



IMAGE SUPERRESOLUTION

Autori:

Andrea Đukanović 105/2020

Nemanja Divljaković 56/2020

Beograd 2024.



SADRŽAJ

I	Uvod	3
II	Konvolutivna mreža	3
III	Super-rezolucija korišćenjem konvolutivne mreže	4
	1. Priprema podataka	4
	2. Treniranje modela	5
IV	Evaluacija	7
V	Rezultati	8
VI	Zaključak	10
	Literatura	11



I Uvod

Problem „image superresolution“ odnosi se na proces povećanja rezolucije slike rekonstrukcijom njenih detalja. Cilj je generisanje slike visoke rezolucije (u nastavku: HR) iz date slike niske rezolucije (u nastavku: LR). Izazov je tačno predvideti nedostajuće detalje kako bi se postigla određena tačnost ali je potrebno i da rezultat bude što više realističan. Ovaj problem je dodatno otežan jer više HR slika može odgovarati jednoj LR slici, pa je tako određivanje tačnih detalja izgubljenih tokom procesa smanjenja rezolucije – još zahtevnije. O super-rezoluciji se govori još od 1952. godine koristeći razne načine popravljavanja kvaliteta fotografija, a u ovom radu je fokus na korišćenju konvolutivnih mreža.

II Konvolutivna mreža

Konvolutivne neuronske mreže (CNN) su specifična vrsta neuronskih mreža dizajnirana za obradu podataka u obliku slika ili signala. Ključna komponenta CNN-a je **konvolucija**.

Konvolucija matrica $A \in \mathbb{R}^{m \times n}$ i $B \in \mathbb{R}^{p \times q}$ je matrica dimenzija $(m - p + 1) \times (n - q + 1)$ koja se označava $A * B$ i definiše na sledeći način:

$$(A * B)_{i,j} = \sum_{k=0}^{p-1} \sum_{l=0}^{q-1} A_{i+k,j+l} B_{k,l} \quad i = 0, \dots, m-p, j = 0, \dots, n-q$$

Konvolutivna mreža se obučava tako da minimizuje greške svojih izlaza na datim ulazima iz trening skupa, a u tom procesu uči filtere koji realizuju korisne transformacije nad ulaznom slikom. Često je potrebno imati više ulaznih matrica. Niz takvih matrica iste dimenzije je **tenzor**, a pojedinačne matrice su njegovi kanali.

Konvolucija tenzora $A \in \mathbb{R}^{m \times n \times c}$ i $B \in \mathbb{R}^{p \times q \times c}$ je matrica dimenzija $(m - p + 1) \times (n - q + 1)$ koja se definiše se na sledeći način

$$(A * B)_{i,j} = \sum_{k=0}^{p-1} \sum_{l=0}^{q-1} \sum_{t=0}^{c-1} A_{i+k,j+l,t} B_{k,l,t} \quad i = 0, \dots, m-p, j = 0, \dots, n-q$$



Neuronska mreža u svakom sloju uči određeni broj filtera. Skup filtera koji deluju nad istim ulazom nazivamo **konvolutivnim slojem** (eng. convolutional layer). On na svom izlazu daje tenzor čiji svaki kanal odgovara izlazu jednog filtera. Tenzor koji predstavlja izlaz konvolutivnog sloja formira se slaganjem matrica koje su rezultat konvolucije filtera tog sloja sa izlazom prethodnog sloja.

Slaganje matrica je operacija kojom se od K matrica A_1, \dots, A_K dimenzija $m \times n$ dobija tenzor $[A_l | l = 1, \dots, K]$ dimenzija $m \times n \times K$ i definiše se na sledeći način:

$$[A_l | l = 1, \dots, K]_{i,j,k} = (A_k)_{i,j} \quad i = 1, \dots, m; j = 1, \dots, n; k = 1, \dots, K$$

Konvolutivna neuronska mreža je funkcija f_w definisana na sledeći način:

$$h_0 = x$$

$$h_i = a\left(g([W_{ij} * h_{i-1} + B_{ij} | j = 1, \dots, K_i])\right), \quad i = 1, \dots, L$$

$$f_w(x) = \tilde{g}_{\tilde{W}}(h_L)$$

gde je x ulazni tenzor, L broj konvolutivnih slojeva, K_i je broj filtera u sloju i , h_i za $i = 1, \dots, L$ predstavlja izlazni tenzor i -tog sloja, W_{ij} je tenzor parametara j -tog filtera i -tog sloja, B_{ij} matrica slobodnih koeficijenata koja po dimenzijama odgovara matrici $W_{ij} * h_{i-1}$, a je funkcija agregacije, g je aktivaciona funkcija, $\tilde{g}_{\tilde{W}}$ je potpuno povezana neuronska mreža. Tenzori W_i , B_i i vektor \tilde{W} čine parametre W .¹

