Zbornik radova Fakulteta tehničkih nauka, Novi Sad

**UDK: 004.8.032.26**
**DOI: https://doi.org/10.24867/08BE26Maletin**

# GENERISANJE SLIKA SA VIŠE OZNAČENIH MNIST CIFARA UPOTREBOM GENERATIVNIH NEURONSKIH MREŽA

# GENERATING IMAGES WITH MULTIPLE LABELED MNIST DIGITS USING GENERATIVE NEURAL NETWORKS

Luka Maletin, *Fakultet tehničkih nauka, Novi Sad*

**Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

**Kratak sadržaj –** *Za treniranje generativne neuronske mreže za generisanje slike već postoje ustaljeni šabloni i principi, ali trenutna istraživanja bave se pre svega generisanjem slike sa jednim objektom. Generisanje slike sa više željenih objekata koji se nalaze na određenim pozicijama, i koji su u koheziji sa pozadinom, značajno bi proširilo primenu ovakvih modela. U ovom radu predstavljena je arhitektura modela za generisanje slike sa više označenih cifara na jednostavnoj pozadini. Kao ulaz modela prosleđuju se labele i granični okviri cifara koje je potrebno iscrtati. Istrenirani model uspešno generiše slike sa redosledima cifara koje je video tokom treniranja, a u slučaju novih redosleda ne generiše uvek očekivane cifre. Predloženi su koraci za poboljšanje i dalji razvoj arhitekture.*

**Ključne reči:** *GAN, MNIST, Generisanje slike, Generisanje objekta, Generativni modeli*

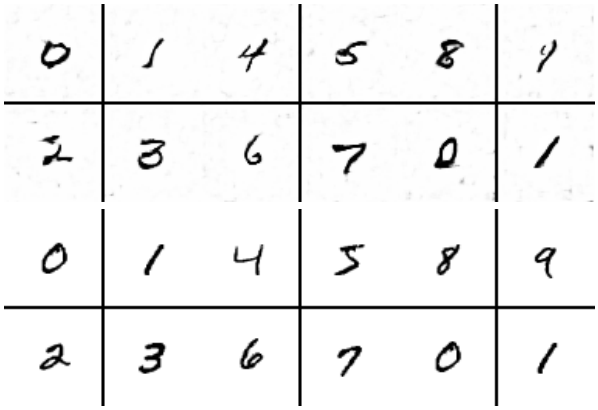
**Abstract** – *When it comes to generating images using generative neural networks, there are already many common methods and principles. However, the current research is focused primarily on generating an image with a single object. Generating images with multiple desired objects on specified locations, and that are in cohesion with the background, would significantly increase the use of such models. In this paper, we present an architecture for generating images with multiple labeled digits on a simple background. The model's inputs are the labels and bounding boxes of the digits that should be painted. The trained model successfully generates images for the orders of digits it has seen during training, while in the case of new orders, it doesn't always generate the specified digits. Steps for further improving the architecture are discussed.*

**Keywords:** *GAN, MNIST, Generating images, Generating objects, Generative models*

## 1. UVOD

Dok diskriminativni modeli mašinskog učenja modeluju uslovnu verovatnoću klase $Y$ datog uzorka $x$, odnosno $P(Y|X=x)$, generativni modeli modeluju zajedničku raspodelu, $P(X, Y)$.

**NAPOMENA:**
**Ovaj rad proistekao je iz master rada čiji mentor je bila dr Jelena Slivka, docent.**



Slika 1. *Primer generisanih slika (gornji red) i primer slika iz skupa podataka sa istim ciframa (donji red). Labele slika su redom 0-1-2-3, 4-5-6-7 i 8-9-0-1.*

Ukoliko je na ovaj način dobro modelovan određeni skup podataka, model je moguće koristiti za generisanje novih uzoraka koji bi odgovarali raspodeli posmatranog skupa podataka. Neki od generativnih modela su model Gausovih mešavina, *Latent Dirichlet allocation* [1], *Variational autoencoder* [2], a u ovom radu korišćene su generativne suparničke mreže (eng. *GAN – Generative Adversarial Network*) [3].

*GAN* čine generator, koji generiše nove uzorke, i diskriminator, koji odlučuje da li je uzorak pravi (iz nekog skupa podataka) ili lažni (generisan od strane generatora). U cilju generisanja obeleženih podataka, odnosno uzoraka određene klase, korišćen je uslovljeni *GAN* (eng. *Conditional GAN*) [4], u kojem i generator i diskriminator na ulazu očekuju labelu.

