

SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

*Luka Mucko*

**Generiranje radarskih slika  
difuzijskim modelima**

Zagreb, 2024.

Ovaj rad izrađen je na Zavodu za komunikacijske i svemirske tehnologije pod vodstvom izv. prof. dr. sc. Daria Bojanjaca i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2023./2024.

# Sadržaj

<b>Popis slika</b>	<b>iii</b>
<b>1 Uvod</b>	<b>1</b>
<b>2 Radarske slike</b>	<b>2</b>
<b>3 Difuzijski modeli</b>	<b>4</b>
<b>4 Implementacija</b>	<b>9</b>
4.1 Priprema podataka . . . . .	9
4.2 HuggingFace diffusers . . . . .	9
4.3 Bezuvjetni model . . . . .	10
4.4 Uvjetni model . . . . .	11
<b>5 Treniranje modela</b>	<b>13</b>
5.1 Bezuvjetni model . . . . .	13
5.2 Uvjetni model . . . . .	14
<b>6 Rezultati</b>	<b>15</b>
6.1 Generiranje bezuvjetnim modelom . . . . .	15
6.2 Generiranje uvjetnim modelom . . . . .	16
<b>7 Zaključak</b>	<b>19</b>
<b>8 Literatura</b>	<b>20</b>

## Popis slika

1	GBSAR-Pi i objekti sa scene . . . . .	2
2	Radarska snimka i rekonstrukcija . . . . .	3
3	Pregled generativnih modela . . . . .	5
4	Grafički model difuzijskog proces . . . . .	5
5	Stopa učenja za bezuvjetni model. . . . .	13
6	Gubitak bezuvjetnog modela na logaritamskoj skali. . . . .	14
7	Gubitak uvjetnog modela na logaritamskoj skali. . . . .	15
8	Primjeri radarskih snimki generiranih bezuvjetnim modelom. . . . .	16
9	Radarske snimke bezuvjetnog modela i njihove rekonstrukcije. . . . .	17
10	Primjeri radarskih snimki generiranih uvjetnim modelom . . . . .	17
11	Rekonstrukcija prvih 4 generiranih slika . . . . .	18

# 1 Uvod

Radarski sistem sa sintetičkom aperturom (eng. *Synthetic Aperture Radar* ili SAR) vrsta je radara kod kojega se senzor kreće za vrijeme snimanja radarske slike. Na taj način umjetno se povećava efektivna površina antene što direktno utječe na povećanje rezolucije radara. Radari sa sintetičkom aperturom koriste se u satelitima ili zrakoplovima za snimanje Zemlje iz svemira ili zraka, ali se koriste i za monitoriranje proizvodnih ili građevinskih procesa u GB-SAR izvedbi (eng. *Ground Based Synthetic Aperture Radar*). Na Zavodu za komunikacijske i svemirske tehnologije razvijen je jedan takav GB-SAR sustav, GBSAR-Pi [1], koji se koristi za monitoriranje procesa [2, 3]. Do sada su razvijeni modeli dubokog učenja za prepoznavanje objekata u radarskim slikama [4, 5, 6, 7], za prepoznavanje objekta s greškom i slično. Za razvoj svakog od tih modela dubokog učenja potrebno je kreirati bazu slika snimljenih radarom i do sada su kreirane 4 takve baze [8, 9, 10, 11]. No, za svaku pojedinu snimku u bazi potrebno je oko 10 minuta što predstavlja ograničavajući faktor na brzinu kreiranja i različitost kreiranih baza. Zato je potrebno kreirati generativni model kojim će se na temelju prethodnih snimki generirati nova snimka. Od postojećih pristupa generativnim modelima u ovom radu odabrani su difuzijski modeli zbog njihove sposobnosti stvaranja visokokvalitetnih slika.

U ovom radu kreirana su dva generativna modela za generiranje radarskih snimki, uvjetni i bezuvjetni model. Uvjetni model kreira radarske slike iz tekstualnog opisa scene dok bezuvjetni model generira radarske slike bez dodatnih ulaznih informacija. Kvalitetu generiranih slika provjeravamo kvalitetom njihovih rekonstrukcija. Oba modela generiraju radarske slike čije su rekonstruirane slike vrlo slične stvarnim slikama što predstavlja odličan prvi korak u treniranju dubokih modela na temelju generiranih radarskih slika.

