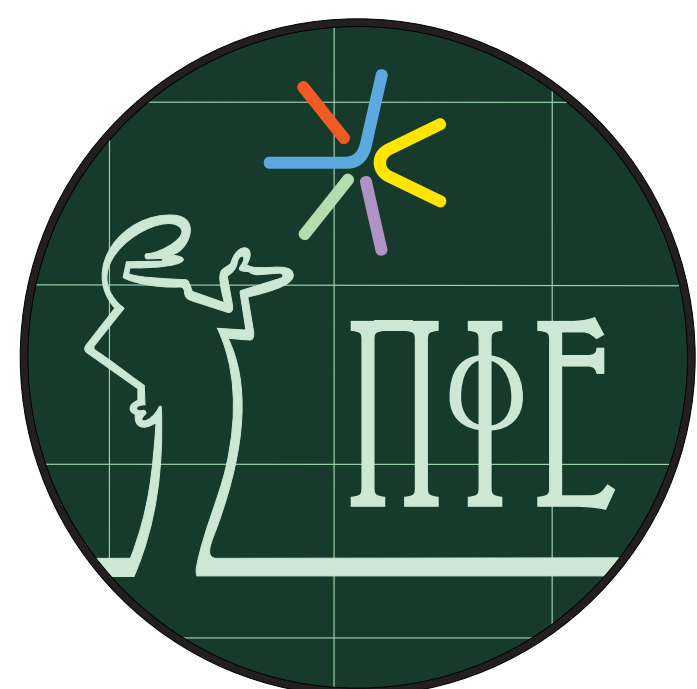


# Generisanje slika suparničkim mrežama



Mihailo Grbić

mihailogrbic99@gmail.com

Matematička gimnazija Beograd

mentor : Miloš Stojanović

milos-stojanovic@outlook.com

## Uvod

Jedan od glavnih prepreka mašinskog učenja jeste sakupljanje i označavanje podataka. Prilikom treniranja modela kompjuterske vizije koriste se preko stotine hiljada slika, sam proces pripremanja tih slika za treniranje modela zahteva veliku količinu vremena, novca i radne snage. Međutim, sa skorašnjim napretcima u tehnologiji i grafici javlja se mogućnost korišćenja veštački stvorenih slika. Korišćenje sintetičkih slika stvorenih kompjuterskim programima rešava probleme sa realnim podacima i obećava brže i jeftinije treniranje modela. Nažalost, veštačke ne postižu dovoljan nivo fotorealizma i zato modeli trenirani na sintetičkim podacima postižu lošije rezultate nego oni trenirani na realnim podacima. Kao rešenje za ovaj problem javilo se nekoliko tehnika boljeg sintetisanja slika. Cilj ovog rada jeste ispitivanje napredne tehnike sintetisanja slika bazirane na upotrebi modela Generativnih suparničkih mreža (GAN).