Jedna od mogućih upotreba *GAN*-ova je generisanje slike primenom arhitekture poput dubokih konvolucionih *GAN*-ova (eng. *DCGAN – Deep Convolutional GAN*) [5]. Kombinovanjem uslovljenog *GAN*-a sa ovakvom arhitekturom, moguće je generisanje slike na kojoj se nalazi željeni objekat, dok je generisanje slike sa više obeleženih objekata značajno manje istražena oblast.

U [6] je predloženo rešenje koje razdvaja generator i diskriminator na po dve putanje (podmreže): globalnu putanju, zaduženu za pozadinu; i lokalnu putanju, zaduženu za pojedinačne objekte. Izlazi konvolucionih slojeva obe putanje spajaju se (konkateniraju) i potom nastavljaju.

Tabela 1. *Arhitektura modela. Za svaki sloj mreže data je dimenzija njegovog izlaza.*

| | | GLOBAL | LOCAL |
|---|---|---|---|
| GENERATOR | **Noise vector** | 100 | |
| | **Label vector** | / | 10 |
| | **Dense, BN, PReLU** | 32x32x256 | 7x7x256 |
| | **Reshape** | 32x32x256 | 7x7x256 |
| | **Conv, BN, PReLU** | 32x32x256 | 7x7x256 |
| | **Upsample** | 64x64x256 | 14x14x256 |
| | **Conv, BN, PReLU** | 64x64x128 | 14x14x128 |
| | **Upsample** | 128x128x128 | 28x28x128 |
| | **Conv, BN, PReLU** | 128x128x64 | 28x28x64 |
| | **Conv** | 128x128x1 | 28x28x1 |
| | **Write object images** | 128x128x1 | |
| DISCRIMINATOR | **Input image** | 128x128x1 | 28x28x1 |
| | **Label vector** | / | 10 |
| | **Conv, PReLU, DO** | 128x128x32 | 28x28x32 |
| | **Conv, PReLU, DO** | 128x128x64 | 28x28x64 |
| | **Conv, PReLU, Flatten, DO** | 128x128x128 | 28x28x128 |
| | **Dense, PReLU, DO** | 128 | 128 |
| | **Dense** | 2 | 2 |
| | **Reshape** | 2 + 8 | |
| | **Write object decisions** | 10 | |
| | **Dense** | 2 | |

Rešenje predloženo u našem radu takođe primenjuje koncept globalne i lokalne putanje. Dok je u [6] zadatak ovih putanja određivanje *feature* mapa, koje se ubrzo konkateniraju i dalje nastavljaju jednim tokom, u našem rešenju obe putanje predstavljaju gotovo ceo generator/diskriminator iz klasične *GAN* arhitekture, čiji rezultati se spajaju na samom kraju. Rezultati treninga nad skupom podataka baziranom na *MNIST* [7] slikama prikazani su u gornjem redu slike 1.U narednom poglavlju detaljnije je predstavljena postojeća literatura, kao i poređenje sa našim rešenjem. Arhitektura modela i korišćene tehnike opisane su u poglavlju 3, a poglavlje 4 sadrži analizu dobijenih rezultata i mogućih poboljšanja. Poglavlje 5 predstavlja sumarizaciju celog rada.

## 2. PRETHODNA REŠENJA

Generisanje slike na kojoj se nalazi više objekata moguće je i primenom običnog *GAN*-a [3], s tim da tada nemamo potpunu kontrolu nad time koji objekti se generišu i gde. U [8] je predloženo rešenje koje ovaj problem delimično prevazilazi. Predloženi model primenom rekurentne neuronske mreže određuje karakteristike prosleđenog teksta (opisa slike), na osnovu kojih se potom primenom *DCGAN* arhitekture dolazi do slike. Na ovaj način omogućeno je upravljanje osobinama objekata, ali ne i njihovih položaja i veličina na slici.