## 2 Radarske slike

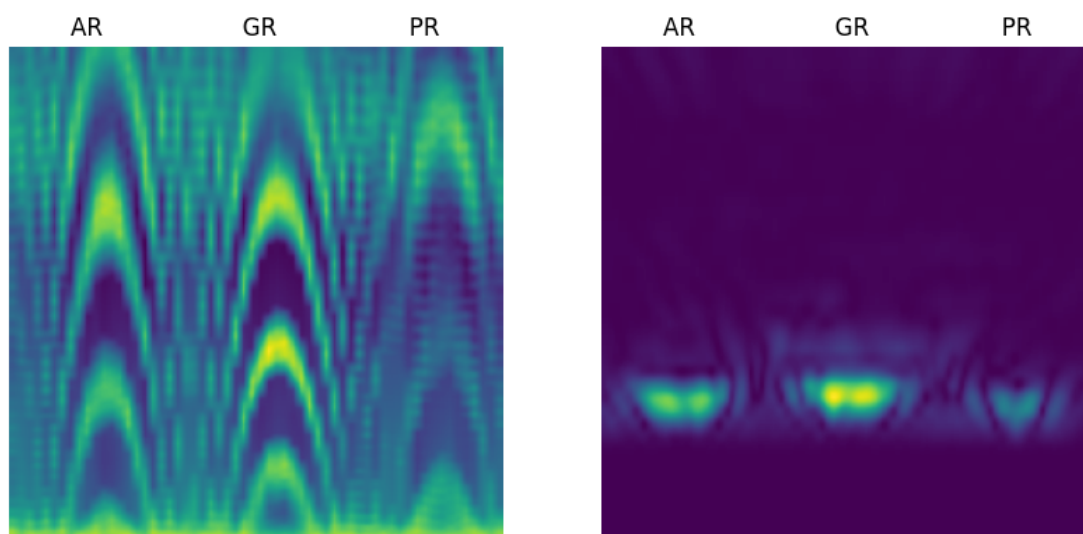
SAR slike koje generiramo snimljene su uz pomoć sustava GBSAR-Pi. Sustav radi u stop-and-go modu što znači da se senzor za snimanje postavi u poziciju za snimanje, pošalje i primi signal te se nakon toga pomakne u sljedeći položaj za snimanje. Senzor se kreće po liniji duljine 60 cm u koracima od po 1 cm. Za snimanje se koristi FMCW radar (eng. *Frequency Modulated Continuous Wave* koji radi na 24 GHz. Na Slici 1 prikazan je GBSAR-Pi sustav zajedno sa standardnom scenom za snimanje. Scena se sastoji od 3 jednako razmaknuta objekta 30cm od radara. Objekti koji se nalaze na sceni mogu biti PR - plastični cilindar, PS - plastični kvadar, GR - stakleni cilindar, GS - stakleni kvadar, AR - aluminijski cilindar i AS - aluminijski kvadar.



Slika 1: GBSAR-Pi i objekti sa scene

Dimenzija pojedine radarske slike je 1024x60 gdje 1024 predstavlja broj frekvencijskih točaka koje karakteriziraju signal dok je 60 je broj koraka senzora za vrijeme snimanja radarske slike. Sirova radarska slika snimljena GBSAR-Pi

sustavom prikazana je na Slici 2. Tako snimljenu sliku moguće je rekonstruirati nekim od algoritama za rekonstrukciju slike snimljene SAR-om od kojih je najpopularniji omega-k algoritam [11]. Na rekonstruiranoj slici nisu prikazani objekti nego radarski otisci objekata što znači da su objekti koji jače raspršuju elektromagnetske valove prikazani većim intenzitetom na rekonstruiranoj slici od onih objekata koji slabije raspršuju elektromagnetske valove koje generira radar. Na Slici 2 se nalazi radarska snimka i njena rekonstrukcija.



Slika 2: Radarska snimka i rekonstrukcija

Na radarskoj snimci sa Slike 2 prikazana su tri objekta. S lijeva na desno to su aluminijski, stakleni i plastični cilindar, isto kao i na Slici 1. Radarski otisci tih objekata u skladu su s materijalom od kojih su ti objekti napravljeni.

### 3 Difuzijski modeli

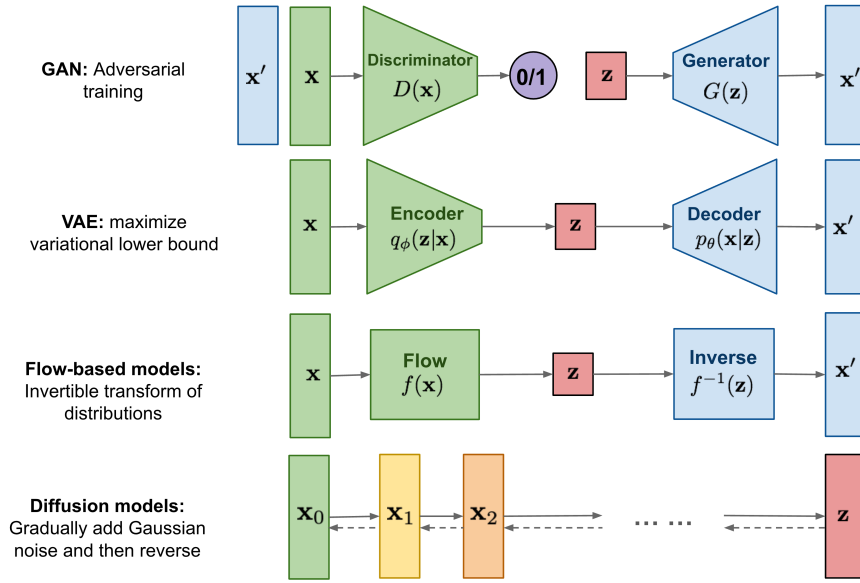
Difuzijski modeli su danas najšire korišteni generativni modeli zbog kvalitete generiranih slika. Ostali popularni generativni modeli koji se danas koriste su GAN-ovi, varijacijski autoenkoderi (VAE) i normalizacijski tokovi. Ali svi imaju svoje specifične probleme: GAN-ovi su potencijalno nestabilni tijekom treniranja i imaju slabu raznolikost generiranih slika. Varijacijski autoenkoderi maksimiziraju donju granicu log izglednosti što ne mora implicirati maksimizaciju izglednosti, a normalizacijski tokovi su ograničeni na invertibilne transformacije u svojoj arhitekturi kao što se vidi na Slici 3 preuzeto iz [12]. Difuzijski modeli imaju jednostavan proces treniranja, dodaj gaussov šum slici u ukupno  $T$  koraka i istreniraj neuronsku mrežu da za danu zamućenu sliku u trenutku  $t$  predvidi šum dodan između trenutka  $t - 1$  i  $t$ , jedini cilj takve neuronske mreže je minimizirati  $L^2$  grešku između predviđenih šumova. Ova jednostavnost u treniranju čini difuzijske modele daleko boljima od ostalih generativnih modela. Također, neuronske mreže koje predviđaju šum možemo uvjetovati dodatnim informacijama kao što je tekst [12]. Najveća mana difuzijskih modela jest što su spori, generiranje zahtjeva  $T$  prolaza kroz neuronsku mrežu, i što zamućena slika  $z_t$  nema korisnu interpretaciju kao što ima latentni prostor kod varijacijskih autoenkodera. [13].

U ovom radu koristimo kanonski difuzijski model DDPM (Denoising Diffusion Probabilistic Models) [14] za bezuvjetno generiranje GB-SAR slika i generiranje GB-SAR slika bazirano na tekstu. Difuzijski model mapira ulaz  $x$  u latentnu varijablu  $z_t$  dodavanjem gaussovog šuma s očekivanjem 0 i varijancom  $\beta_t$ ,

$$z_1 = \sqrt{1 - \beta_1}x + \sqrt{\beta_1}\epsilon_1, \quad (1)$$

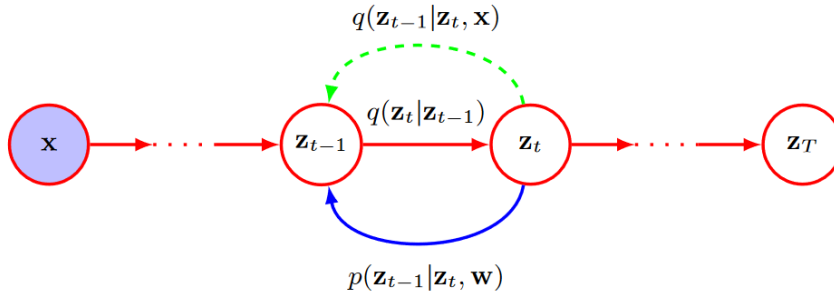
$$z_t = \sqrt{1 - \beta_t}z_{t-1} + \sqrt{\beta_t}\epsilon_t, \quad (2)$$





