

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

**Generativno modeliranje u
računalnom vidu**

Dominik Barukčić

Voditelj: *prof. dr. sc. Tomislav Hrkać*

Zagreb, svibanj 2025.

SADRŽAJ

1. Uvod	1
2. Osnovni pojmovi	2
2.1. Generativno modeliranje	2
2.2. Generativna suparnička mreža	3
2.3. Varijacijski autoenkoder	4
2.4. Difuzijski modeli	6
3. Primjene u računalnom vidu	7
3.1. Generiranje sintetičkih slika	7
3.2. Povećanje rezolucije	8
3.3. Prijenos stila	9
3.4. Uklanjanje šuma	10
4. Praktični eksperiment	12
4.1. Generiranje sintetičke slike lica korištenjem StyleGAN	12
4.2. Povećanje rezolucije korištenjem Real-ESRGAN	13
4.3. Prijenos stila korištenjem CycleGAN	14
4.4. Uklanjanje šuma korištenjem VAE	15
5. Zaključak	16
6. Sažetak	17
7. Literatura	18

1. Uvod

U ovom radu obrađuje se tema generativnog modeliranja u računalnom vidu. Generativni modeli predstavljaju podskup metoda dubokog učenja koji imaju sposobnost stvaranja novih podataka sličnih onima iz stvarne domene. Takvi modeli postaju sve važniji alat u analizi i obradi slikovnih podataka jer omogućuju stvaranje realističnih slika, njihovu rekonstrukciju ili poboljšanje kvalitete te prijenos stilova i karakteristika između slika. Njihova široka primjenjivost očituje se u raznim područjima, uključujući umjetničku produkciju, medicinsku dijagnostiku, računalnu grafiku, pa čak i sigurnosne sustave. Cilj je seminarског rada prikazati osnovne metode generativnog modeliranja, kao što su generativne suparničke mreže (GAN), varijacijski autoencoderi (VAE) i difuzijski modeli, uz njihove primjene u području računalnog vida. Rad uključuje i praktične eksperimente korištenjem dostupnih modela te analizu rezultata generacije. Naglasak je stavljen na razumijevanje osnovnih principa svakog modela, njihovu analizu u smislu kvalitete rezultata i složenosti treniranja, kao i mogućnosti koje pružaju u stvarnim aplikacijama.

2. Osnovni pojmovi

2.1. Generativno modeliranje

Generativni model je model koji u smislu probabilističkog pristupa opisuje kako stvaramo novi skup podataka. Cilj je generativnog modeliranja izgraditi model koji što vjernije oponaša distribuciju stvarnih podataka. Kada naučimo distribuciju tih podataka, iz nje možemo uzorkovati nove različite primjere slične onima u skupu za treniranje.

Prepostavimo da imamo model koji sadrži slike automobila. Zadatak je dizajnirati model koji generira jednu novu ili skup slika automobila. To su podaci koji dosad nisu postojali, isto kao ni automobil koji bi se nalazio na generiranoj slici. Zadatak modela je naučiti pravila koja opisuju izgled automobila, tako da novi generirani uzorci budu realistični. Ovom formulacijom, opisali smo problem koji se rješava generativnim modeliranjem.

Potrebna nam je velika količina primjera objekta kojeg model treba naučiti generirati, tzv. skup za treniranje (engl. *training set*). Jedan primjer iz skupa nazivamo opažanjem, ono se sastoji od velikog broja značajki. Kod slika to su najčešće vrijednosti piksela. Slike sadrže velik broj piksela, a samo manji broj kombinacija pikselsnih vrijednosti čini semantički smislen sadržaj, zbog čega je problem generiranja novih slika veoma složen.

Generativni model je probabilistički model. Proizvodi izlaz uzorkovan iz naučene distribucije podataka. Radi se o nedeterminističkom pristupu, što znači da ne možemo unaprijed točno znati kakav će uzorak model generirati, za razliku od determinističkih pristupa. Zbog toga model mora sadržavati stohastičku komponentu koja utječe na svaki pojedini izlaz.

Zamislimo da postoji nepoznata vjerojatnosna distribucija koja opisuje kolika je vjerojatnost da se neka slika nalazi u skupu za treniranje. Gradimo model koji što vjernije oponaša tu distribuciju i omogućuje uzorkovanje novih, raznolikih uzoraka koji izgledaju poput onih iz izvorne domene.