Uslovljeni *GAN* [4] pruža kontrolu klase generisanog objekta, ali njegova klasična arhitektura podržava generisanje samo jednog objekta na slici. Arhitektura predložena u [6] proširuje ovaj koncept time da model na ulazu prihvata proizvoljan broj labela objekata, kao i njihovih graničnih okvira (eng. *bounding box*). Ove informacije se potom pretprocesiraju, kako bi se dobila ulazna slika za generator, na kojoj su iscrtani granični okviri objekata obojeni u različite boje zavisno od njihovih labela. Globalni tok generatora (eng. *global pathway*) na osnovu ove slike, kao i vektora šuma (eng. *noise*), generiše karakteristike slike, a pre svega njene pozadine. Za svaku od labela, primenom lokalnog toka generatora (eng. *local pathway*), generišu se karakteristike objekata. Primenom odvojene neuronske mreže [9], *feature* mape objekata se transformišu u dimenzije odgovarajućih graničnih okvira. Izlazi globalnog i lokalnog toka mreže se potom konkateniraju, nakon čega sledi nekoliko slojeva standardnih za *GAN* arhitekturu. Diskriminator se takođe sastoji iz globalnog i lokalnog toka, gde globalni izvlači karakteristike cele slike, a lokalni, na osnovu labela i graničnih okvira izvlači pojedinačno karakteristike svakog objekta. Ponovo, izlazne karakteristike oba toka se konkateniraju, nakon čega slede slojevi kojima se dolazi do konačne odluke diskriminatora.

Rešenje predstavljeno u ovom radu oslanja se na arhitekturu sa globalnim i lokalnim putanjama iz [6], ali uz razliku u tome šta izlazi tih putanja predstavljaju. Izlazi putanja generatora direktno se upisuju u rezultujuću sliku, najpre izlaz globalne putanje, a potom izlazi lokalnih putanja svakog objekta posebno. Izlazi putanja diskriminatora direktno predstavljaju odluke o pozadini i pojedinačnim objektima, na osnovu kojih se potom donosi konačna odluka za celu sliku.

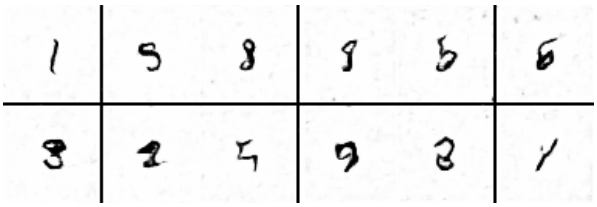
## 3. METOD

U narednim poglavljima izloženi su skup podataka, arhitektura i trening modela.

### 3.1. Skup podataka

Kako je cilj bio isprobati novu arhitekturu za generisanje više objekata na jednoj slici, najpre je kreiran jednostavan skup podataka. Izgenerisano je 15000 slika dimenzija 128x128, sa po četiri 32x32 slike cifara iz *MNIST* skupa podataka [7] na svakoj od njih. Cifre se nalaze u

posebnim ćelijama, međusobno razdvojenim crnim okvirima. Uz svaku sliku zabeležene su i labele sve četiri cifre, kao i njihovi granični okviri. Primer slika iz skupa podataka dat je u donjem redu slike 1.



Slika 2. *Primeri slika sa loše generisanim ciframa. Labele slika su redom 1-9-8-6, 3-2-5-8 i 5-2-7-4*

### 3.2. Arhitektura

Glavni izazov bio je napraviti model koji omogućava generisanje različitih obeleženih objekata na slici sa pozadinom. Arhitektura se oslanja na ideju globalne i lokalne podmreže predloženu u [6], a slojevi mreža prikazani su u tabeli 1. Globalni generator, počevši od vektora šuma (**Noise vector**) ima zadatak da napravi pozadinu. Ovo se postiže naizmeničnim primenjivanjem konvolucionih slojeva (**Conv**) i uveličavanjem prostornih dimenzija, tj. širine i visine (**Upsample**).

Konvolucioni slojevi praćeni su *batch* normalizacionim slojevima (**BN**) [10], a za aktivacionu funkciju korišćen je parametrizovani *ReLU* (**PReLU**) [11]. Potom se lokalni generator primenjuje posebno za svaki objekat, pri čemu su korišćene iste težine u slojevima (eng. *shared weights*). Generator objekata, odnosno lokalni generator, koristi šablon uslovljenog *GAN*-a [4], gde se na ulazu prosleđuje labela željenog objekta (**Label vector**).

Labela se prosleđuje u vidu vektora dužine broja klasa (10 u slučaju cifara), gde se na poziciji koja odgovara željenoj klasi nalazi vrednost 1, a na ostalim pozicijama vrednost 0. Izlazi lokalnog generatora se potom mapiraju na izlaz globalnog generatora na pozicije odgovarajućih graničnih okvira objekata, što daje konačan rezultat mreže generatora.

Diskriminator je takođe podeljen na globalnu i lokalnu podmrežu. Globalni tok posmatra celu sliku i donosi odluku (u vidu 2 neurona) o tome da li je slika prava ili lažna (generisana). Lokalni tok primenjuje se redom na svaki od graničnih okvira i donosi posebnu odluku o verodostojnosti svakog objekta (po 2 neurona).