Slika 3: Pregled generativnih modela

gdje su  $\epsilon_t \sim \mathcal{N}(0, I)$   $\beta_t < 1$ . Odabir  $\beta_t$  i koeficijenata  $\sqrt{1 - \beta_t}$  i  $\sqrt{\beta_t}$  je takav da je srednja vrijednost  $z_t$  bliža nuli nego  $z_{t-1}$  i varijanca bliža jediničnoj matrici nego  $z_{t-1}$ . Dakle kada  $t \rightarrow \infty$  vrijedi da  $z_t \sim \mathcal{N}(0, I)$



Slika 4: Grafički model difuzijskog proces

Budući da je  $q(z_1, \dots, z_t|x) = q(z_1|x) \prod_{\tau=2}^t q(z_\tau|z_{\tau-1})$  ako marginaliziramo preko varijabli  $z_1, \dots, z_{t-1}$  dobivamo analitičku formu za  $q(z_t|x)$ ,

$$q(z_t|x) = \mathcal{N}(\sqrt{\alpha_t}x, (1 - \alpha_t)I), \quad (3)$$

gdje je  $\alpha_t = \prod_{\tau=1}^t (1 - \beta_\tau)$  ili ekvivalentno

$$z_t = \sqrt{\alpha_t}x + \sqrt{1 - \alpha_t}\epsilon_t. \quad (4)$$

Cilj difuzijskog modela je naučiti inverzni proces  $q(z_{t-1}|z_t, x)$ . Korištenjem Bayesovog teorema imamo

$$q(z_t|z_{t-1}, x) = \frac{q(z_t|z_{t-1}, x)q(z_{t-1}|x)}{q(z_t|x)}. \quad (5)$$

Iz (5) se može pokazati da se  $q(z_{t-1}|z_t, x)$  ravna po normalnoj razdiobi s očekivanjem  $m_t(x, z_t)$  i varijancom  $\sigma_t^2$ ,

$$m_t(x, z_t) = \frac{(1 - \alpha_{t-1})\sqrt{1 - \beta_t}z_t + \sqrt{\alpha_{t-1}}\beta_t x}{1 - \alpha_t}, \quad (6)$$

$$\sigma_t^2 = \frac{\beta_t(1 - \alpha_{t-1})}{1 - \alpha_t}. \quad (7)$$

Uvrstimo li izraz za latentnu varijablu  $z_t$  (4) u (6) dobijemo očekivanje

$$m_t(x, z_t) = \frac{1}{\sqrt{\alpha_t}}\left(z_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}}\epsilon_t\right). \quad (8)$$

Budući da ne možemo derivirati slučajno generirane brojeve da bi smo odstranili šum moramo koristiti Langevinovu dinamiku stohastičkog gradijenta (stochastic gradient langeving dynamics) [15]. Langevinova dinamika stohastičkog gradijenta uzorkuje iz distribucije  $p(x)$  znajući samo  $\nabla(\log p(x))$

$$x_t = x_{t-1} + \frac{\delta}{2}\nabla_x \log p(x_{t-1}) + \sqrt{\delta}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{I}), \quad \delta > 0, \quad (9)$$

gdje dodajemo šum  $\delta\epsilon_t$  tako da uzorkovanje ne stane u lokalnom minimumu.

Tablica 1: Algoritmi treniranja i uzorkovanja, preuzeto iz [14]

---

**Algorithm 1** DDPM Treniranje

---

- 1: **Ulaz:** Distribucija podataka  $q(x_0)$ ,  $\epsilon_\theta(x_t, t)$  neuronska mreža
  - 2: **for**  $t = 1$  **do**  $T$  **do**
  - 3:     Uzorkuj  $x_0 \sim q(x_0)$
  - 4:     Uzorkuj  $\epsilon \sim \mathcal{N}(0, I)$
  - 5:     Uzorkuj  $t \sim \text{U}(\{1, \dots, T\})$
  - 6:     Izračunaj  $x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon$
  - 7:     Izračunaj gubitak  $L = \|\epsilon - \epsilon_\theta(x_t, t)\|^2$
  - 8:     Ažuriraj parametre modela s gradijentom  $\nabla_\theta L$
  - 9: **end for**
- 

