije možemo očito pokvariti naše podatke za treniranje i sam model.

## **Kontrast**

Ovom transformacijom menjamo osvetljenost različitih delova slike, tj. slika sa velikim kontrastom ima veoma značajnu razliku u osvetljenosti svojih najsvetlijih i najtamnijih delova.

# **Tehnike brisanja i maskiranja**

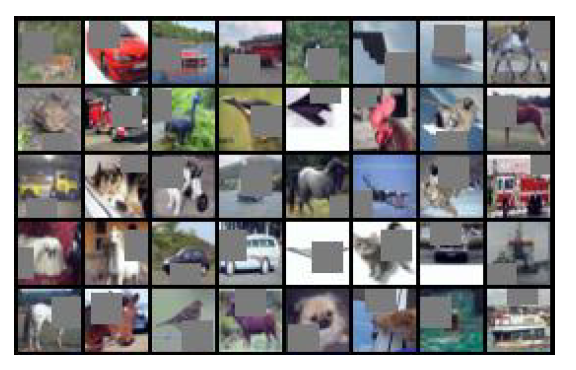
Ove tehnike pomažu da model nauči još neke dodatne osobine, pomažu mu da ne generalizuje na osnovu malog broja naučenih osobina. Kako bi se ovo postiglo, možemo obrisati neke regione sa slike ili ih maskirati. Ovde se ubrajaju tehnika nasumičnog brisanje, cutout i hide-and-seek.

## **Cutout**

Ova tehnika je dosta slična dropout metodi. Dropout funkcionise tako što se neki neuroni u mreži nasumično isključuju pri obučavanju, ovo se ponavlja za svaki prolaz kroz mrežu. Na ovaj način se mreža uči da bolje generalizuje i da uzima veći broj karakteristika u obzir, ovo je veoma približno obučavanju ansamblu mreža.

Cutout funkcioniše veoma slično, ali se primenjuje na same slike da bi se postigao ovaj efekat, tj – primenjuje se samo na prvom sloju mreže I pikseli koji se biraju moraju biti jedan pored drugog. Ova tehnika ima sličan uticaj kao i tehnike regularizacije. Cutout podrazumeva generisanje nasumičnih regiona slike (koji takođe mogu biti nasumične veličine) koji se iseču tako što se ceo region oboji jednom bojom. Na ovaj način mreža može da sazna šta treba da se nalazi na tim regionima samo na osnovu konteksta slike.

Sama tehnika može da se kombinuje sa ostalim tehnikama regularizacije, kao i da se upotrebom paralelizacije primeni pri učitavanju podataka.

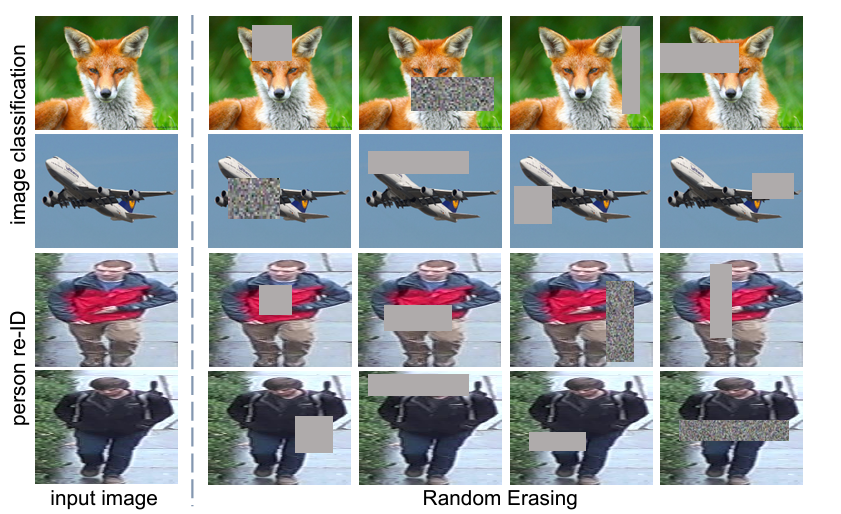


*Slika 3. Cutout* [3]

## **Nasumično brisanje (Random Erasing)**

Ova tehnika je veoma slična cutout tehnici, ali takođe unosi i neke novine. Možemo podešavati veličine regiona, takođe možemo da imamo veći broj regiona na jednoj slici. Sami regioni mogu da se pored potpunog popunjavanja jednom bojom popune i nasumičnim šumom.

