



Proizvoljan Real-time Neuronski Prenos Stila

Stefan Kandić
stefan.kandic@yahoo.com | Soft Kompjuting 2021

Uvod

Neuronski prenos stila je tehnika dubokog učenja koja nam omogućava generisanje slike na osnovu dve ulazne, jedne koja će predstavljati njen sadržaj, i druge koja će predstavljati njen stil.



Postiže se minimizovanjem razlika feature mapa između ulaznih i generisane slike dobijenih na različitim slojevima pretreniranih konvolucionih neuronskih mreža.

Problemi

Nažalost, prvobitna rešenja [1] su bila iterativne prirode, gde je za konvergenciju potrebno više stotina iteracija od kojih svaka uključuje forward i backward pass kroz duboku neuronsku mrežu. Na današnjim procesorima ovo može da traje i do nekoliko sati te su ova rešenja praktično neupotrebljiva za slike velike rezolucije i video zapise.

Do poboljšanja je došlo treniranjem neuronske mreže tako da u jednom prolazu generiše izlaznu sliku [2, 3]. Iako su ovi pristupi doveli do velikog ubrzanja, oni su nažalost ograničeni na unapred zadati skup stilova na kojima je mreža trenirana te su dosta izgubili na fleksibilnosti u odnosu na originalni pristup.

Međutim, pojavio se pristup koji tvrdi da je ipak moguće trenirati neuronsku mrežu tako da prenos stila radi u jednom prolazu i za stilove koje mreža nije nikad videla u toku treniranja, pomoću Adaptivne Normalizacije Instanci (AdaIN) [4]. Moj rad se fokusira na njegovu analizu i implementaciju.

Metodologija

Rešenje se zasniva na enkoder-dekoder arhitekturi, gde su enkoder i dekoder duboke neuronske mreže simetrične jedna u odnosu na drugu.

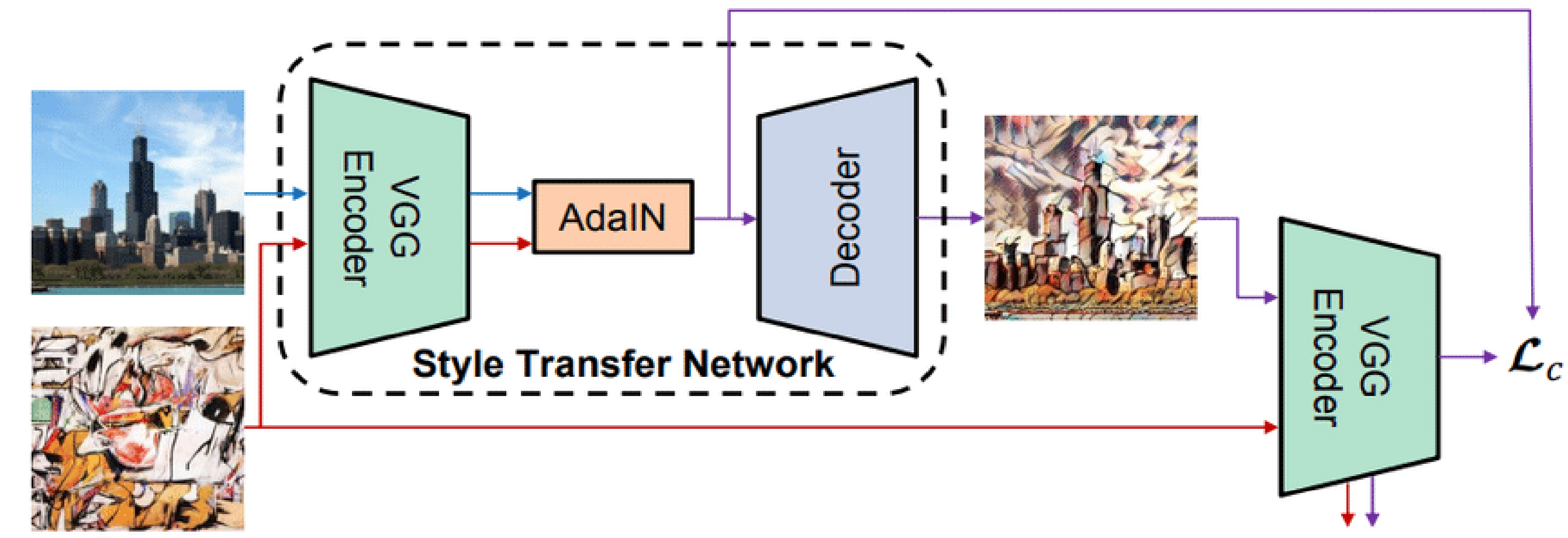
Enkoder je pretrenirana VGG mreža sa dodatim 3x3 konvolucionim slojem na početku koji služi za dodatno preprocesiranje ulazne slike.