(a) Algoritam Treniranja

---

**Algorithm 2** DDPM Uzorkovanje

---

- 1: **Ulaz:** Treniran model  $\epsilon_\theta$ , raspored šuma  $\{\alpha_t\}_{t=1}^T$
  - 2: Uzorkuj  $x_T \sim \mathcal{N}(0, I)$
  - 3: **for**  $t = T$  **do** 1 **do**
  - 4:     Uzorkuj  $\epsilon \sim \mathcal{N}(0, I)$  ako je  $t > 1$ , inače  $\epsilon = 0$
  - 5:     Izračunaj  $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(x_t, t) \right) + \sqrt{1 - \alpha_{t-1}}\epsilon$
  - 6: **end for**
  - 7: **Povrat:**  $x_0$
- 

(b) Algoritam Uzorkovanja

Algoritam treniranja dodaje šum originalnom podatku za neki trenutak  $t$  te trenira neuronsku mrežu da predvidi taj šum, a algoritam uzorkovanja uzorkuje  $x_t \sim \mathcal{N}(0, I)$  te ga odšumljava koristeći korak 5 koji nalikuje langevinovoj dinamici i vraća  $x_0$  što predstavlja generiranu sliku. Ako želimo generirati podatke uz dani tekst onda bismo mogli koristiti novu neuronsku mrežu koja će u svakom koraku uzorkovanja predviđati klasu [16], a možemo i parametrizirati neuronsku mrežu koja predviđa šum sa klasom  $\epsilon_\theta(z_t, t, y)$  i tijekom treniranja povremeno zanemariti značajku  $y$ . Takav model može generirati podatke bez klase  $\epsilon_\theta(z_t, t, y) = \epsilon_\theta(z_t, t, y = \emptyset)$  (eng. *Classifier-free guidance*) [17].

Najčešće korišteni model neuronske mreže u difuzijskim modelima je U-Net [18] jer se pokazao izrazito preciznim u generiranju maski za segmentaciju, stoga je prikladan za generiranje slika iste veličine, a modeli s tekstom koriste embedding iz već istreniranog CLIP modela od OpenAI-a, [19], [13].

Evaluiranje generativnih modela je predmet istraživanja sam za sebe. Kod jednostavnih problema kao što je linearna regresija ili klasifikacija možemo imati zaseban skup podataka na kojem ćemo mjeriti kvalitetu generalizacije. Kod generativnih modela to nije primjenjivo jer u većini slučajeva ne postoji "ground-truth". Generativni model je distribucija  $q(x)$  koje bi trebala modelirati distribuciju podataka  $p(x)$ . Kod uspoređivanja dviju distribucija vrlo često koristimo divergencije kao što je Kullback-Leibler koja gubi korist u velikim dimenzijama npr. slike. Kod evaluiranja generiranja slika često se koristi Fréchet Inception Distance (FID)

$$\text{FID} = \|\mu_m - \mu_d\|^2 + \text{tr}(\Sigma_d + \Sigma_m - 2(\Sigma_d \Sigma_m)^{1/2}). \quad (10)$$

Problem s FID je to što mjeri samo udaljenost između distribucija bez obzira na kvalitetu slika. Visoki FID može značiti da model ne generira kvalitetne podatke, da previše koncentrira masu oko distribucije podataka ili da model samo generira dio podataka, a ne cijeli podatkovni skup [20]. Budući da je korišteni skup podataka relativno jednostavan nećemo koristiti nikakve numeričke mjere za mjerenje kvalitete generiranih slika nego ćemo ručno usporediti originalne slike i njihove rekonstrukcije sa generiranim slikama i njihovim rekonstrukcijama.

## 4 Implementacija

### 4.1 Priprema podataka

U izvornim podacima snimke su podijeljene na 3 isječka, svaki s jednim objektom iz scene. Također, tijekom snimanja se razlikuje horizontalna i vertikalna polarizacija, za ovaj rad to nije bitno. Svaka snimka dolazi u obliku: `{klasa}_{pozicija na sceni (1,2,3)}_{polarizacija}_{id snimanja}.npy` veličine matrice `[1024, 20]`. 50 je različitih id-jeva, ukupno 300 isječaka ili 100 cijelih snimki. Za svaki id snimanja i svaku polarizaciju stvoreno je ukupno  $3!$  permutacija pozicija i time je broj snimki povećan sa 100 na 600 snimki veličine `[1024, 60]`. U-Net model korišten u implementaciji zahtjeva da slika bude kvadratna. Sve snimke su transformirane iz veličine `[1024, 60] → [256, 256]`. Snimke su pretvorene u crno-bijeli png format gdje poprimaju cjelobrojne vrijednosti `[0, 255]`. Za uvjetni model koji koristi tekst svaka slika ima svoj atribut "text" koji se sastoji od triju skraćenih naziva klasa:

```
{"file_name": "PR_PS_GR_46.png", "text": "pr ps gr"},  
{"file_name": "GS_GR_PR_90.png", "text": "gs gr pr"},  
{"file_name": "AR_PS_GR_43.png", "text": "ar ps gr"}  
...
```