Razlikujemo dvije osnovne paradigme: *diskriminativno* i *generativno modeliranje*. Diskriminativno modeliranje procjenjuje uvjetnu vjerojatnost $\mathbb{P}(y | \mathbf{x})$, odnosno vjerojatnost da opažanje \mathbf{x} pripada klasi y . Fokus je na učenju granice između klasa. Generativno modeliranje procjenjuje vjerojatnost $\mathbb{P}(\mathbf{x})$, koja opisuje koliko je vjerojatno opažanje \mathbf{x} . Ako radimo s označenim podacima, moguće je procijeniti uvjetnu vjerojatnost $\mathbb{P}(\mathbf{x} | y)$ koja nam opisuje koliko je vjerojatno da se opažanje \mathbf{x} pojavi unutar klase y .

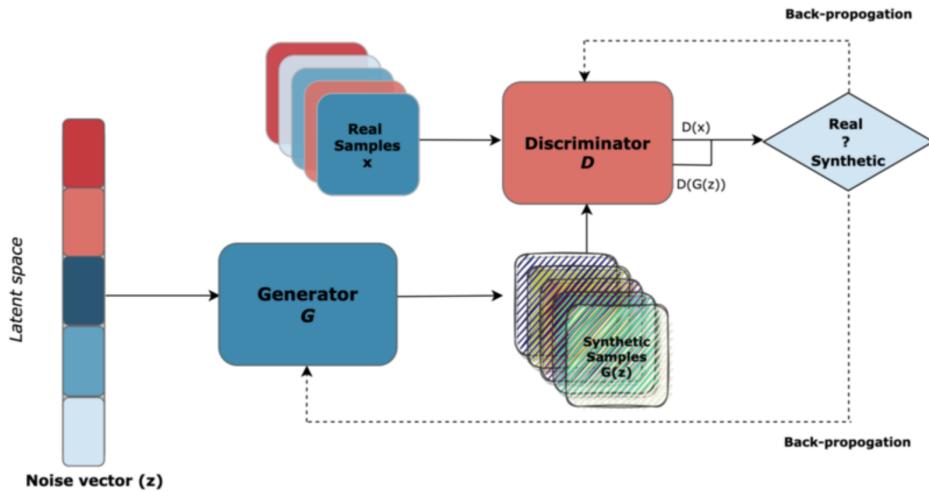
2.2. Generativna suparnička mreža

Generativna suparnička mreža (eng. *Generative Adversarial Network*, GAN) je klasa generativnih modela koju su 2014. godine predstavili Ian Goodfellow i suradnici. GAN sadrži dvije neuronske mreže koje se treniraju u suparničkom odnosu: **generator** G i **diskriminator** D . Cilj generatora je naučiti distribuirati podatke tako da proizvodi sintetičke uzorke $G(z)$ koji nalikuju stvarnim. Zadatak diskriminatora je razlikovati stvarne uzorke \mathbf{x} iz skupa za treniranje od lažnih generiranih uzoraka.

Generator G prima slučajni vektor šuma \mathbf{z} iz latentnog prostora i generira sintetski uzorak $G(\mathbf{z})$. Diskriminator D nastoji maksimizirati točnost klasifikacije između stvarnih i generiranih uzoraka. Obje mreže treniraju se istovremeno, pri čemu generator pokušava zavarati diskriminator, dok diskriminator pokušava pravilno razlikovati ulaze. Proces možemo interpretirati kao minimax igru s funkcijom cilja:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

Idealno, ravnoteža između G i D dovodi do toga da se generirani podaci ne razlikuju od stvarnih, tj. $D(\mathbf{x}) = 0.5$ za sve ulaze. Na taj način model uči implicitnu distribuciju podataka bez potrebe za izričitom specifikacijom funkcije vjerojatnosti.



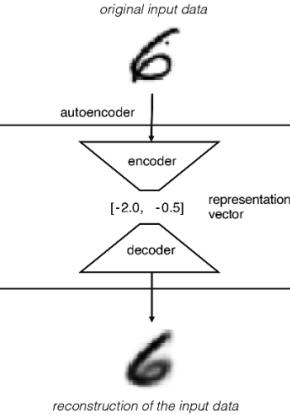
Slika 2.1: Shema rada generativne suparničke mreže (GAN) [6]

Kao što je istaknuto u literaturi [1] i [2], GAN-ovi su iznimno moćni u generiranju realističnih slika, tekstura i glazbe. Važno je spomenuti da treniranje GAN-ova može biti zahtjevno. Razlozi mogu biti nestabilnost, koja se javlja zbog oscilirajuće dinamike između generatora i diskriminatora te problem *mode collapse*, koji se očituje kada generator uči proizvoditi samo ograničen broj varijacija.

Unatoč izazovima, GAN arhitektura temelj je mnogim modelima poput StyleGAN, CycleGAN, BigGAN, itd. Primjenjuju se u zadacima prijenosa stila, povećanja rezolucije, obradi medicinskih snimaka i u mnogim drugim primjenama.