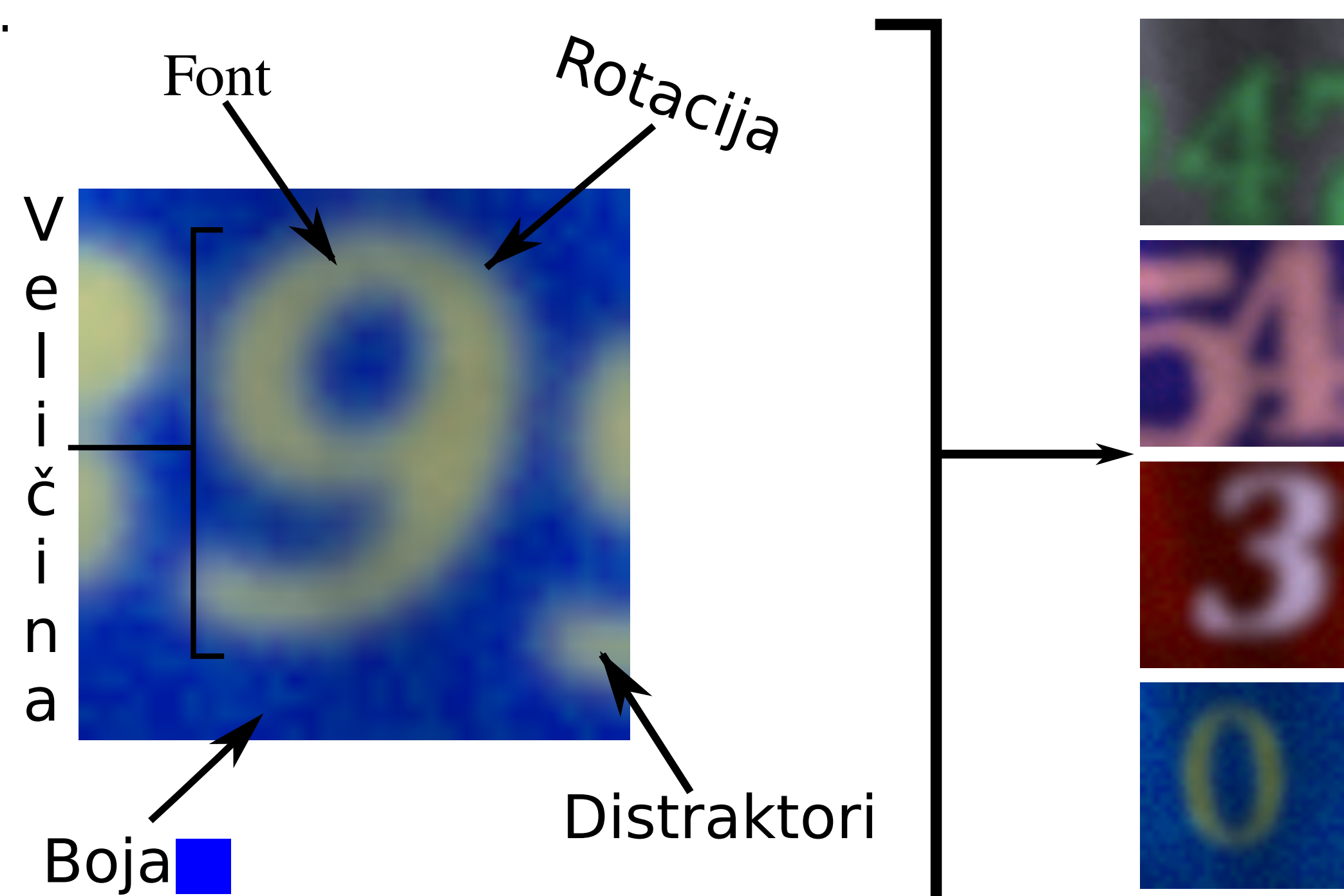
## Slike kućnih brojeva

Problem za koji pokušavamo da generišemo podatke je klasifikacija broja na osnovu 64x64 RGB slike kućnog broja uzete iz Google Street View aplikacije. Korišćena je Street View House Numbers (SVHN) baza podataka sa 73 257 slika kućnih brojeva.