III Super-rezolucija korišćenjem konvolutivne mreže

1. Priprema podataka

U ovoj fazi koristi se skup slika visoke rezolucije da bi se generisale slike niske rezolucije za treniranje modela. HR slike se prvo smanjuju pomoću bikubne interpolacije, a zatim se vraćaju u prvobitnu veličinu kako bi se simulirali LR izlazi.

¹ Veći deo ovog poglavlja preuzet je iz knjige „Veštačka inteligencija“ autora dr Predraga Janičića i dr Mladena Nikolića, profesora na Matematičkom fakultetu u Beogradu



Bikubna interpolacija korišćena je jer za razliku od jednostavnijih metoda poput najbližeg suseda ili bilinearne interpolacije, uzima u obzir vrednosti piksela iz šireg okruženja, što daje rezultat glatke i prirodne slike.

Ovaj pristup koristi težinsku sumu piksela iz okoline od 4x4 piksela oko ciljnog piksela (ukupno 16 piksela) da bi izračunala novu vrednost piksela. Koristi polinom četvrtog stepena kako bi napravila glatku krivu između piksela, osiguravajući da prelazi između boja budu što prirodniji.

$$f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} P(i, j)$$

Gde je:

- $f(x, y)$ nova vrednost piksela na poziciji (x, y) ,
- $P(i, j)$ vrednosti 16 susednih piksela (4x4 okolina),
- a_{ij} koeficijenti koje određuje bikubna interpolacija, zasnovani na pozicijama piksela i njihovoj udaljenosti od ciljnog piksela.

Nije korišćen unapred generisan LR skup da bi bilo lakše menjati factor smanjenja rezolucije, radi bolje kontrole i testiranja.

2. Treniranje modela

Implementirana je konvolutivna mreža koja se sastoji iz tri konvolutivna sloja:

1. **Prvi konvolutivni sloj:** Ima 64 filtera veličine 9x9 i koristi ReLU aktivacionu funkciju. On ekstrahuje nelinearne karakteristike iz ulazne slike niske rezolucije.
2. **Drugi konvolutivni sloj:** Sadrži 32 filtera veličine 1x1 i takođe koristi ReLU aktivacionu funkciju. Ovaj sloj funkcioniše kao mapiranje nelinearnih karakteristika koje je prethodni sloj izvukao.
3. **Treći konvolutivni sloj:** Ima 3 filtera veličine 5x5 i služi za rekonstrukciju izlazne slike. Konačni rezultat je slika visoke rezolucije koja je poboljšana u poređenju sa ulaznom slikom.



Konvolucija se koristi da prepozna obrasce unutar slike (kao što su ivice i teksture) i koristi te informacije za poboljšanje rezolucije slike.

Loss funkcija

Koristi se kombinovana funkcija greške koja se sastoji od dve komponente:

1. **Srednjekvadratna greška (MSE loss):** MSE meri razliku između piksela slike visoke rezolucije i generisane slike mreže. MSE je formulisana kao:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (I_{SR,i} - I_{HR,i})^2$$

gde je $I_{SR,i}$ piksel generisane slike, $I_{HR,i}$ piksel originalne slike visoke rezolucije, a N ukupan broj piksela.

2. **Perceptual Loss:** Koristi se unapred trenirani VGG19 model za ekstrakciju karakteristika slika i poređenje tih karakteristika između slike visoke rezolucije i generisane slike (SR). Na ovaj način mreža ne samo da uči piksel po piksel, već i semantičke karakteristike slike.
3. Konačna funkcija gubitka je kombinacija *MSE loss* i *perceptual loss*:

$$L_{total} = w_{MSE} \cdot L_{MSE} + w_{preceptual} \cdot L_{preceptual}$$

gde su w_{MSE} i $w_{preceptual}$ težinski koeficijenti koji podešavaju udeo svake komponente u ukupnom gubitku.

Validacija: Nakon svakog koraka treniranja, model se testira na validacionom skupu kako bi se izračunao validacioni gubitak. Ovaj korak pomaže u proceni da li se model prilagođava.

