



Univerzitet u Novom Sadu

Fakultet tehničkih nauka

## Principi prezentacije i prepoznavanje oblika

### Generative Adversarial Networks (GAN) za generisanje slika LEGO figurica

**Profesori:** prof. dr Marko Arsenović,  
prof. dr Anderla Andraš

**Asistent:** Sofija Đorđević

**Student:** Anđela Mrđa IT7-2022

**Smjer:** Inženjerstvo informacionih sistema

Novi Sad, januar 2026.

## Sadržaj

1. Uvod .....	3
2. Osnovni pojmovi i teorijska osnova GAN-a .....	4
3. Pregled istraživanja o primjeni GAN modela u generisanju slika .....	7
4. Najčešće korišćeni algoritmi i modeli generativnih adversarijalnih mreža .....	8
5. Skupovi podataka, metrike i ograničenja.....	10
6. Vizija i komentar .....	11
7. Literatura .....	12

# 1. Uvod

Razvoj dubokog učenja doveo je do značajnog napretka u oblasti vizuelne percepcije računarskih sistema, naročito u zadacima kao što su klasifikacija, detekcija i segmentacija slika. Pored ovih zadataka, sve veću pažnju privlače generativni problemi, čiji je cilj stvaranje novih podataka koji su statistički i vizuelno slični postojećim uzorcima. Generisanje realističnih slika predstavlja jedan od najsloženijih problema u ovoj oblasti, budući da zahtijeva razumijevanje strukture, teksture, boje i međusobnih odnosa unutar vizuelnog sadržaja.

Generativni modeli imaju široku primjenu u savremenim sistemima vještačke inteligencije. Oni se koriste za sintezu slika, proširenje skupova podataka (data augmentation), poboljšanje kvaliteta postojećih podataka, kao i u kreativnim i industrijskim oblastima, uključujući dizajn, animaciju i igre. Među različitim pristupima generativnom modelovanju, generativne adversarijalne mreže (GAN – *Generative Adversarial Networks*) izdvajaju se kao jedan od najuticajnijih i najuspješnijih modela za generisanje vizuelnog sadržaja visoke rezolucije i realističnog izgleda.

Kada govorimo o adversarijalnim mrežama, dolazimo do ideje o postojanju dva takmičarska neuronska modela - generatora i diskriminatora. Generator je dio mreže koji ima zadatak da proizvodi nove uzorke podataka, dok je uloga diskriminatora da razlikuje generisane uzorke od stvarnih podataka iz skupa za treniranje. Kroz čitav ovaj proces, oba modela se istovremeno unapređuju, što kao rezultat daje realističnije generisane podatke.

U ovom radu se razmatra primjena generativnih adversarijalnih mreža na problem generisanja slika LEGO figurica. LEGO figurice predstavljaju zanimljiv i moderan vizuelni domen zbog svoje velike raznolikosti u bojama, oblicima, pozama i dodatnim detaljima, kao i zbog jasne strukture objekata. Takav skup podataka omogućava istraživanje sposobnosti GAN modela da generišu koherentne i vizuelno uvjerljive slike objekata koji imaju prepoznatljivu formu, ali istovremeno i značajan stepen varijabilnosti.

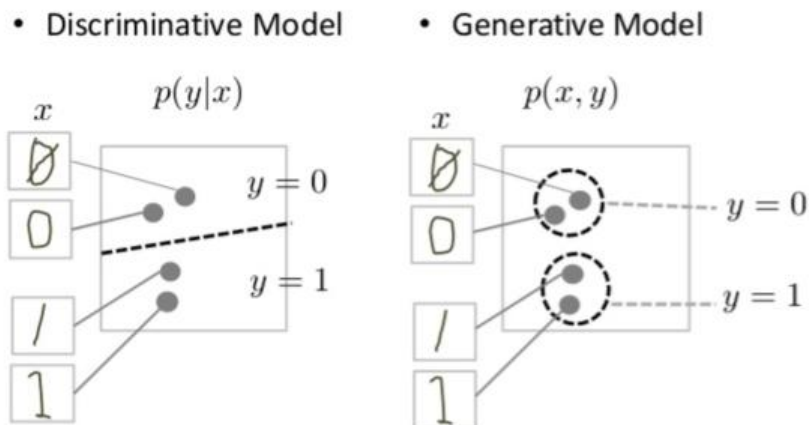
Cilj prve faze projekta je proučavanje teorijskih osnova generativnih adversarijalnih mreža, kao i postojećih pristupa i varijanti ovih modela koje su do sada korišćene za generisanje slika. Poseban akcenat stavljen je na razumijevanje principa rada GAN arhitektura, tipičnih skupova podataka i metrika za evaluaciju kvaliteta generisanih slika. Ova teorijska osnova predstavlja temelj za kasnije faze projekta, u kojima će biti razmatrana implementacija modela i analiza dobijenih rezultata.

