

Image Style Transfer Using Convolutional Neural Networks

Projekat iz Računarske Inteligencije

Matematički fakultet, Univerzitet u Beogradu

Sanja Nedeljković

Sep 2023.

Sadržaj

1 Uvod	2
2 Opis problema	2
3 Metodologija	3
3.1 VGG19	3
3.2 Reprezentacija sadržaja	4
3.3 Reprezentacija stila	4
4 Algoritam	5
5 Rezultati	6
5.1 Jednake vrednosti parametara α i β	6
5.2 Veći fokus na sadržaj	8
5.3 Poređenje fokusa na sadržaj i stil	9
6 Zaključak	10
7 Literatura	10

1 Uvod

U prošlosti je oponašanje specifičnog stila bio težak poduhvat koji je zahtevao veštog umetnika, mnogo vremena i razumevanje umetničkog doba. Umetnik je morao pažljivo da reprodukuje svaki potez četkicom, izbor boje i tekture kako bi tačno uhvatio suštinu stila. Međutim, pred kraj prošlog veka, teorija umetnosti postaje interesantna i informatičarima. Ovo je dovelo do brojnih studija i tehnika, a sve sa zajedničkim ciljem da se obične slike automatski transformišu u umetnost.

U ovom radu prikazana je tehnika za prenos stilova slika, inspirisana radom[1] kojim je predstavljena mogućnost primene ove tehnike pomoću konvolutivnih neuronskih mreža.

2 Opis problema

Prenos stila slike je tehnika koja podrazumeva korišćenje dve slike — sliku sadržaja i referentnu sliku stila (kao što je umetničko delo poznatog slikara) — i njihovo spajanje tako da izlazna slika izgleda kao slika sadržaja, ali je „oslikana“ u stilu referentne slike stila. Prenošenje stila sa jedne na drugu sliku, iako kompleksniji, može se smatrati problemom prenosa tekture slike, čiji je cilj da se izvuku tekture iz slike čiji stil želimo da primenimo, dok sadržaj originalne ulazne slike treba da ostane što jasniji. Ovim procesom nastaje nova slika, koja odgovara datom stilu, dok se odbacuju informacije o globalnom rasporedu scene slike iz koje je taj stil dobijen.

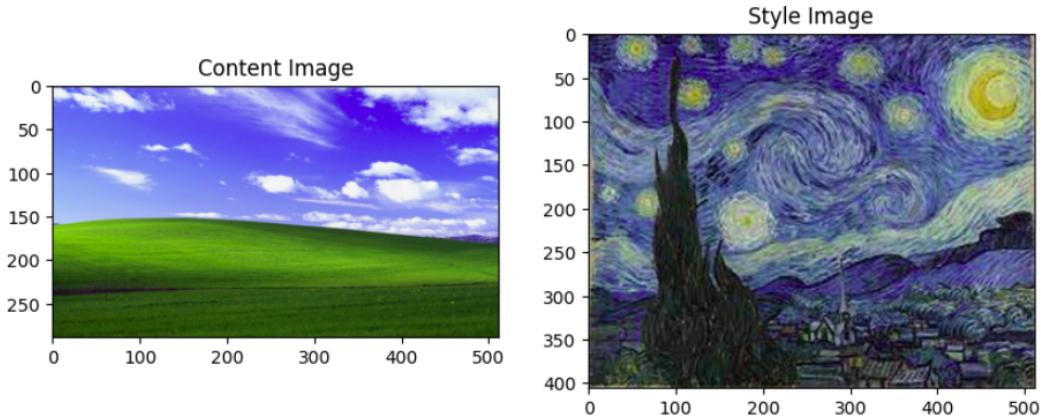


Figure 1: Slika sadržaja (levo), slika stila (desno)