Ova tehnika, kao i ostale njoj slične, se trude da upotrebom okluzije povećaju prediktabilnu moć neuronskih mreža. Pod okluzijom se smatra na postojanje drugih objekta ispred našeg objekta od interesa.

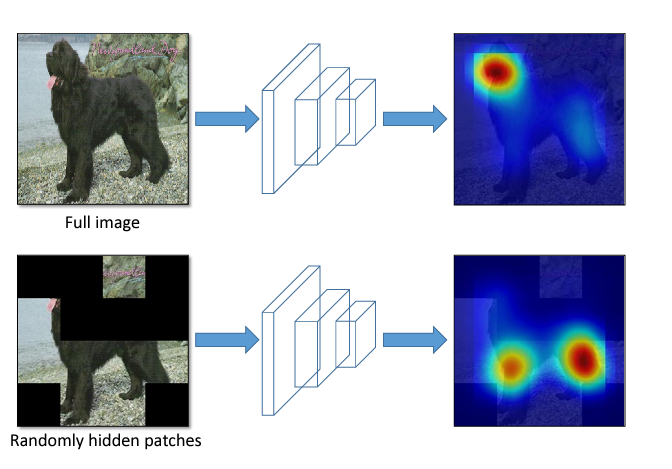


*Slika 4. Random Erasing*[4]

## **Hide-and-Seek**

Ova tehnika je veoma slična prethodnoj. Razlika je u tome što se ovde slika deli na mrežu (grid), pa se zatim nasumično biraju ćelije koje će biti popunjene crnom ili belom bojom. Ovaj proces se ponavlja pri svakoj iteraciji. Na ovaj način se bolje simulira scenario iz stvarnog sveta kada imamo više objekta koji se nalaze ispred glavnog predmeta naše slike. Primenom ove tehnike možemo promeniti fokus same mreže. Pri testiranju se koriste početne slike.

U odnosu na random erasing, ova tehnika može da izabere veći broj ćelija za maskiranje (ako imamo red mreže sa tri ćelije, moguće je da prva bude sakrivena, druga otkrivena/nemodifikovana, i treća sakrivena), pa zbog toga može da poboljša rezultate više od nasumičnog brisanja.



*Slika 5. Hide-and-Seek* [5]

# **Tehnike dubokog učenja**

Glavni cilj ovih tehnika je se zasniva na ideji da se augmentacija podataka zasniva na generisanju novih primera podataka, ove tehnike na osnovu postojećih, prave nove. Takođe, pošto su neuronske mreže dobre u predstavljanju ulaza sa velikim brojem dimenzija pomoću reprezentacija manjih dimenzija (feature mape), ovde se uključuju i tehnike koje modifikuju te reprezentacije.

## **5.1 Autoenkoderi**

Ova arhitektura se sastoji od dve mreže – enkodera i dekodera.

Enkoder služi za mapiranje ulaza u feature. Ovi feature-i treba da budu najkarakterisnije osobine za ulazne podatke, tj. da je moguće da se na osnovu njih rekonstruišu ulazni podaci iste dimenzionalnosti. Ovde takođe imamo i redukciju dimenzionalnosti, zato što su dimenzije izlaza enkodera manje od dimenzija ulaza

Dekoder za ulaz dobija izlaz enkodera, i na osnovu tih feature-a pokušava rekonstruiše podatke.

Sami ulazni podaci ne sadrže labele, što ovo čini nenadgledanim učenjem. Nakon treniranja možemo odbaciti dekoder i koristiti enkoder za nadgledano učenje. Takođe, nad samim izlazom enkodera, tj. vektorskom reprezentacijom podataka možemo dodati šum, vršiti interpolaciju i druge vektorske manipulacije.

## **5.2 Varijacioni autoenkoderi**

Kod varijacionih autoenkodera, za izlaz dobijamo srednju vrednost latentne distribucije i njenu standardnu devijaciju. Latentni proctor predstavlja prostor niske dimenzije u koji su mapirani ulazni podaci tako da se zadrže samo ključne osobine. Jedna tačka u ovom prostoru je vektor koji predstavlja neki od uzoraka podataka. Tačke koje su blizu jedna druge predstavljaju slične ulazne podatke. Ovaj latentni prostor je neka distribucija, uglavnom normalna/gausova (tj gausova distribucija za svaki uzorak sa ulaza).

Ako nasumično izaberemo neku tačku u ovom prostoru, možemo generisati podatke slične ulaznim podacima. Izlaz iz ovih mreža može takođe biti manipuliran kao i kod običnih autoenkodera, ili možemo iskoristiti kao ulaz za GAN mrežu.

## **5.2 Generativne protivničke mreže (GAN – Generative Adverserial Networks)**

Kod ove arhitekture takođe imamo dve mreže – generator i diskriminator. Generator pokušava da prevari diskriminator i on generiše slike dok diskriminator pogađa da li su slike prave ili lažne/generisane. Generisane slike imaju labelu 0 dok prave imaju labelu 1. Ovaj postupak takođe može da se posmatra kao i minmax igra gde je cilj diskriminatora da prepozna lažne slike a cilj generatora da prevari diskriminator (diskriminator hoće da njegov izlaz bude blizu jedinice, a izlaz generatora nula, dok je cilj generatora obrnut).

Nakon njihovog zajedničkog treniranja koristimo generator da generišemo slike.

**5.3 DCGAN**

Ovde se koriste CNN mreže za generator i diskriminator umesto MLP-a. Generator pomoću dekonvolucionih slojeva može da poveća prostorne dimenzije i da generiše sliku/izlaz koji je detaljan i sa velikom rezolucijom. Diskriminator pomoću konvolucionih slojeva smanjuje prostorne dimenzije tokom ekstrakcije karakteristika.

Pošto osnovna GAN mreža generiše izlaze niske rezolucije, ova mreža se pokazala kao značajno poboljšanje.

# **Zaključak**

Pri rešavanju problema računarskog vida, tehnike augmentacije su nezaobilazne – geometrijske transformacije je skoro uvek moguće upotrebiti dok su ostale tehnike i arhitekture navedene u ovom radu često prateća dopuna.

Neki od problema sa upotrebom ovih tehnika je to što postoji nedostatak teorijskog istraživanja za ovu oblast, kao i to što neki od rezultata nama nisu u potpunosti jasni, dok mrežama poboljšavaju performanse.

# **Literatura**

[1] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, *Deep Image: Scaling up Image Recognition*. 2015. [Online]. Available: https://arxiv.org/abs/1501.02876

[2] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>

[3] T. DeVries and G. W. Taylor, *Improved Regularization of Convolutional Neural Networks with Cutout*. 2017. [Online]. Available: <https://arxiv.org/abs/1708.04552>

[4] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, *Random Erasing Data Augmentation*. 2017. [Online]. Available: <https://arxiv.org/abs/1708.04896>

[5] K. K. Singh, H. Yu, A. Sarmasi, G. Pradeep, and Y. J. Lee, *Hide-and-Seek: A Data Augmentation Technique for Weakly-Supervised Localization and Beyond*. 2018. [Online]. Available: https://arxiv.org/abs/1811.02545