Rano zaustavljanje (Early Stopping): Implementiran je mehanizam ranog zaustavljanja koji prati validacioni gubitak. Ako se validacioni gubitak ne smanji za unapred definisanu vrednost tokom određenog broja epoha, treniranje se prekida kako bi se izbeglo prilagođavanje modela.

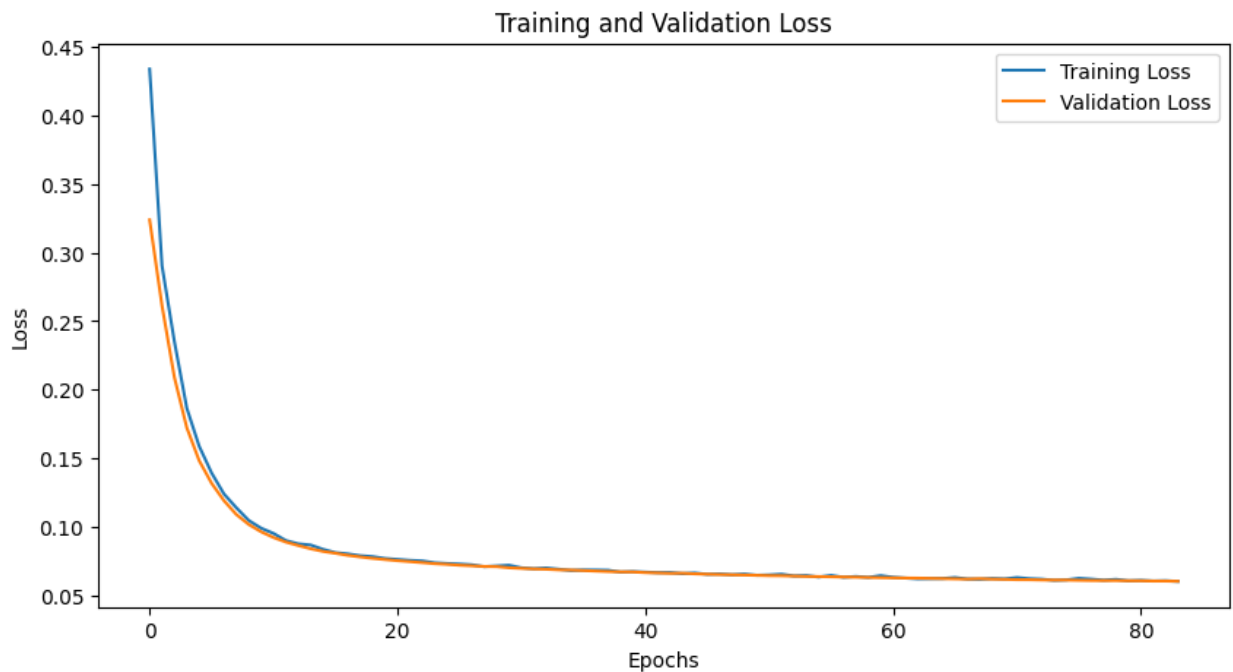


IV Evaluacija

PSNR (Peak Signal-to-Noise Ratio) je jedna od najčešće korišćenih kvantitativnih mera za ocenu kvaliteta slika kod problema super-rezolucije. Ona meri odnos između maksimalne moguće snage signala (u ovom slučaju slike visoke rezolucije) i šuma (razlike između originalne slike i generisane slike). PSNR se računa pomoću sledeće formule:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_{pixel}}{\sqrt{MSE}} \right)$$

gde je MAX_{pixel} maksimalna moguća vrednost piksela (u normalizovanom opsegu od 0 do 1, ova vrednost je 1), MSE je srednja kvadratna greška između slike visoke rezolucije i slike generisane modelom. PSNR daje bolju ocenu što je vrednost viša, što ukazuje na manju razliku između generisane slike i originala.



Slika 1 Grafik loss funkcije kroz epohe



V Rezultati

Low Resolution



Super Resolution



High Resolution



Low Resolution



Super Resolution



High Resolution



Low Resolution



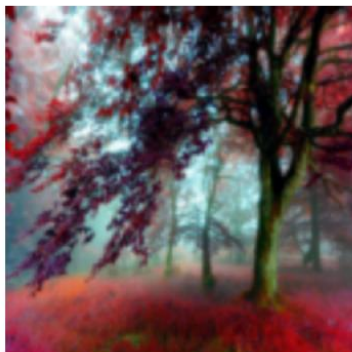
Super Resolution



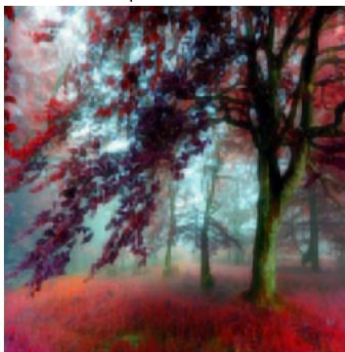
High Resolution



Low Resolution



Super Resolution



High Resolution





Slika 2 PSNR vrednost za test skup

Postignuti rezultati ukazuju na jasno poboljšanje rezolucije slika u poređenju sa ulaznim slikama niske rezolucije, ali kao što se vidi iz niskih vrednosti PSNR kod nekih slika, postoje znantna ograničenja ovog modela.

Uticaj veličine skupa podataka

Jedno od ključnih ograničenja ovog modela je veličina skupa podataka korišćenog za treniranje. Zbog hardverskih ograničenja računara, nije bilo moguće koristiti veći i raznovrsniji skup fotografija što bi pomoglo da se prepozna širi spektar vizuelnih karakteristika, čime bi se poboljšala rekonstrukcija detalja.

Kao što je pokazano u literaturi, veći skupovi podataka značajno doprinose boljoj generalizaciji i kvalitetu generisanih slika. U ovom radu smo izabrali ograničeni skup slika radi optimizacije brzine treniranja i efikasnog korišćenja dostupnih resursa, ali je jasno da bi veći dataset mogao rezultirati još boljim performansama modela.



VI Zaključak

Naprednije metode, poput Generativnih Adverzarijalnih Mreža (GAN), kao što je SRGAN, predstavljaju efikasniji pristup problemu super-rezolucije. Za razliku od klasičnih konvolutivnih mreža, GAN modeli koriste adverzarijalnu funkciju gubitka, što omogućava generisanje vizuelno realističnijih slika sa preciznijim teksturama i detaljima.

Glavna razlika između ovog modela i GAN pristupa leži u funkciji gubitka i načinu obučavanja. CNN model koristi srednju kvadratnu grešku (MSE) kao osnovnu metriku, fokusirajući se na minimiziranje pikselske razlike između slike niske i visoke rezolucije. Iako to daje dobre kvantitativne rezultate (npr. PSNR), očigledno dovodi do "zamazljenih" slika kojima nedostaju visokofrekventni detalji.

GAN metod opisan u literaturi koristi adverzarijalnu funkciju gubitka, koja uključuje dodatni diskriminativni model. Ovakav pristup forsira generator da proizvodi slike koje izgledaju realističnije, dodajući vizuelno važne detalje koje MSE funkcija ne može da uhvati. Kao rezultat, generisane slike često imaju bogatije texture i deluju bliže originalnim slikama visoke rezolucije.

Ova arhitektura zahteva znatno više resursa za obučavanje, uključujući veću količinu memorije i moćniji hardver. Ovakva ograničenja su uticala na izbor modela, s obzirom na dostupnost resursa. Iako model pokazuje solidne rezultate, prelazak na GAN arhitekturu mogao bi značajno unaprediti kvalitet generisanih slika.

Sledeći koraci uključuju istraživanje naprednijih metoda u kontekstu super-rezolucije, korišćenje većih skupova podataka i primenu dodatnih metrika, kao što je SSIM (Structural Similarity Index), za sveobuhvatniju procenu kvaliteta slike.



Literatura

1. Dong, C., Loy, C. C., He, K. & Tang, X. (2016). *Image Super-Resolution Using Deep Convolutional Networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence
2. Maiseli, B. & Abdalla, A. T. (2024). *Seven decades of image super-resolution: achievements, challenges, and opportunities*. EURASIP Journal on Advances in Signal Processing,
3. Yang, J. & Huang, T. *Image super-resolution: Historical overview and future challenges*.
4. Dabović, M., & Tartalja, I. (2017). *Duboke konvolucijske neuronske mreže – koncepti i aktuelna istraživanja*.
5. dr Jančić, P, dr Nikolić, M. (2024). *Veštačka inteligencija*. Univerzitet u Beogradu, Matematički fakultet.
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). *Generative Adversarial Networks*