3 Metodologija

3.1 VGG19

Konvolutivne neuronske mreže su kategorija neuronskih mreža koje se ističu zbog svoje efikasnosti, posebno u domenu obrade slika. Ključna prednost jeste njihova mogućnost učenja hijerarhijske (višeslojne i strukturirane) reprezentacije direktno iz slika. Upravo ovakva reprezentacija omogućava mreži da obuhvati i detalje niskog nivoa (npr. ivice, teksture) i koncepte visokog nivoa (npr. objekte, scene) unutar slike. Kada se ove mreže treniraju za prepoznavanje objekata, informacije koje mreža nosi o objektima na slici postaju sve eksplisitnije i kompleksnije sa povećanjem dubine mreže. Zbog ovog svojstva, za problem prenosa stila slike korišćena je VGG19 arhitektura duboke konvolutivne neuronske mreže, koja je obučena za prepoznavanje i lokalizaciju objekata na ImageNet skupu podataka. Sastoji se od 16 slojeva konvolucije razdvojenih sa 5 pooling slojeva i 3 potpuno povezana sloja na kraju. Mreža je korišćena bez tri poslednjia sloja.

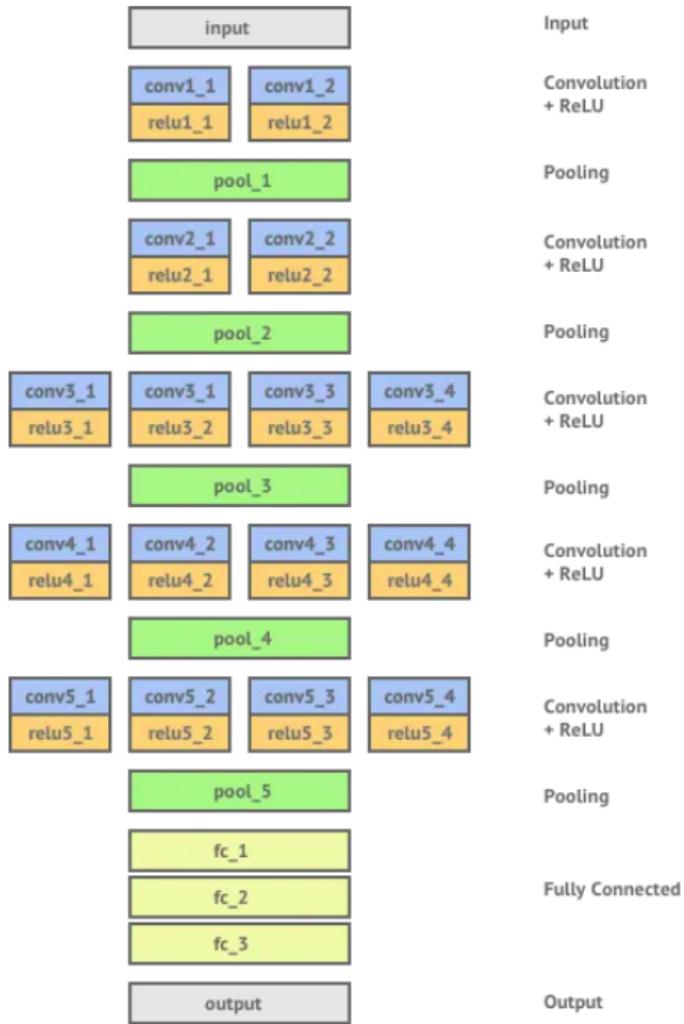


Figure 2: Arhitektura VGG19 mreže

3.2 Reprezentacija sadržaja

U kontekstu ovog problema, reprezentacija sadržaja predstavlja ključni aspekt procesa gde je sadržaj slike neophodno "kodirati" na određeni način kako bi se kasnije preneo na ciljnu sliku. Ova reprezentacija se obično izdvaja iz dubljih slojeva neuronske mreže, jer, kao što je već napomenuto, dublji slojevi su vešti u hvatanju karakteristika visokog nivoa i semantičkih informacija unutar slike. Izdvajanjem prikaza sadržaja iz ovih slojeva je dobijena suština onoga što slika sadrži, u smislu objekata, struktura i njihovih prostornih rasporeda, dok su detalji poput boja i tekstura piksela već uklonjeni.