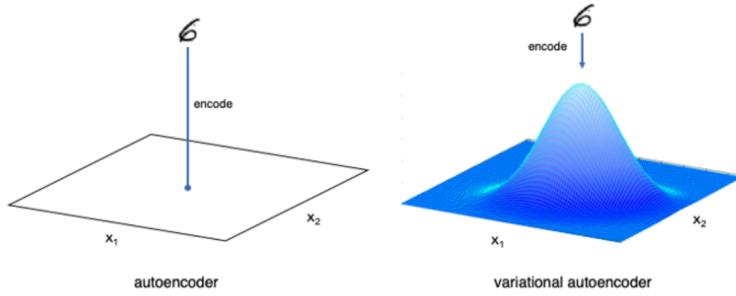
2.3. Varijacijski autoenkoder

Autoenkoder je neuronska mreža koja se sastoji od dva dijela: enkodera i dekodera. Enkoder je dio mreže koji preslikava višedimenzionalni ulaz u reprezentacijski vektor niže dimenzije (latentni prostor). Dekoder je dio mreže koji iz latentnog vektora rekonstruira podatke natrag u izvornu domenu. Ideja je da iz latentnog prostora možemo odabrati bilo koji vektor i rekonstruirati podatak u izvornoj domeni.



Slika 2.2: Dijagram autoenkodera [2]

Varijacijski autoenkoder (eng. Variational Autoencoder, VAE) proširuje klasični autoenkoder tako da latentni prostor promatra probabilistički. Koristi naučeno približno zaključivanje i trenira se metodama temeljenim na gradijentnim postupcima.



Slika 2.3: Razlika između autoenkodera i varijacijskog autoenkodera – deterministički vs. probabilistički prikaz latentnog prostora [2]

Za generiranje uzorka, VAE najprije uzme uzorak \mathbf{z} iz latentne distribucije $p_{\text{model}}(\mathbf{z})$, a trenira se maksimizacijom donje granice dokaza (eng. evidence lower bound, ELBO), definirane izrazom:

$$\mathcal{L}(q) = \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})} [\log p_{\text{model}}(\mathbf{z}, \mathbf{x})] + \mathcal{H}(q(\mathbf{z}|\mathbf{x})) \quad (2.4.1)$$

$$= \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})} [\log p_{\text{model}}(\mathbf{x}|\mathbf{z})] - D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}) \parallel p_{\text{model}}(\mathbf{z})) \quad (2.4.2)$$

$$\leq \log p_{\text{model}}(\mathbf{x}) \quad (2.4.3)$$

Kako bi se omogućilo propagiranje gradijenata kroz stohastički uzorak $\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})$, koristi se **reparametrisacijski trik**. Umjesto direktnog uzorkivanja iz latentne distribucije, latentna varijabla \mathbf{z} se izražava kao funkcija:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$$

gdje su $\boldsymbol{\mu}$ i $\boldsymbol{\sigma}$ parametri distribucije koje daje enkoder, a $\boldsymbol{\epsilon}$ je nezavisni slučajni šum. Postupak omogućuje izračun gradijenata kroz funkciju uzorkovanja, čime je omogućeno učenje parametara gradijentnim spustom.

Funkcija gubitka sastoji se od dvije komponente:

- **Rekonstrukcijski gubitak** koji mjeri koliko dobro dekoder rekonstruira ulaz \mathbf{x} iz latentne reprezentacije \mathbf{z} ,
- **KL divergencija** predstavlja kaznu koja mjeri razliku između približne posteriorne distribucije $q(\mathbf{z}|\mathbf{x})$ i prethodno definirane priori distribucije $p(\mathbf{z})$, za koju obično uzimamo standardnu normalnu distribuciju.

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{x}|\mathbf{z})] - D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}))$$

Rekonstrukcijski gubitak potiče model da uči preciznu rekonstrukciju ulaza, a KL divergencija osigurava da je latentni prostor strukturiran na način pogodan za generiranje novih uzoraka. Time je omogućena **generativna sposobnost** varijacijskog autoenkodera. Model zatim generira nove uzorke jednostavnim uzorkovanjem iz distribucije $p(\mathbf{z})$ i dekodiranjem kroz dekoder.

Varijacijski autoenkoderi nalaze primjenu u zadacima poput generiranja novih slika i uzoraka, interpolacije između podataka u latentnom prostoru, uklanjanja šuma i rekonstrukcije slike te detekcije anomalija.

2.4. Difuzijski modeli

Difuzijski modeli (eng. Diffusion models) su novija klasa generativnih modela koji su iznimno učinkoviti u generiranju visoko kvalitetnih i realističnih slika. Temelje se na postupku učenja generativnog procesa kroz dvije faze:

- **Difuzija** - postupno dodavanje šuma na podatke
- **Denoising** - postupna rekonstrukcija podataka iz šuma

Treniranjem model uči kako ukloniti šum iz slike, a u fazi generacije počinje od nasumičnog šuma i iterativno stvara sliku.

Za razliku od GAN-ova, koji imaju probleme poput nestabilnosti u treniranju i mode collapsea, difuzijski modeli imaju veću stabilnost i visoku razlučivost generiranih uzoraka. Nedostatak im je sporo generiranje zbog velikog broja iteracija potrebnih za postupno uklanjanje šuma.

3. Primjene u računalnom vidu

Zahvaljujući sposobnosti stvaranja realističnih i korisnih vizualnih podataka, generativni modeli imaju široku primjenu u računalnom vidu. U nastavku su opisane sljedeće primjene.

3.1. Generiranje sintetičkih slika

Strmoglavim razvojem područja pojavljuju se zahtjevi za velikom količinom kvalitetnih podataka. Model je onoliko dobar koliko su i podaci koje koristimo za treniranje i validaciju. Prikupljanje takvih podataka u praksi ispada vremenski zahtjevno, a time i skupo. Ako tome dodamo zakonska ograničenja, dolazimo do još većih poteškoća (npr. GDPR, etika u medicini ili nadzornoj sigurnosti). Jedno od rješenja jest generiranje sintetičkih podataka. U tu svrhu možemo koristiti bilo koji generativni model ili kombinaciju s alatima jezgara igre (eng. game engine) ili 3D modeliranja. [4]

Sintetičke slike možemo klasificirati u dvije glavne kategorije:

- **Složene (kompozitne) slike** - kombinacija stvarnih pozadina s različitim objektima
- **Potpuno virtualne slike** - generirane u potpunosti

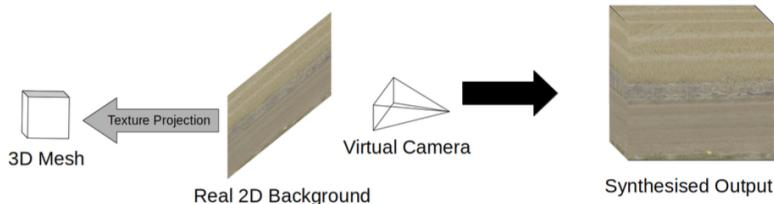
Sintetički podaci mogu dopuniti ili proširiti stvarne skupove. Također, možemo im smanjiti pristrandost ili izbaciti neke skupe realne podatke. GAN modeli su kvalitetni u generiranju slika lica, a velika je korist što ne narušavamo tuđu privatnost ili sigurnost. Postoje drugi modeli koji služe za simuliranje 3D okruženja kojima možemo testirati teške uvjete bez stvarnog rizika. Mogućnosti su brojne, ali ne dolaze bez plaćanja cijene. [4]

Prednosti:

- automatizirano označavanje podataka
- skalabilnost i kontrola
- očuvanje privatnosti i sigurnosti

Nedostaci:

- često sintetički podaci imaju lošije rezultate u odnosu na realne
- zahtjevni računalni resursi
- nema univerzalne metodologije za ocjenu kvalitete podataka



Slika 3.1: Sinteza projekcijom stvarne 2D slike na 3D mrežu [4]

3.2. Povećanje rezolucije

Povećanje rezolucije slike (eng. Super-resolution) problem je u računalnom vidu kojim nastojimo iz ulazne slike niske razlučivosti generirati sliku visoke razlučivosti. Dostupne su klasična metode interpolacije, ali ne pružaju nikakve nove informacije. Metode generativnog modeliranja, posebice generativne suparničke mreže, omogućuju učenje distribucija visokih razlučivosti u cilju nadopune izgubljenih detalja.

GAN pristupi za povećanje rezolucije sastoje se od 2 modela:

- **Generatorka** koji generira visoko-razlučivu sliku iz nisko-razlučivog ulaza
- **Diskriminatorka** koji razlikuje generiranu sliku od stvarne visoko-razlučive slike

Diskriminatorka kao suparnički proces generatoru, omogućuje da s vremenom poboljša kvalitetu generiranih slika kako bi zavarale diskriminatorku.

Metode možemo klasificirati prema vrsti učenja:

- **Polunadzirano učenje** - kombinira ograničene označene podatke s velikim brojem neoznačenih podataka
- **Nenadzirano učenje** - koristi samo neoznačene podatke, često uz dodatne tehnike poput cikličke konzistencije

Neki istaknuti GAN modeli za super-rezoluciju:

- **SRGAN** - prvi GAN model za super-rezoluciju koji koristi perceptualnu funkciju gubitka za generiranje vizualno uvjerljivih slika

- **ESRGAN** - poboljšanje SRGAN-a koje uvodi Residual-in-Residual Dense Block (RRDB) za bolju rekonstrukciju detalja i koristi relativistički pristup u diskriminatoru
- **Real-ESRGAN** - proširenje ESRGAN-a koje se fokusira na stvarne degradacije slika, omogućujući bolje rezultate na slikama iz stvarnog svijeta

Prilikom treniranja, dostupne su različite funkcije gubitaka:

- **Adversarialni gubitak** - potiče generator da stvara slike koje su teško razlikovljive od stvarnih
- **Perceptualni gubitak** - koristi značajke iz unaprijed treniranih mreža (npr. VGG) za očuvanje vizualnih karakteristika
- **Funkcija gubitka sadržaja** - mjeri razliku između generirane i stvarne slike na pikselnoj razini

Evaluacija performansi modela provodi se pomoću metrika kao što su PSNR (Peak Signal-to-Noise Ratio) i SSIM (Structural Similarity Index), iako ove metrike ne koreliraju uvijek s percepcijском kvalitetom slike.

3.3. Prijenos stila

Prijenos stila (eng. Style transfer) tehnika je kojom se sadržaj slike (npr. boje, teksture, debljina kista) prenosi na drugu sliku, zadržavajući stil originalne slike. Generativni modeli, specifično GAN-ovi, iznimno su učinkoviti u ovoj domeni. Osnovni cilj je generirati sliku koja zadržava semantički sadržaj originalne slike vizualno oblikovan prema ciljnog stilu. Primjene su u umjetnosti, produkciji, animaciji, društvenim mrežama i medicinskoj vizualizaciji. Tri istaknuta pristupa prijenosu stila temeljena na GAN arhitekturi - CycleGAN, StyleGAN i TL-GAN - analizirana su u radu Bo i sur. [7]

Svaki od ovih modela koristi različite arhitekture i strategije za prijenos stila:

- **CycleGAN** omogućuje stilizaciju između dviju domena bez uparenih primjera koristeći cikličku konzistenciju.
- **StyleGAN** uvodi mapirajuću mrežu i višestupanjsko umetanje latentnog vektora, tako postiže bolju kontrolu nad generiranim stilovima.
- **TL-GAN** omogućuje preciznu kontrolu nad atributima slike kombiniranjem pretreniranih klasifikatora s GAN-ovima, bez potrebe za ponovnim treniranjem generatora.

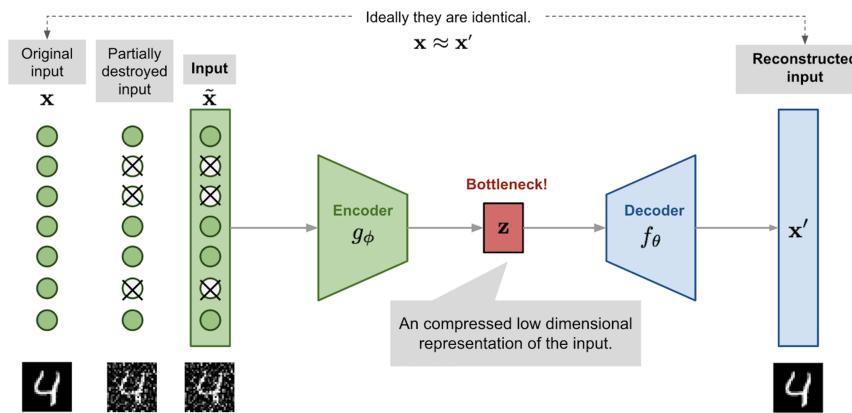
Idealni model za prijenos stila treba omogućiti:

- Preciznu kontrolu stilskih atributa
- Efikasnost treniranja i izvedbe
- Skalabilnost na slike visoke rezolucije
- Mogućnost integracije s detektorima objekata radi selektivne primjene stila
- Naknadnu optimizaciju rezultata radi korekcije neželjenih deformacija

3.4. Uklanjanje šuma

Uklanjanje šuma sa slika (eng. denoising) jedan je od klasičnih zadataka u računalnom vidu gdje model mora naučiti rekonstruirati izvornu sliku iz njezine oštećene verzije. Denoising autoencoder (DAE) predstavlja robusnu arhitekturu neuronske mreže dizajniranu kako bi se spriječila prenaučenost (eng. overfitting) i potaknuto učenje korisnih značajki za dobru generalizaciju modela.

Denoising autoencoder, predstavili su Vincent i suradnici 2008. godine. Za razliku od klasičnog autoenkodera koji pokušava naučiti funkciju identiteta, DAE uzima šumom kontaminiran ulaz i trenira mrežu tako da rekonstruira prvobitni ulaz. [9]



Slika 3.2: Arhitektura autoenkodera za uklanjanje šuma [9]

Prilikom treniranja, ulazne slike se kontaminiraju različitim vrstama šuma (npr. Gaussov, "salt and pepper", zamračenje piksela). Model tada uči rekonstrukciju izvorne, čiste slike. Cilj je da mreža ne pamti šablonе, nego da nauči relevantne apstraktnе značajke koje pomažu u oporavku slike.

Takav pristup ima dvostruku korist:

- **Robusnost** - otpornost na male varijacije i šum u ulazu

- **Generalizacija** - lakše učenje strukturalnih odnosa između piksela i ne oslanjanje na pojedinačne vrijednosti

Izvorni eksperiment uključivao je maskiranje piksela (nasumično postavljanje dijela ulaznih vrijednosti na nulu). Primijetimo kako ovo podsjeća na dropout, iako je DAE predstavljen 4 godine prije dropout tehnike. [9]

Kod visokodimenzionalnih podataka (npr. slika), ova metoda omogućuje modelu da:

- nauči ovisnosti između piksela
- raspozna semantičke informacije iz nepotpunih ulaza
- poboljša stabilnost i robusnost latentnog prostora

Kombinacija DAE sa varijacijskim autoenkoderom (VAE) proširuje mogućnosti. Model osim uklanjanja šuma, koristi probabilistički latentni prostor kako bi omogućio generativne mogućnosti.

4. Praktični eksperiment

4.1. Generiranje sintetičke slike lica korištenjem StyleGAN

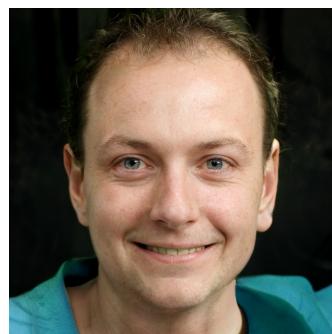
Za generiranje realističnih slika lica korišten je model StyleGAN3-R, kojeg je razvila NVIDIA. Treniran je na skupu FFHQ (Flickr-Faces-HQ) koji sadrži slike visoke kvalitete ljudskih lica u rezoluciji od 1024×1024 piksela. StyleGAN3 predstavlja značajno poboljšanje u odnosu na prethodne verzije (StyleGAN i StyleGAN2) jer rješava problem aliasinga i omogućuje glatke interpolacije u latentnom prostoru.

Model je preuzet sa službenog NVIDIA repozitorija te lokalno pokrenut korišteњem Pythona i biblioteke PyTorch. Generiranje se temelji na uzorkovanju latentnog vektora z iz višedimenzionalnog Gaussovog prostora, koji se potom prosljeđuje kroz generativnu mrežu (inferencija). Rezultat je slika visoke rezolucije koja prikazuje lice osobe koje zapravo ne postoji.

Proces generiranja prikazan je u nastavku:

- Učitavanje pretreniranog modela `stylegan3-r-ffhq-1024x1024.pkl`
- Generiranje latentnog vektora slučajnim uzorkovanjem iz $N(0, I)$
- Prolazak latentnog vektora kroz mrežu i dobivanje slike

Dobivena slika prikazana je na slici 4.1.



Slika 4.1: Generirana slika lica

4.2. Povećanje rezolucije korištenjem Real-ESRGAN

Za eksperiment povećanja rezolucije korišten je pretrenirani model Real-ESRGAN (Real-Enhanced Super-Resolution Generative Adversarial Network). To je unaprijeđena verzija ESRGAN-a dizajnirana za bolje generaliziranje nad realističnim slikama. Model koristi RRDB (Residual-in-Residual Dense Block) arhitekturu i sposoban je rekreirati izgubljene detalje i oštrinu u slikama niske kvalitete.

Korišten je model `Realesrgan_x4plus`, preuzet iz službenog repozitorija `xinntao/Real-ESRGAN` i pokrenut lokalno na CPU-u. Budući da FP16 preciznost nije podržana na CPU-u, korišten je parametar `-fp32`, uz `-tile 128` za optimizaciju memorijske potrošnje. [10]

Eksperiment je proveden na slici u niskoj rezoluciji prikazujući sliku grada Zagreba iz davnina. Cilj je bio povećati razlučivost i istovremeno rekonstruirati oštrinu i detalje koji su izgubljeni zbog degradacije.

Rezultati eksperimenta prikazani su na slici 4.2.



Izvorna slika niske razlučivosti



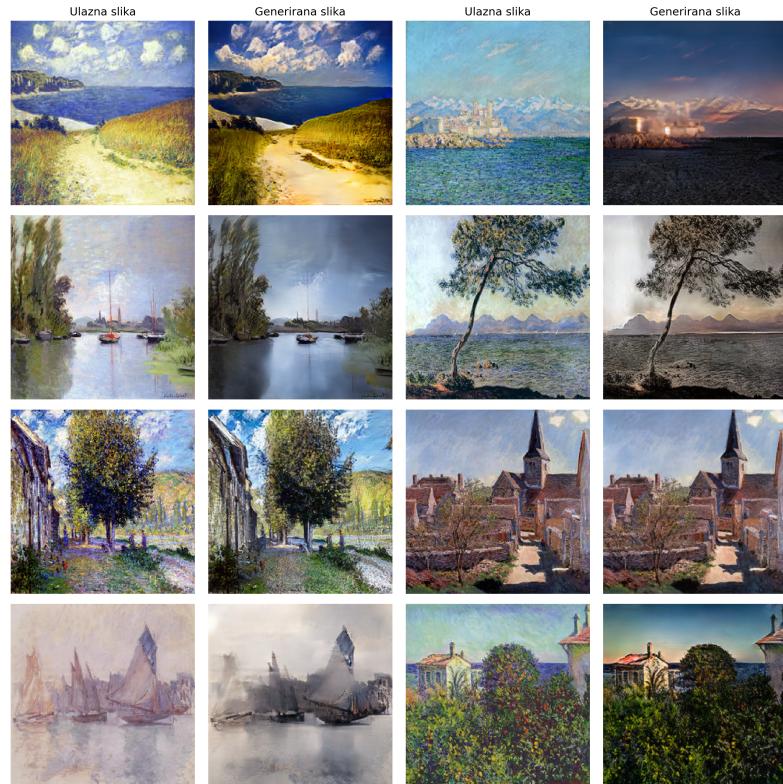
Generirana slika visoke razlučivosti

Slika 4.2: Rezultati povećanja rezolucije korištenjem Real-ESRGAN modela

Model je uspješno poboljšao vizualnu jasnoću slike. Povećao je oštrinu na prometnim znakovima, rubovima zgrada te linijama ceste. Model je donekle rekonstruirao izgubljene teksture i eliminirao dio kompresijskih artefakata, iako se mogu primjetiti manja iskrivljenja u područjima gdje nedostaje dovoljno informacija u izvornim pikselima. Ovo je veoma korisno za obnovu starih fotografija ili unapređenje ulaznih podataka za računalni vid (npr. detekcija objekata).

4.3. Prijenos stila korištenjem CycleGAN

Za potrebe eksperimenta korišten je pretrenirani CycleGAN model za prijenos stila iz umjetničkog stila Claudea Moneta u domenu realističnih fotografija pejzaža (model monet2photo). [8]



Slika 4.3: Prijenos stila: Monetova slika (lijevo) i generirani realistični pejzaž (desno)

Rezultati generiranja prikazani na slici pokazuju različite stupnjeve uspješnosti prijenosa stila iz Monetovog umjetničkog izraza u fotorealistični prikaz. U prva dva retka možemo primijetiti uspješnu transformaciju gdje su generirane slike zadržale semantičku strukturu ulazne slike uz realistični efekt prijenosa stila. Kod određenih primjera razlike između ulazne i generirane slike su minimalne ili teško uočljive. U takvim slučajevima model očito nije dovoljno dobro "prepoznao" koje elemente treba transformirati, što može biti posljedica:

- nedostatne razlike između stilova iz skupa za treniranje
- ograničenja modela u generalizaciji na određene kompozicije
- nedostatka informacija o dubini ili strukturi scene u slici

Primjećujemo da model bolje funkcioniра na slikama s jasnim kompozicijama, kontrastima i osvjetljenjem, dok slike s difuznijim konturama i teksturama zadržavaju neke Monetove karakteristike.

vaju više karakteristika Monetovog stila iako bi trebale biti realističnije. Zaključujemo da model pokazuje sposobnost prijenosa stila, ali rezultati nisu u svim slučajevima konzistentni te ovise o karakteristikama ulazne slike.

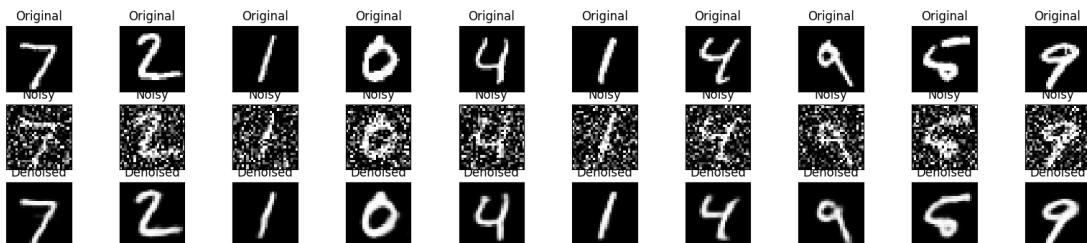
4.4. Uklanjanje šuma korištenjem VAE

U ovom eksperimentu korišten je varijacijski autoenkoder (VAE) za uklanjanje šuma s rukom pisanih znamenki iz skupa MNIST. Model je treniran tako da iz slike s dodatnim Gaussovim šumom nauči rekonstruirati originalnu i čistu inačicu slike. VAE koristimo kao denoising autoencoder koji je specijalizirana arhitektura za obradu degradiranih slika. [9]

Podaci su prethodno normalizirani i prošireni šumom s Gaussovom distribucijom, nakon čega treniramo model kojemu su ulazne slike one s dodanim šumom, a ciljne slike su originali. Model koristi konvolucijske slojeve u enkoderu i dekoderu te latentni prostor dimenzije 128.

Rezultati su prikazani na slici 4.4 i sastoje se od tri reda:

- Original - originalne slike iz skupa MNIST
- Noisy - slike s dodatnim šumom
- Denoised - slike rekonstruirane pomoću VAE-a



Slika 4.4: Uklanjanje šuma s MNIST slika pomoću varijacijskog autoenkodera

Model uspješno uklanja većinu šuma i vraća prepoznatljivu strukturu znamenaka, čime vidimo učinkovitost VAE arhitekture u zadatku uklanjanja šuma. Unatoč jednostavnosti modela, postignuti rezultati su zadovoljavajući, uzimajući u obzir nisku razlučivost ulaznih slika (28×28 piksela).

5. Zaključak

U cilju demistifikacije generativnog modeliranja, prikazane su tri ključne arhitekture generativnih modela: generativne suparničke mreže (GAN), varijacijski autoenkoder (VAE) i difuzijski modeli. Difuzijski modeli nisu detaljno obrađeni radi njihove složenosti. Cilj rada postignut je arhitekturama GAN i VAE. Pojašnjeno je kako omogućiti stvaranje novih podataka iz naučene distribucije i kako naučiti tu distribuciju. Analizirane su temeljne ideje, osnovna matematička podloga i prednosti u praktičnim primjenama računalnog vida kroz 4 eksperimenta. Eksperimenti prikazuju generiranje sintetičkih slika, povećanje rezolucije, prijenos stila i uklanjanje šuma. Eksperimentalni dio potvrdio je učinkovitost implementiranih arhitektura. Prikazano je kako se pretrenirani modeli mogu iskoristiti za konkretne zadatke bez potrebe za vlastitim treniranjem na velikim skupovima podataka. Iako je postignut veliki napredak, generativni modeli i dalje se suočavaju s izazovima poput stabilnosti treniranja (posebice kod GAN-ova), računalnih zahtjeva (posebno kod difuzijskih modela) i pouzdanosti generiranih podataka. U budućnosti se očekuje daljnje poboljšanje kontrole nad generiranim uzorcima, hibridizacija modela (npr. VAE-GAN) te šira primjena u domenama poput medicine, industrijske vizualne inspekcije i kreativnih industrija.

6. Sažetak

Generativno modeliranje kao suvremenii pristup u računalnom vidu omogućuje stvaranje novih slika i poboljšanje kvalitete postojećih. Obrađujemo osnovne metode generativnog modeliranja poput GAN-ova, VAE-ova i difuzijskih modela. Prikazat ćemo njihove glavne značajke, usporediti ih te prikazati primjenu u generiranju sintetičkih slika, povećanju rezolucije i prijenosu stila. Uključeno je i nekoliko praktičnih primjera dobivenih iz dostupnih pretreniranih modela.

7. Literatura

- [1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.
- [2] David Foster. *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play*. O'Reilly Media, Sebastopol, CA, 2019.
- [3] Jakub Langr and Vladimir Bok. *GANs in Action: Deep Learning with Generative Adversarial Networks*. Manning Publications, Shelter Island, NY, 2019.
- [4] Keith Man and Javaan Chahl. A review of synthetic image data and its use in computer vision. *Journal of Imaging*, 8(11), 2022.
- [5] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Generative adversarial networks for image super-resolution: A survey. *arXiv preprint arXiv:2204.13620*, 2022.
- [6] Ghadeer Ghosheh, Li Jin, and Tingting Zhu. A review of generative adversarial networks for electronic health records: applications, evaluation measures and data sources. 03 2022.
- [7] Xihao Bo, Xiaoyang Jing, and Xiaojian Yang. Style transfer analysis based on generative adversarial networks. pages 27–30, 2021.
- [8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017.
- [9] Lilian Weng. From autoencoder to beta-vae. *lilianweng.github.io*, 2018.
- [10] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*.