## 2. Osnovni pojmovi i teorijska osnova GAN-a

**Generativni modeli** predstavljaju klasu modela mašinskog učenja čiji je cilj učenje distribucije podataka, kako bi se na osnovu naučenog obrasca mogli generisati novi, prethodno neviđeni uzorci. Za razliku od drugih diskriminativnih modela (*klasifikacija, segmentacija, detekcija*) koji su usmjereni na razlikovanje ili klasifikovanje postojećih podataka, generativni modeli nastoje da opišu kako su podaci raspoređeni u prostoru osobina. Upravo ova sposobnost je ono što čini generativne mreže posebno pogodnim za zadatke koji uključuju sintezu teksta, govora, muzike, slika i drugih kompleksnih oblika podataka.

U formalnom smislu, generativni modeli nastoje da aproksimiraju zajedničku raspodjelu podataka i oznaka, označenu kao  $p(X, Y)$ , ili samo raspodjelu podataka  $p(X)$  u slučajevima kada oznake nisu dostupne. Nasuprot tome, diskriminativni modeli uče uslovnu vjerovatnoću  $p(Y | X)$ , odnosno fokusirani su isključivo na predviđanje oznake na osnovu ulaznih podataka. Ova razlika nam govori da generativni modeli moraju naučiti znatno složeniju strukturu podataka, jer obuhvataju i međusobne zavisnosti između osobina, a ne samo granice razdvajanja između klasa.

Ilustracija razlike između generativnih i diskriminativnih modela prikazana je na slici (*Slika 1*), gdje diskriminativni model uči granicu razdvajanja između klasa, dok generativni model nastoji da modeluje raspodjelu podataka unutar svake klase.



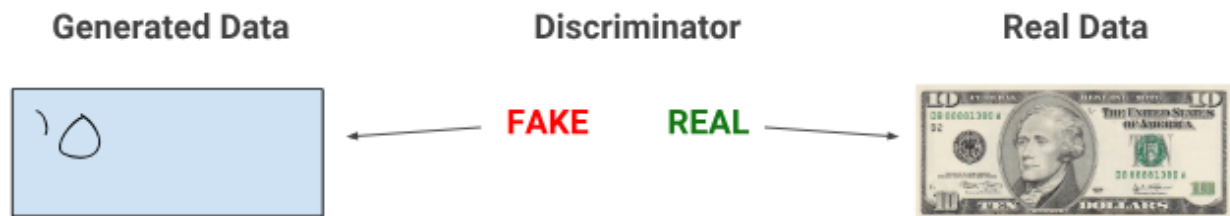
Slika 1: Razlika između diskriminativnih i generativnih modela

Generativne adversarijalne mreže (GAN) predstavljaju specifičnu klasu generativnih modela, prvi put predstavljenu u radu Goodfellowa <sup>1</sup> i saradnika. Osnovna ideja GAN arhitekture zasniva se na adversarijalnom učenju dva neuronska modela: **generatora** i **diskriminatora**.

Generator ima zadatak da proizvodi sintetičke uzorke koji podsjećaju na stvarne podatke iz skupa za treniranje, dok je diskriminator obučen da razlikuje stvarne podatke od onih koje je proizveo generator.

Tokom procesa učenja, generator i diskriminator se nalaze u stalnoj konkurenciji. Generator nastoji da proizvede uzorke koji će zavarati diskriminator, dok diskriminator pokušava da unaprijedi svoju sposobnost razlikovanja stvarnih i generisanih podataka. Ovakva postavka vodi ka situaciji u kojoj generator, ukoliko je proces učenja stabilan i uspješan, proizvodi sve realističnije uzorke koji se teško razlikuju od stvarnih podataka.

Proces treniranja GAN-a može se intuitivno objasniti kroz faze učenja. Na početku treniranja, generator proizvodi očigledno nerealistične uzorke, te diskriminator vrlo lako prepoznaje generisane podatke kao lažne (Slika 2).



Slika 2: Početak generisanja podataka

Kako treniranje napreduje, generator postepeno unapređuje kvalitet svojih izlaza, dok diskriminator postaje sve manje siguran u svoje odluke (Slika 3).



Slika 3: Generator počinje da generiše vjerodostojnije slike

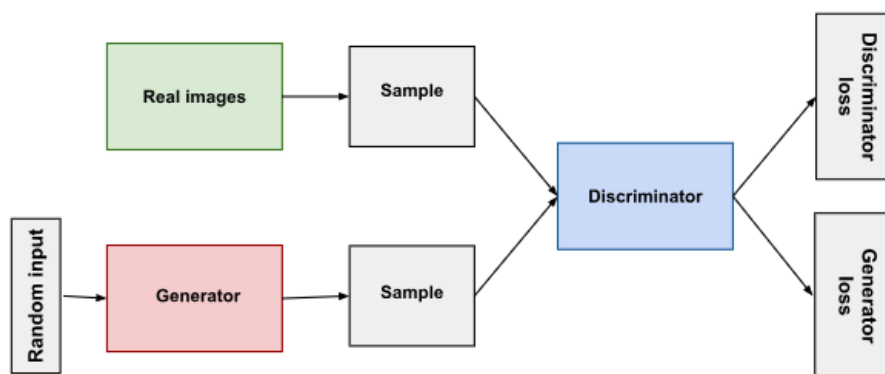
U idealnom slučaju, dolazi se do ravnoteže u kojoj diskriminator više ne može pouzdano razlikovati stvarne i generisane uzorke (Slika 4).

<sup>1</sup> I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville i Y. Bengio, "Generative Adversarial Networks," *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672–2680.



Slika 4: Diskriminator ne može razlikovati generisanu sliku od stvarne

Cjelokupna arhitektura generativne adversarijalne mreže sastoji se od dva osnovna toka podataka. Prvi tok obuhvata stvarne podatke koji se uzorkuju iz skupa za treniranje i proslijeđuju diskriminatoru. Drugi tok započinje slučajnim ulazom, najčešće iz latentnog prostora, koji generator transformiše u sintetičke uzorke. I stvarni i generisani uzorci se zatim proslijeđuju diskriminatoru, čiji izlaz predstavlja procjenu vjerovatnoće da je ulazni uzorak stvaran. Povratna propagacija greške omogućava ažuriranje parametara oba modela tokom treniranja (Slika 5).



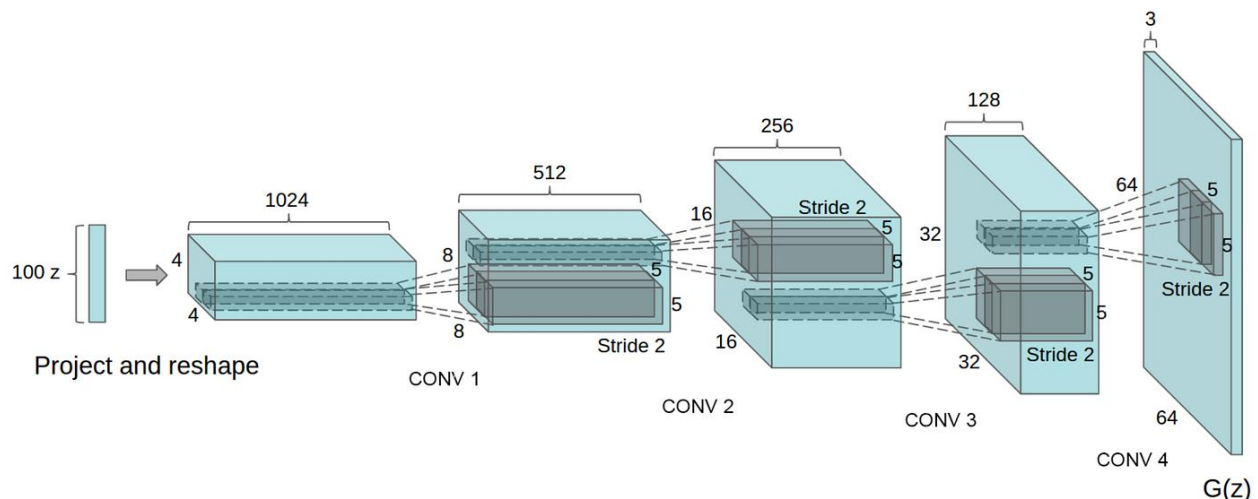
Slika 5: Cjelokupna arhitektura generativne adversarijalne mreže

Važno je naglasiti da su i generator i diskriminator najčešće realizovani kao **neuronske mreže** (*modeli mašinskog učenja inspirisani načinom rada bioloških neuronskih sistema*), te da njihova arhitektura može varirati u zavisnosti od tipa podataka i konkretne primjene. Ova fleksibilnost omogućava primjenu GAN modela ne samo na slike, već i na druge vrste podataka, uključujući vremenske serije, grafove i trodimenzionalne strukture. Međutim, upravo zbog složenosti zadatka modelovanja cjelokupne raspodjele podataka, generativni modeli, a naročito GAN-ovi, suočavaju se sa brojnim izazovima u pogledu stabilnosti treniranja i pokrivenosti modova podataka, što će biti detaljnije razmatrano u narednim poglavljima.

### 3. Pregled istraživanja o primjeni GAN modela u generisanju slika

Razvoj generativnih adversarijalnih mreža započinje radom Goodfellowa i saradnika iz 2014. godine, koji je uveo adversarijalni okvir učenja generativnih modela kroz istovremeno treniranje generatora i diskriminatora. Ovaj pristup je pokazao da se kompleksne distribucije podataka mogu implicitno naučiti bez eksplicitnog modelovanja vjerovatnoće, što je predstavljalo značajan iskorak u oblasti generativnog modelovanja. Rani eksperimenti u ovoj oblasti fokusirali su se na relativno jednostavne skupove podataka, ali su jasno ukazali na potencijal GAN pristupa u sintezi vizuelnih sadržaja.

Kako su se istraživanja razvijala, fokus literature se pomjerao ka unapređenju stabilnosti treniranja i kvaliteta generisanih slika. Posebno značajan doprinos u tom kontekstu predstavlja uvođenje konvolucionih arhitektura u GAN modele (*Slika 6*), što je dovelo do boljeg iskorišćenja prostorne strukture slika i znatno realističnijih rezultata. Ovi radovi su pokazali da arhitektonski izbori imaju ključnu ulogu u uspješnosti GAN modela, te su postavili temelje za njihovu širu primjenu u zadacima obrade slika.



*Slika 6: Uvođenje konvolucionih arhitektura u GAN modele*

Paralelno sa arhitektonskim unapređenjima, veliki dio literature bavio se teorijskim i optimizacionim izazovima adversarijalnog učenja. Predloženi su različiti pristupi koji nastoje da obezbijede stabilniju dinamiku učenja i pouzdaniji signal za optimizaciju, čime se ublažavaju problemi poput nestabilnosti i oscilovanja tokom treniranja. Ovi radovi su doprinijeli boljem razumijevanju ponašanja GAN modela i omogućili njihovu primjenu na složenijim i raznovrsnijim skupovima podataka.

Sa razvojem stabilnijih GAN varijanti, istraživanja su se proširila i na širi spektar slikovnih zadataka. GAN-ovi su se afirmisali ne samo u sintezi slika „od nule“, već i u zadacima transformacije između vizuelnih domena, kao što su prevođenje slike u sliku, prenos stila i manipulacija vizuelnim atributima. Ovakvi pristupi su pokazali da GAN modeli mogu efikasno naučiti relacije između različitih domena čak i u uslovima ograničene ili neuparene obuke, što je značajno proširilo njihovu praktičnu upotrebljivost.

Napredak u kvalitetu generisanih slika dodatno je podstaknut arhitekturama koje omogućavaju finiju kontrolu procesa generisanja i bolju organizaciju latentnog prostora. Ovi modeli su demonstrirali da se razdvajanjem globalnih strukturalnih karakteristika od lokalnih detalja može postići veća raznovrsnost i viši perceptivni kvalitet generisanih uzoraka. Takvi pristupi su naročito značajni za vizuelne domene u kojima je važno očuvati prepoznatljivu strukturu objekata, uz istovremeno generisanje novih varijacija.

U kontekstu ovog projekta, prethodna literatura jasno pokazuje da su GAN modeli posebno pogodni za probleme generisanja slika sa izraženom strukturom i vizuelnom raznolikošću. Iako su izazovi poput stabilnosti treniranja i pokrivenosti modova i dalje prisutni, dosadašnja istraživanja ukazuju da se pažljivim izborom arhitekture i pristupa treniranju mogu postići smisleni i vizuelno uvjerljivi rezultati. Ova saznanja predstavljaju teorijsku osnovu za dalju primjenu GAN modela na domen slika LEGO figurica, koji kombinuje jasnu strukturu objekata sa velikim brojem mogućih varijacija.

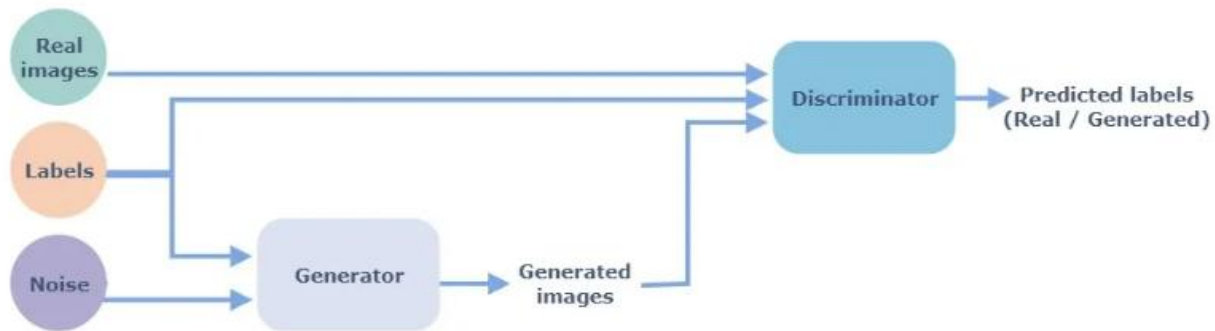
## 4. Najčešće korišćeni algoritmi i modeli generativnih adversarijalnih mreža

Nakon uvođenja osnovnog GAN okvira, istraživanja su se brzo usmjerila ka razvoju različitih varijanti i arhitektura koje nastoje da unaprijede kvalitet generisanih slika, stabilnost treniranja i kontrolu nad procesom generisanja. Tokom godina, predložen je veliki broj GAN modela, pri čemu su se neki izdvojili kao posebno značajni zbog svoje praktične primjene u generisanju slika. U ovom poglavlju prikazani su najčešće korišćeni GAN modeli koji su obilježili razvoj oblasti i koji su relevantni za probleme sinteze vizuelnih podataka.

Jedna od prvih i najuticajnijih varijanti je **Deep Convolutional GAN (DCGAN)**, koji uvodi sistematsku upotrebu konvolucionih neuronskih mreža u arhitekturu generatora i diskriminatora. Umjesto klasičnih potpuno povezanih slojeva, DCGAN koristi konvolucije i dekonvolucije, čime se omogućava efikasnije učenje prostorne strukture slika. DCGAN se često smatra polaznom tačkom za savremene GAN eksperimente na slikama, jer kombinuje relativnu jednostavnost sa dobrim performansama.



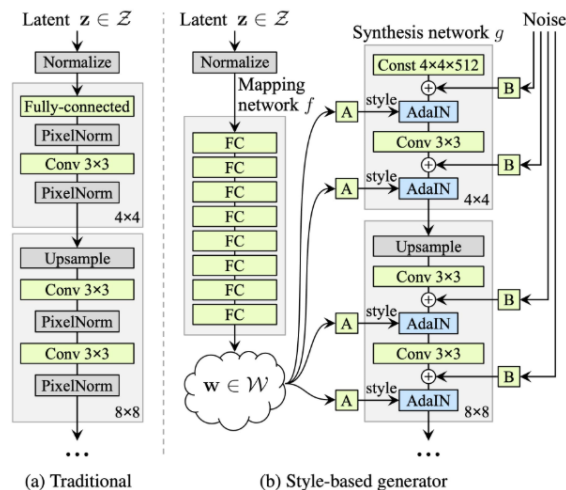
Dalji razvoj GAN modela doveo je do uvođenja **Conditional Generative Adversarial Networks (cGAN)**. Ove mreže predstavljaju proširenje osnovnog GAN modela, u kojem se dodatne informacije u vidu oznaka (*labela*) uvode kao uslov i za generator i za discriminator (*Slika 7*). Ove oznake omogućavaju kontrolisano generisanje podataka, jer usmjeravaju generator ka stvaranju uzoraka sa određenim karakteristikama, dok diskriminator koristi iste informacije kako bi preciznije razlikovao stvarne i generisane podatke. Na taj način, cGAN modeli postižu bržu konvergenciju i efikasnije učenje u poređenju sa neuslovljenim GAN pristupima.



*Slika 7: Conditional GAN*

Posebno značajan pravac razvoja predstavljaju GAN modeli za **transformaciju slike-u-sliku**, među kojima se ističu **pix2pix** i **CycleGAN** pristupi. Pix2pix modeli su namijenjeni situacijama u kojima postoje upareni ulazno-izlazni podaci, te omogućavaju učenje preslikavanja između dva vizuelna domena. Nasuprot tome, CycleGAN uvodi koncept ciklične konzistentnosti, čime omogućava učenje transformacija između domena i u odsustvu uparenih podataka. Ovi modeli su pokazali da se adversarijalno učenje može koristiti ne samo za generisanje novih slika, već i za realističnu vizuelnu transformaciju postojećih.

Posebno napredan pristup u okviru generativnih adversarijalnih mreža predstavlja **StyleGAN arhitektura** (*Slika 8*), koja uvodi koncept razdvajanja stilskih komponenti tokom procesa generisanja slike. Umjesto direktnog mapiranja latentnog vektora na izlaznu sliku, StyleGAN omogućava kontrolu različitih nivoa vizuelnih karakteristika, pri čemu se globalna struktura objekta, srednji elementi i fini detalji modeluju na različitim slojevima mreže. Ovakav pristup omogućava kombinovanje vizuelnih osobina iz različitih uzoraka, što je posebno interesantno za domene poput LEGO figurica, gdje se mogu razdvajati i kombinovati oblik figurice, stil odjeće i kolorni detalji, uz zadržavanje koherentne strukture generisanog objekta.



Slika 8: StyleGAN arhitektura

Pored navedenih modela, u literaturi se pojavljuju i brojne druge varijante GAN-ova koje se fokusiraju na specifične aspekte problema, kao što su stabilnost treniranja, raznovrsnost generisanih uzoraka ili efikasnost učenja. Ipak, DCGAN, cGAN, CycleGAN i StyleGAN predstavljaju ključne tačke razvoja GAN pristupa za slike i često služe kao referentni modeli u savremenim istraživanjima. Razumijevanje njihovih osnovnih principa i razlika pruža solidnu osnovu za izbor odgovarajućeg modela u konkretnim primjenama generisanja slika.

## 5. Skupovi podataka, metrike i ograničenja

Kao i drugi modeli, tako i evaluacija generativnih adversarijalnih mreža u velikoj mjeri zavisi od kvaliteta i karakteristika skupa podataka na kojem se trenira. Kada govorimo o generisanju slika, u literaturi se često koriste standardni skupovi podataka kao što su *MNIST*, *CIFAR-10* i *CelebA*, koji omogućavaju poređenje različitih pristupa u kontrolisanim uslovima. Pored ovih skupova, značajnu ulogu imaju i prilagođeni skupovi podataka, naročito u slučajevima kada se model primjenjuje na specifične vizuelne domene, poput slika LEGO figurica.

Za procjenu kvaliteta generisanih slika koriste se različite metrike, među kojima se najčešće navode **Inception Score (IS)** i **Fréchet Inception Distance (FID)**. Ove metrike nastoje da kvantitativno ocijene sličnost između stvarnih i generisanih podataka, ali se u praksi često kombinuju sa subjektivnom vizuelnom procjenom, budući da numeričke vrijednosti ne moraju uvijek u potpunosti odražavati perceptivni kvalitet slika.

Uprkos stalnom unapređivanju, GAN modeli se i dalje suočavaju sa određenim ograničenjima. Najčešći izazovi uključuju nestabilnost tokom treniranja, pojavu mode collapse fenomena, kao i osjetljivost na izbor arhitekture i hiperparametara. Ova ograničenja ukazuju na potrebu pažljivog dizajna eksperimentalnog okvira i predstavljaju važan motiv za dalje istraživanje i unapređenje generativnih modela.

## 6. Vizija i komentar

Na osnovu pregleda teorijskih osnova i postojeće literature, generativne adversarijalne mreže mogu se posmatrati kao snažan i fleksibilan okvir za generisanje vizuelnih podataka. Dosadašnja istraživanja pokazuju da GAN modeli nisu ograničeni samo na sintezu slika visoke vizuelne kvalitete, već omogućavaju i kontrolisano generisanje, manipulaciju vizuelnim atributima i kombinovanje karakteristika iz različitih uzoraka. Ovakve osobine čine GAN pristup posebno pogodnim za primjenu na domenima koji posjeduju jasnu strukturu, ali i visok stepen varijabilnosti.

Posebno interesantan aspekt problema predstavlja činjenica da se slike LEGO figurica mogu grupisati prema tematskim cjelinama, pri čemu pojedine grupe dijele zajedničke vizuelne karakteristike, poput odjeće, boja i prepoznatljivih detalja. Takva struktura podataka otvara mogućnost istraživanja kontrolisanog generisanja slika, gdje bi model mogao da nauči razlike između različitih vizuelnih domena i generiše uzorke koji odgovaraju zadatoj temi. Na primjer, figurice iz istog tematskog univerzuma mogu posjedovati zajednički vizuelni stil, što omogućava ispitivanje sposobnosti generativnog modela da internalizuje i reprodukuje ove karakteristike.

Ovaj projekat predstavlja priliku za dublje razumijevanje generativnih modela i izazova koji prate njihovu praktičnu primjenu. Pored samog generisanja slika, značajan aspekt rada predstavlja i analiza ograničenja GAN modela, poput stabilnosti treniranja i raznovrsnosti generisanih rezultata.

Stečena teorijska osnova predstavlja temelj za dalji rad na ovom projektu, koji će u narednim kontrolnim tačkama biti usmjeren na implementaciju odabranog modela i analizu dobijenih rezultata. Na taj način, projekat će povezati teorijsko razumijevanje generativnih adversarijalnih mreža sa njihovom praktičnom primjenom u konkretnom vizuelnom domenu.

## 7. Literatura

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville i Y. Bengio, "Generative Adversarial Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [2] A. Radford, L. Metz i S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *International Conference on Learning Representations (ICLR)*, 2016.
- [3] J. Zhu, T. Park, P. Isola i A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *International Conference on Computer Vision (ICCV)*, 2017.
- [4] T. Karras, S. Laine i T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [5] "Overview of GAN Structure," Google Developers, Machine Learning, 2025. [Online]. Available: [https://developers.google.com/machine-learning/gan/gan\\_structure](https://developers.google.com/machine-learning/gan/gan_structure)
- [6] "What is a Conditional Generative Adversarial Network (cGAN)?," DataScientest, 2025. [Online]. Available: <https://datascientest.com/en/what-is-a-conditional-generative-adversarial-network-cgan>
- [7] "Discover NVIDIA's StyleGAN: Revolutionary Image Generation," Viso.ai, Deep Learning Insights, 2025. [Online]. Available: <https://viso.ai/deep-learning/stylegan/>
- [8] "22.2-GAN Theory," University at Buffalo, 2025. [Online]. Available: <https://cedar.buffalo.edu/~srihari/CSE676/22.2-GAN%20Theory.pdf>
- [9] "Pix2pix: Key Model Architecture Decisions," Neptune.ai, 2025. [Online]. Available: <https://neptune.ai/blog/pix2pix-key-model-architecture-decision>