Moguće je konstruisati slike čije mape atributa(kerneli) na izabranom sloju konvolucije odgovaraju odgovarajućim mapama atributa date slike sadržaja. U tom slučaju loss je definisan kao srednje kvadratna greška između mape atributa slike sadržaja(C_c) i mape atributa generisane slike(T_c). Izbor sloja (ili u nekim slučajevima, slojeva) originalne slike utiče na stepen strukturiranosti slike.

$$\mathcal{L}_{content} = \frac{1}{2} \sum (\mathcal{T}_c - \mathcal{C}_c)^2$$

Figure 3: Content Loss

3.3 Reprezentacija stila

Reprezentacija stila predstavlja nešto kompleksniji zadatak. U kontekstu ovog problema, reprezentacija stila kvantifikuje korelacije između različitih atributa u različitim slojevima mreže. Atributi odgovaraju obrascima i teksturama koji su uhvaćeni na slici, i koji se nalaze na različitim nivoima apstrakcije unutar slojeva. Konstruisanje stila se može odraditi uzimanjem u obzir različitih podskupova slojeva mreže. Povećanjem broja slojeva koji se koriste u ovom procesu se dobija poboljšanje i verniji prikaz umetničkog stila, dok je prostorni raspored manje prisutan.

Korelacija se računa pomoću Gramove matrice, koja je suštinski skalarni proizvod vektora jedne mape atributa. Vrednosti bliske nuli označavaju da između vektora ne postoji korelacija, tj. to su oblasti u kojima su vizuelni obrasci različiti ili manje izraženi u stilu. Sa druge strane, velike vrednosti ukazuju na oblasti slike gde su obrasci ili teksture veoma izraženi, što značajno doprinosi umetničkom stilu. U ovom slučaju, loss je definisan slično kao i kod sadržaja, samo se za računanje srednje kvadratne greške umesto tensora koriste Gramove matrice slika.

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

Figure 4: Gram matrica kernela F

$$\mathcal{L}_{style} = a \sum_i w_i (\mathcal{T}_{s,i} - \mathcal{S}_{s,i})^2$$

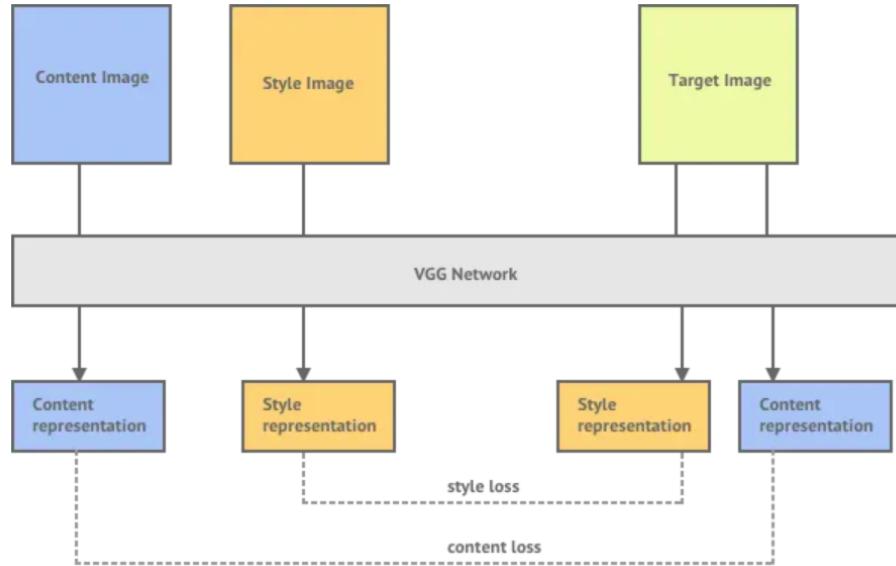
Figure 5: Style Loss

4 Algoritam

Ključni koraci algoritma:

- Slike sadržaja, stila, i ciljna slika:** Odabir dve slike. Slika sadržaja služi kao platno na koje će stil biti primenjen, a slika stila sadrži umetničke karakteristike koje se primenjuju. Ciljna slika se može inicijalizovati na dva načina:
 - Generisanje slučajnog šuma (poželjno dimenzije slike sadržaja).
 - Kopija slike sadržaja.
- Neuronska mreža:** Koristi se unapred obučena mreža, kao što je VGG19, za dobijanje atributa.
 - Biraju se slojevi (uglavnom jedan sloj) za reprezentaciju sadržaja, slika sadržaja se prosleđuje modelu i rezultati se čuvaju.
 - Biraju se slojevi za reprezentaciju stila, slika stila se prosleđuje modelu i rezultati se čuvaju.
 - Ciljna slika se prosleđuje modelu i rezultati se čuvaju.
- Total Loss:** Definišemo Total Loss kao linearnu kombinaciju već definisanih Content Loss i Style Loss. Parametri α i β se koriste da bi balansirali uticaj ove dve funkcije, koje su zbog različitog načina računanja znatno razlikuju u razmeri (ukoliko se ciljna slika generiše nasumično). U praksi, parametar β obično ima veću vrednost.