Između njih se nalazi AdaIN sloj, čiji ulaz predstavljaju enkodovane slike sadržaja x i stila y i koji poravnava srednju vrednost i standardnu devijaciju sadržaja kako bi bili isti kao i kod stila:

$$AdaIN(x, y) = \sigma(y) \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y)$$

Na ovaj način AdaIN sloj vrši transfer stila u feature prostoru i omogućava dekoderu da generiše sliku za proizvoljnu kombinaciju ulaza sadržaj-stil koje mreža nije videla tokom treniranja.

Pošto je enkoder pretrenirana mreža, a AdaIN sloj nema parametre koji se uče (μ i σ su samo deskriptivna statistika ulaza), dekoder je jedini deo arhitekture koji se uči prilikom treniranja mreže. Njegov zadatak je da invertuje izlaz AdaIN sloja iz feature prostora nazad u prostor slika.



Podaci

Za treniranje se koriste sledeći skupovi podataka:

- za slike sadržaja MS-COCO dataset koji sadrži svakodnevne scene i predmete u svom prirodnom kontekstu
- a za skup stilova Wikiart dataset, koji sadrži slike umetničkih dela

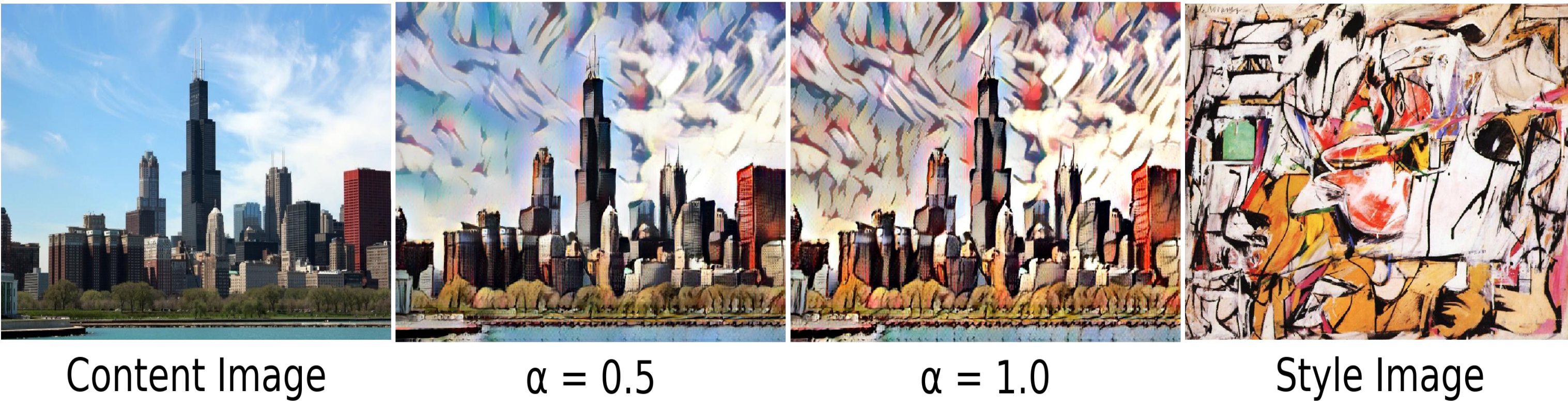
Rezultati

Iako je model treniran samo 20 epoha na skupu od po četrdeset hiljada slika sadržaja i stila, on pokazuje sličan kvalitet generisanih slika kao i [4], koji je treniran čak 160 hiljada epoha na duplo većem broju slika. Takođe, moglo bi se reći da je sam kvalitet slika bolji od slika generisanih u originalnom radu [1].



Balansiranje sadržaja i stila

Takođe, moguće je kontrolisati odnos sadržaja i stila koji će biti preneti na novu sliku pomoću parametra α iz AdaIN sloja. Za $\alpha = 0$, dekoder će pokušati da perfektno rekonstruiše polaznu sliku sadržaja, dok ćemo njegovim povećavanjem dobiti sve veći uticaj stila u izlaznoj slici.



Literatura

1. Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style."
2. Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis."
3. Li, Yijun, et al. "Diversified texture synthesis with feed-forward networks."
4. Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization."