

# Neural Style Transfer

Seminarski rad u okviru kursa  
Računarska inteligencija  
Matematički fakultet

Ivan Vranković

September 2022

### **Sažetak**

Tema ovog rada je tehnika koja transformise fotografije u umetnička dela koristeći stilove i motive poznatih umetnika i njihovih dela. Kompozicija slika se izvodi pomoću grane mašinskog učenja koja se zove Konvolutivne neuronske mreže. Krajnji rezultat kompozicije je fotografija nacrtana stilom nekog umetničkog dela.

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>4</b>
1.1	Opis problema . . . . .	4
<b>2</b>	<b>Rešenje</b>	<b>4</b>
2.1	Ideja . . . . .	4
<b>3</b>	<b>Izračunavanje greške</b>	<b>4</b>
3.1	Greška gubitka sadržaja . . . . .	4
3.2	Greška gubitka stila . . . . .	5
3.3	Ukupna greška . . . . .	6
<b>4</b>	<b>Model</b>	<b>6</b>
<b>5</b>	<b>Zaključak</b>	<b>7</b>
	<b>Literatura</b>	<b>8</b>

# 1 Uvod

## 1.1 Opis problema

Na početku imamo dve slike. Može se zamisliti da se nekako izdvajaju karakteristike niskog nivoa kao što su boja i tekstura jedne slike (koju ćemo nazvati slika stila) i primenjuju se na semantičnije slike višeg (koju ćemo nazvati slika sadržaja) da bismo došli do slike sa prenetim stilom.

Kako bismo uopšte počeli da postizemo ovako nešto? Mogli bismo da uradimo analizu slike na nivou piksela na slici stila kako bismo dobili boje i aspekte njene teksture. Ali se onda postavlja drugo pitanje - kako da sistemu predstavimo da su teksture i boje koje želimo da izvučemo ustvari na skali poteza četkice, a ne na oblicima u globalu na slici stila? Ali čak i da uspemo ovo da uradimo, pitanje je kako bi uspeli da prenesemo stil sa slike stila na sliku sadržaja, istovremeno održavajući semantiku te slike.

Odgovori na ova pitanja su ponudjeni u radu [1] koji govori da je ovo jedan tip optimizacionog problema.

# 2 Rešenje

## 2.1 Ideja

Proces kombinovanja dve slike se sastoji iz izdvajanja karakteristika slike (sadržaja jedne i stila druge slike), izračunavanja greškaka i spuštanja niz gradijent na rezultujućoj slici.

Izdvajanje karakteristika se postiže provlačenjem slike kroz VGG19 neuralnu mrežu koja je unapred obučena za prepoznavanje objekata na ImageNet skupu podataka, izračunavanje gresaka se izvodi nad tim izdvojenim karakteristikama i na kraju se rezultujuća slika dobija izvodjenjem gradijenta na slici sa belim šumom.

# 3 Izračunavanje greške

## 3.1 Greška gubitka sadržaja

Treba nam funkcija koja će težiti nuli kada su njene dve ulazne slike (slika sadržaja i slika stila) veoma blizu jedna drugoj u smislu sadržaja, a koja raste kako njihov sadržaj odstupa. Ovu funkciju nazivamo funkcija gubitka sadržaja.

$$\mathcal{L}_{\text{content}} \left( \text{, \text{, \right) \approx 0$$

Slika 1: Šematski prikaz greške gubitka sadržaja

Rezultati prikazani u glavnom tekstu [1] su generisani na osnovu VGG mreže. VGG je konvolutivna neuronska mreža čije su performanse prepoznavanja objekata jednaki ljudskom prepoznavanju.

Svaki sloj u mreži generiše nelinearnu matricu filtera čija se složenost povećava sa pozicijom sloja u mreži. Stoga je data ulazna slika sadržaja kodirana u svakom sloju KNN-a primenom filtera na tu sliku.

Sloj sa

$$N_l$$

različitih filtera ima

$$N_l$$

mapa karakteristika, svaka veličine

$$M_l$$

, gde je

$$M_l$$

veličina te mape. Filteri u sloju l mogu biti skladišteni u matrici  $F^l \in R^{N_l \times M_l}$ , gde je

$$R_{ij}^l$$

aktivacija filtera i na poziciji j u sloju l. Neka je  $\vec{p}$  originalna slika, a  $\vec{x}$  slika koja se generiše,

$$P_l$$

i

$$F_l$$

njihove reprezentacije u sloju l. Zatim definišemo srednje kvadratnu grešku kao:

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

Slika 2:

ff

### 3.2 Greška gubitka stila

Treba nam i funkcija koja bi nam rekla koliko su dve slike bliske jedna drugoj u smislu različitosti stila. Opet, ova funkcija raste kako njene ulazne slike divergiraju u pogledu stila. Ovu funkciju nazivamo funkcijom gubitka stila.

$$\mathcal{L}_{style} \left( \text{img1}, \text{img2} \right) \approx 0$$

Slika 3: Šematski prikaz greške gubitka stila

Napravili smo reprezentaciju stila koji izračuvana korelacije između različitih rezultata primene filtera. Ove korelacije mape karakteristika su prikazane Gram matricom  $G^l \in R^{N_l \times M_l}$  na slici 4

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$$

Slika 4:

Neka su  $\vec{a}$  i  $\vec{x}$  originalna slika i slika koja je generisana i neka su

$$A^l$$

i

$$G^l$$

reprezentacije njihovih stilova u sloju l. Uticaj tog sloja na ukupnu grešku je:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

Slika 5:

a ukupni gubitak je:

$$\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l$$

Slika 6:

gde je

$$w_l$$

faktor korekcije za svaki sloj.

### 3.3 Ukupna greška

Da bi se generisala slika koja je rezultat mešanja sadržaja jedne slike i stila druge slike zajedno se minimizuju udaljenost slike sa belim šumom od slike sa sadržajem u jednom sloju mreže i reprezentacije stila druge slike u ostalim slojevima KNN-a. Neka su  $\vec{p}$  fotografija iz koje se izvlači sadržaj i  $\vec{a}$  umetnička slika iz koje se izvlači stil. Funkcija ukupnog gubitka koju pokušavamo da minimizujemo je prikazana na slici 7, gde su

$$\alpha$$

i

$$\beta$$

faktori korekcije za sadržaj i stil.

## 4 Model

Rezultati predstavljeni u glavnom tekstu generisani su na osnovu VGG mreže, sa 16 konvolucionih i 5 pooling slojeva VGG19 mreže. Nije korišćen nijedan od potpuno povezanih slojeva. Za sintezu slike je zamenjena operacija max-pooling operacijom average pooling jer je otkriveno da se poboljšava gradijent protoga i blago se postižu realniji rezultati. Kao optimizator korišćen je Adam.

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x})$$

Slika 7:

## 5 Zaključak

U ovom radu prikazan je prenos umetničkog stila koji prati originalni rad [1]. Međutim, postoji mnogo radova na ovu temu koji pokušavaju da poboljšaju rešenje menjanjem parametara KNN-a, korišćenjem nekih drugih funkcija koje računaju greške ili drugih optimizatora.

## Literatura

- [1] Matthias Bethge Leon A. Gatys, Alexander S. Ecker. A Neural Algorithm of Artistic Style. <https://arxiv.org/pdf/1508.06576.pdf>.