$$\mathcal{L} = \alpha \mathcal{L}_{style} + \beta \mathcal{L}_{content}$$



- Optimizacija:** Koristi se tehnika optimizacije kao što je L-BGFS ili Adam da bi se minimizovao ukupni loss menjanjem vrednosti piksela ciljne slike kroz iteracije. Ovim postupkom se ne menjaju reprezentacije sadržaja i stila originalnih slika, već samo ciljne slike.
- Krajnji rezultat:** Konačna izlazna slika predstavlja rezultat rada algoritma. Na kvalitet generisane ciljne slike utiče mnoštvo faktora, uključujući izbor sledećih parametara: broj slojeva za reprezentaciju, broj iteracija, metod inicijalizacije, vrednosti α i β , brzina učenja. Pored toga, kompatibilnost dve početne slike igra značajnu ulogu u konačnom rezultatu, gde slike koje dele sličan sadržaj imaju tendenciju da daju bolje rezultate. Na primer, mešanje dve pejzažne slike može da proizvede prihvatljiviji rezultat u poređenju sa kombinovanjem različitih slika, kao što su fotografije zgrade i mačke.

5 Rezultati

U svim primerima ciljna slika je inicijalizovana kao kopija slike sadržaja. Za reprezentaciju sadržaja je korišćen sloj 'block5_conv2', a reprezentaciju stila slojevi 'block1_conv1', 'block2_conv1', 'block3_conv1', 'block4_conv1', 'block5_conv1'. Fiksirani parametri su i broj iteracija koji je 100, kao i brzina učenja koja je postavljena na 10.

5.1 Jednake vrednosti parametara α i β

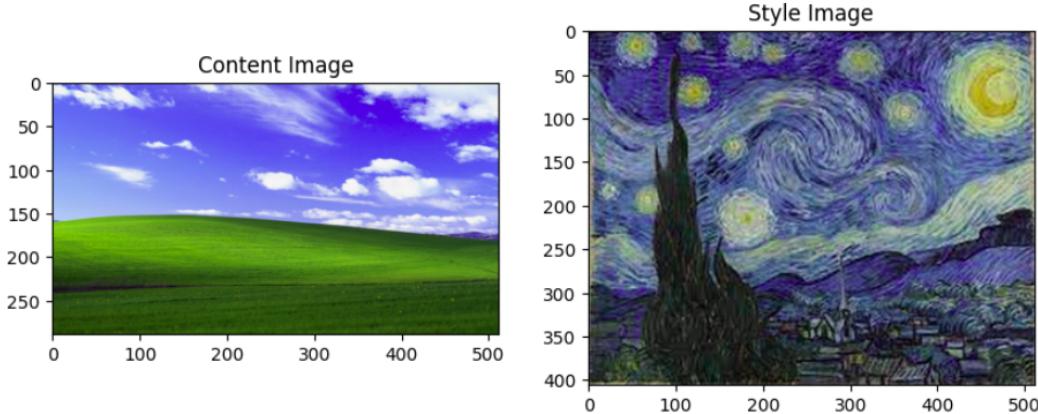


Figure 6: Slika sadržaja (levo), slika stila (desno)

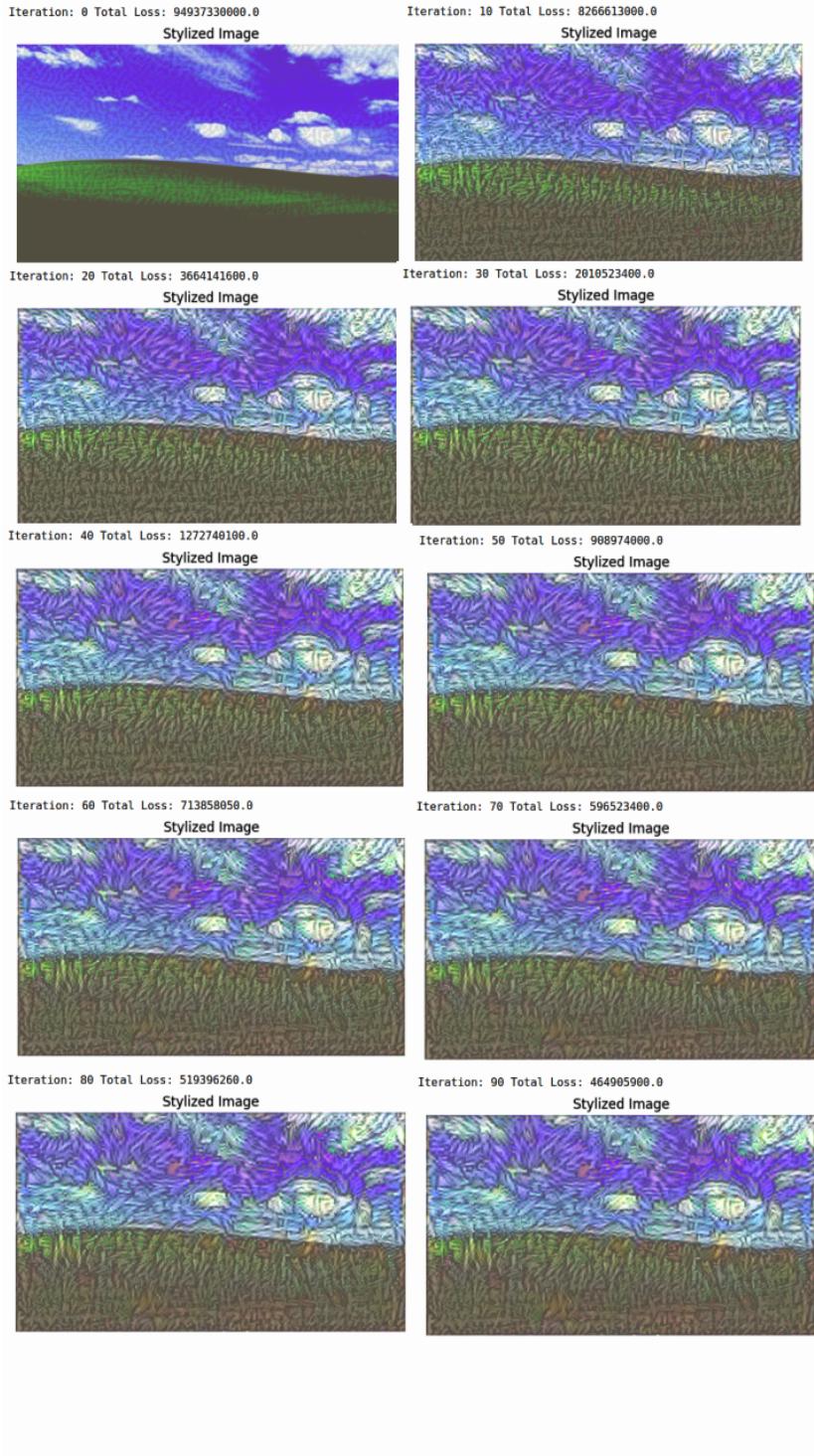


Figure 7: Izmena ciljne slike kroz 100 iteracija.

5.2 Veći fokus na sadržaj

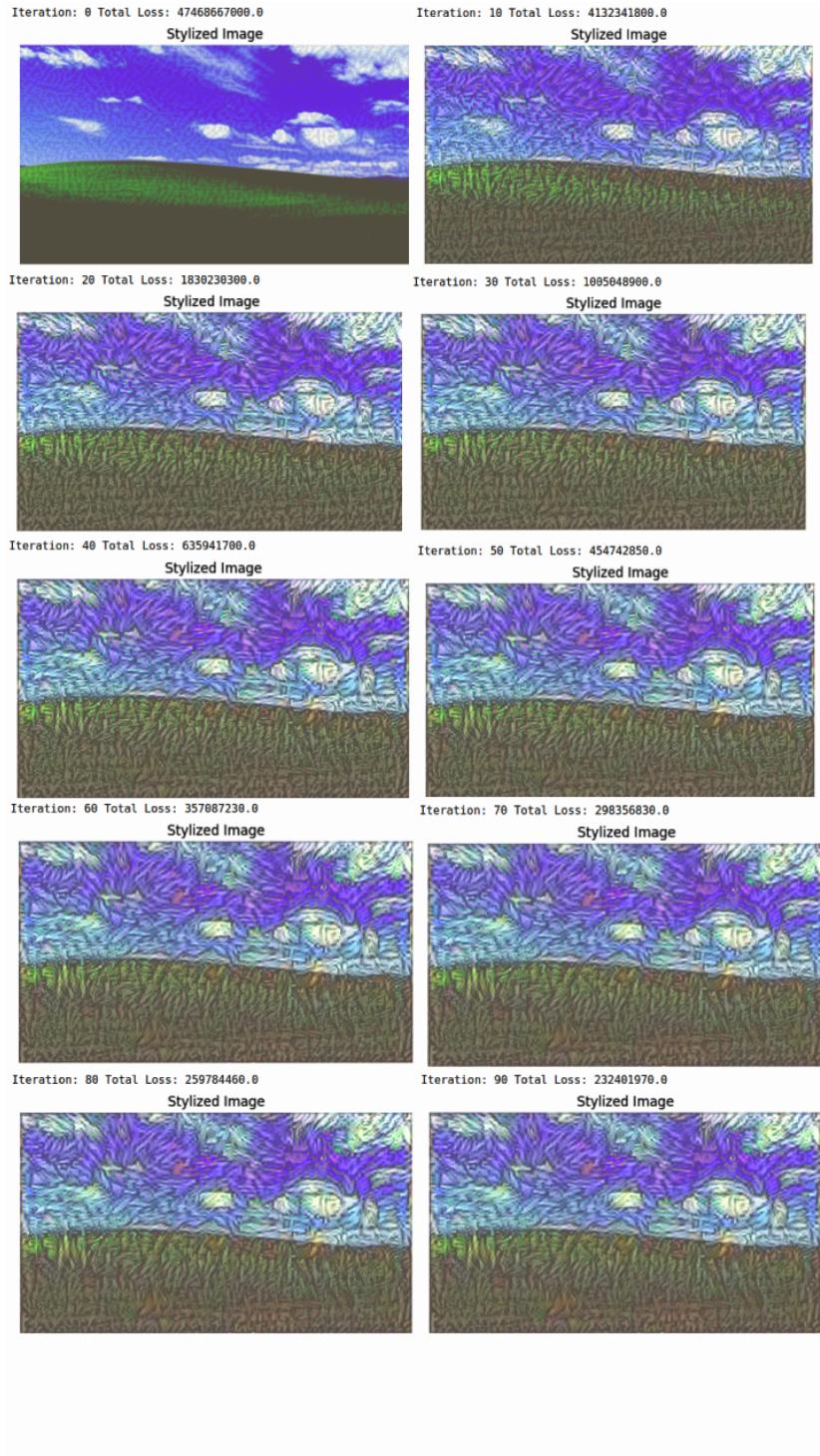


Figure 8: Izmena ciljne slike kroz 100 iteracija.

5.3 Poređenje fokusa na sadržaj i stil

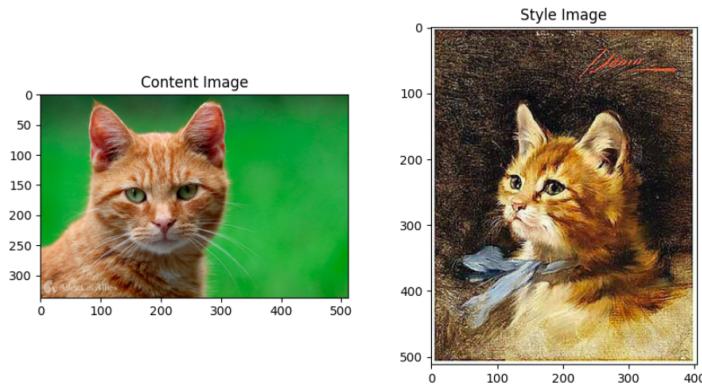


Figure 9: Slika sadržaja (levo), slika stila (desno)