Na ovaj način, u slučaju četiri objekta, dobija se 10 neurona koji primenom još jednog potpuno povezanog sloja (**Dense**) učestvuju u konačnoj odluci diskriminatora da li se radi o pravoj ili lažnoj slici. Radi boljeg odlučivanja, lokalni diskriminator na ulazu takođe očekuje labelu posmatranog objekta. Za smanjenje rizika od preprilagođavanja (eng. *overfitting*) između konvolucionih slojeva intenzivno su korišćeni dropout slojevi (**DO**) [12].

### 3.3. Trening

Za dobijanje rezultata sa odgovarajućom pozadinom i jasnim ciframa potrebno je oko 25000 epoha, ali je u svrhu eksperimentisanja različitih konfiguracija trenirano i značajno više od toga. Za veličinu *batch*-a isprobane su različite vrednosti, a najbolje se pokazala vrednost 32.

Tokom treninga korak učenja (eng. *learning rate*) je menjan, a najbolje rezultate ostvaruje korak između 1e-6 i 1e-5 za mrežu generatora i 1e-5 i 1e-4 za mrežu diskriminatora. Uprkos korišćenju *batch* normalizacionih slojeva trening je osetljiv na inicijalizaciju parametara, te su početne vrednosti parametara određene empirijski.

## 4. REZULTATI I DISKUSIJA

Model je sposoban da generiše slike sa redosledom cifara koji je video više puta tokom treninga, odnosno, ako se neki redosled (npr. 0-1-2-3) nalazi na slikama u skupu podataka, model će naučiti da generiše slike sa tim redosledom (slika 1, gornji red). U slučaju generisanja slike sa potpuno novim, neviđenim redosledom dešava se da model ne generiše uvek očekivane cifre (slika 2). Ovo ukazuje na potrebu za stavljanjem većeg akcenta na labele tokom treninga.

U [6] se labele ne prosleđuju samo kao vektor, već se generiše slika na kojoj su iscrtani granični oblici objekata obojenih na osnovu njihovih labela, što služi kao ulaz u mrežu. Ovim proširenjem naš model bi naučio da bolje razlikuje cifre.

Trenutan skup podataka sadrži veoma jednostavnu pozadinu koja ne pokriva pozadinu pojedinačnih cifara. Kao sledeći korak trebalo bi napraviti slike sa različitim pozadinama različitih šablona/dezena koji pokrivaju celu površinu. Postojeća arhitektura bi morala da se promeni da na pametniji način kombinuje izlaze lokalnog generatora sa izlazom globalnog generatora, a verovatno bi se pojavila i potreba za dodavanjem dodatnih konvolucionih slojeva nakon spajanja tokova, kako bi se osigurala kohezija objekata sa pozadinom.

## 5. ZAKLJUČAK

U ovom radu predstavljen je model koji omogućava generisanje slike sa više označenih objekata. U svrhu validiranja ideje napravljen je jednostavan skup podataka korišćenjem primeraka iz *MNIST* skupa podataka [7]. Arhitektura modela bazirana je na *GAN* arhitekturi [3], a kombinuje ideju uslovljavanja mreža iz [4], principe važne za uspešno generisanje slike iz [5] i koncept globalne i lokalne podmreže iz [6].

Rezultati pokazuju da model uspešno generiše slike sa konfiguracijama objekata koje je video tokom treninga, ali da pravi greške u slučaju novih konfiguracija. Predloženo je proširenje kojim bi se ovaj problem prevazišao, kao i mogući dalji koraci razvoja u ovom smeru, koji bi omogućili generisanje slika sa kompleksnijim pozadinama.

## 6. LITERATURA


[1] Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp.993-1022.

[2] Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[4] Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[5] Radford, A., Metz, L. and Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
[6] Hinz, T., Heinrich, S. and Wermter, S., 2019. Generating Multiple Objects at Spatially Distinct Locations. arXiv preprint arXiv:1901.00686.
[7] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.
[8] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B. and Lee, H., 2016. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396.
[9] Jaderberg, M., Simonyan, K. and Zisserman, A., 2015. Spatial transformer networks. In Advances in neural information processing systems (pp. 2017-2025).
[10] Ioffe, S. and Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
[11] Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).
[12] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), pp.1929-1958.

## Kratka biografija:

**Luka Maletin** rođen je 1995. godine u Novom Sadu. Osnovne akademske studije završio je 2018. godine na Fakultetu tehničkih nauka, na kom brani i master rad 2019. godine iz oblasti Elektrotehnike i računarstva – Softversko inženjerstvo i informacione tehnologije.

kontakt: lukadjmaletin@gmail.com