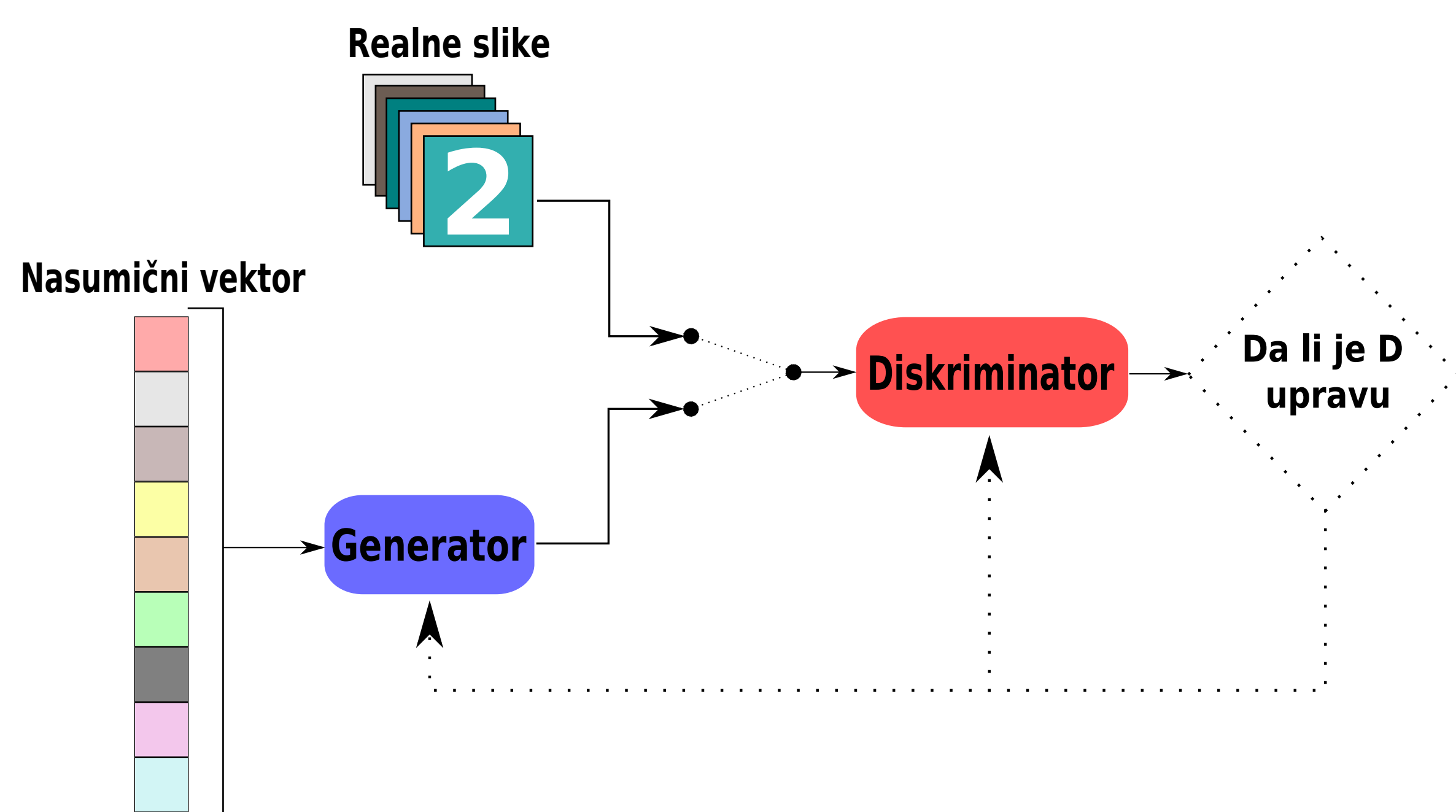
## Bazično sintetisanje

Sintetisanje slika bez korišćenja naprednih tehnika. Slike su generisane u programu Processing. Prilikom stvaranja slika kućnih brojeva varirani su: boja broja i pozadine, font, veličina, položaj i rotacija broja, nasumični distraktori i šum slike.