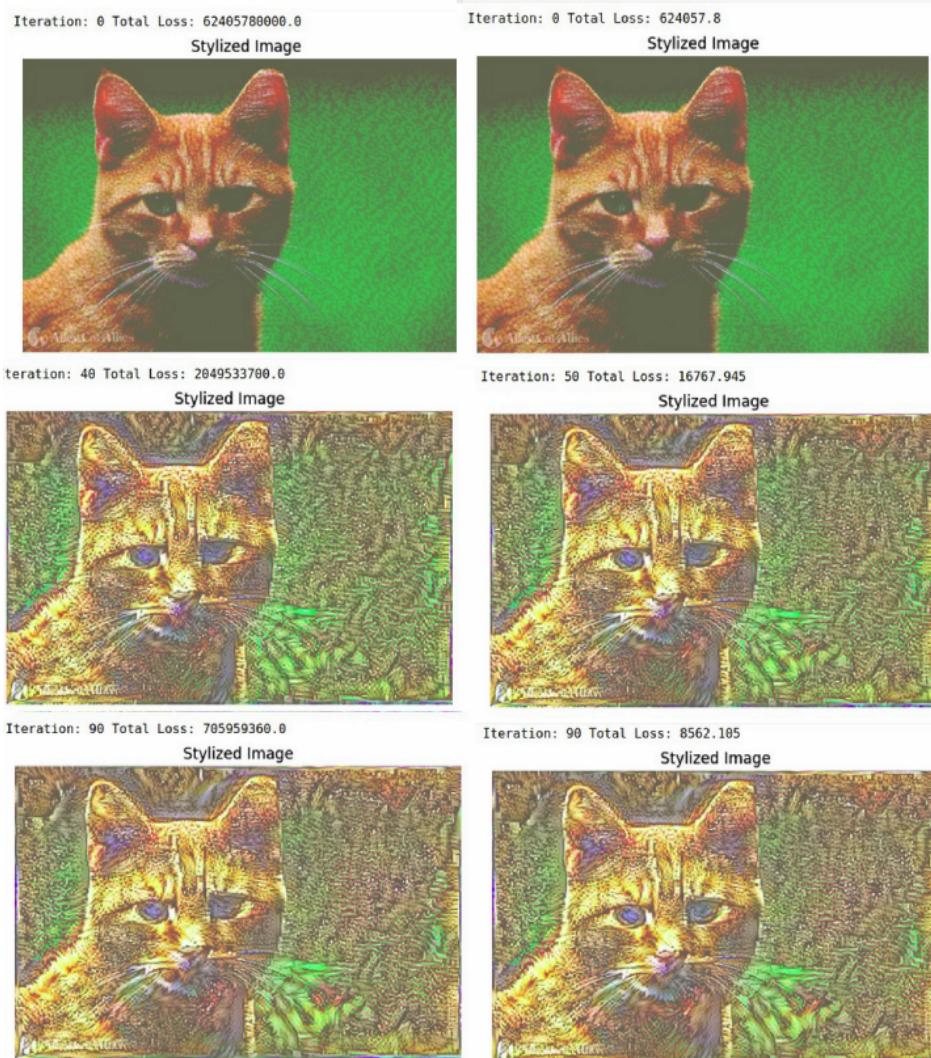


Figure 10: Fokus na stil (levo), fokus na sadržaj (desno)

6 Zaključak

Ključna stvar ovog rada jeste oktriće da su u CNN sadržaj i stil odvojeni. Iako je komplikovano precizno odrediti šta čini stil jedne slike, pomoću ovoga svojstva moguće je generisati potpuno nove slike i imitirati stilove neuporedivo brže nego pomoću ljudskih veština slikanja.

7 Literatura

- [1] [Image Style Transfer Using Convolutional Neural Networks](#)
- [2] [Generative Style Transfer TensorFlow Tutorial](#)