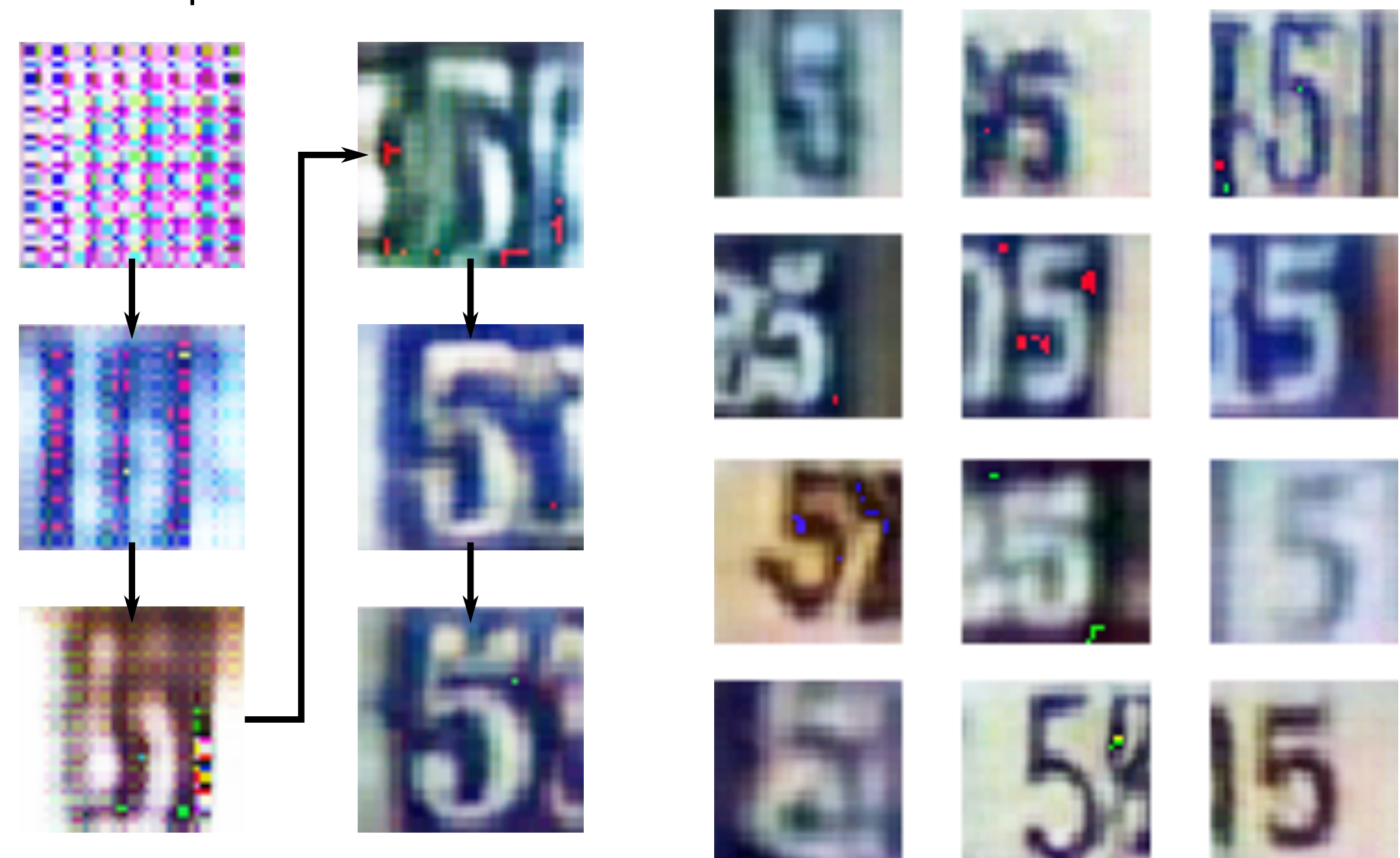
## Generativne suparničke mreže

GAN arhitektura podrazumeva korišćenje i uzajamno treniranje dva modela neuronskih mreža Generatora (G) i Diskriminatora (D). G je mreža dekonvolucionih slojeva koja se trenira da stvori slike što sličnije realnim. G kao ulaz prima nasumični šum od 100 brojeva a kao izlaz izbacuje 64x64 RGB sliku. D je mreža konvolucionih slojeva koja se trenira da razazna prave slike od onih koje je stvorio G. D kao ulaz prima 64x64 RGB sliku, a izbacuje jednu vrednost koja odgovara proceni verovatnoće da je uneta slika realna. G i D se treniraju istovremeno tako da napretku G-a odgovara gubitak D-a i obrnuto. U idealnom slučaju nakon dovoljno treniranja G bi trebalo da stvara slike istog kvaliteta kao realne, a da D postiže tačnost od 50%.



## Rezultati i diskusija

GAN model je nakon treninga uspešno generisao verodostojne slike kućnih brojeva. Troslojna konvoluciona neuronska mreža trenirana na 73 257 slika iz SVHN baze podataka postiže u zadatku prepoznavanja broja postiže 86% tačnosti. Neuronska mreža iste arhitekture trenirana na istom broju bazično sintetisanih slika postiže tačnost od 63% dok mreža trenirana na GAN generisanim slikama postiže tačnost od 77%.



## Literatura

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherilj Ozair, Aaron Courville, Yoshua Bengio; "Generative Adversarial Networks", Universite de Montr'eal

Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet; "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks", Google Inc.