### 4.2 HuggingFace diffusers

Za treniranje modela korištena je open-source biblioteka HuggingFace Diffusers [21] zbog lakoće treniranja i korištenja difuzijskih modela i javno dostupnih skripti za treniranje. Huggingface diffusers koristi takozvani pipeline (DDPMPipeline, StableDiffusionPipeline itd.) za generiranje slika. Pipeline

se sastoji od schedulera i modela neuronske mreže. Scheduler tijekom treniranja definira kako se dodaje šum, a tijekom uzorkovanja definira kako se odstranjuje šum. Za oba istrenirana modela koristimo `DDPMScheduler` [14].

### 4.3 Bezuvjetni model

Model korišten za bezuvjetno generiranje slika (unconditional image generation) je U-Net. U-Net se sastoji od enkodera i dekodera. Enkoder smanjuje veličinu (preslikava nelinearnim funkcijam) slike za 2 puta u svakom bloku enkodera, zatim se smanjena verzija ponovno uvećava do početne veličine. Da bi se smanjio gubitak informacije kod smanjenja slike izlaz svakog bloka u enkoderu se također koristi u odgovarajućem bloku u dekoderu. U svakom bloku enkodera i dekodera se vrijeme u difuzijskom procesu kodira u sliku i izlaz cijelog modela je predviđen šum u trenutku  $t$ . Radarske snimke imaju jedan kanal pa je i to izlaz U-Neta. Korišteni U-Net ima 8 blokova za enkoder i dekoder. Broj kanala u svakom bloku je `block_out_channels`.

```
1 model = UNet2DModel(  
2     sample_size=256,  
3     in_channels=1,  
4     out_channels=1,  
5     layers_per_block=2,  
6     block_out_channels=(64, 64, 128, 128, 256, 256, 512, 512),  
7     down_block_types=(  
8         "DownBlock2D",  
9         "DownBlock2D",  
10        "DownBlock2D",  
11        "DownBlock2D",  
12        "DownBlock2D",  
13        "DownBlock2D",  
14        "AttnDownBlock2D",  
15        "DownBlock2D"  
16    ),  
17    up_block_types=(
```

```

18     "UpBlock2D",
19     "AttnUpBlock2D",
20     "UpBlock2D",
21     "UpBlock2D",
22     "UpBlock2D",
23     "UpBlock2D",
24     "UpBlock2D",
25     "UpBlock2D"
26 )
27 )

```

Scheduler za bezuvjetni model definira difuzijski proces u 500 koraka, a varijance u svakom koraku kreću od  $\beta_1 = 0.0001$  te se povećavaju do  $\beta_{500} = 0.02$  linearno.

## 4.4 Uvjetni model

Uvjetni model prima opis scene da bi izgenerirao sliku. Najšire korišteni uvjetni model je Stable Diffusion [22]. Stable diffusion koristi 3 modela **VAE**, **U-Net** i **Text-encoder**.

**Varijacijski autoenkoder (VAE)** prikazan na Slici 3 sastoji se od dva dijela, enkodera i dekodera. Enkoder transformira sliku u prostor manje dimenzije, latentni prostor. Kod Stable Diffusion veličina latentnog prostora je (4, 64, 64). Enkoder pretvara sliku iz latentnog prostora u slike. Šum se dodaje latentnoj reprezentaciji. Tijekom generiranja odšumljene latentne reprezentacije se pretvore natrag u sliku kroz dekodeer.

**U-Net** u Stable Diffusion je manji po broju parametara jer je latentni prostor manje dimenzije nego što su slike. Korišteni U-Net ima 4 bloka te smanjuje sliku veličine [4, 64, 64] na [1280, 4, 4]

**Text-encoder** daje numeričku reprezentaciju (engl. embedding) za ulazni tekst. On se ne mijenja tijekom treniranja, već se koristi gotovi CLIP model [19]. U ovom radu se koristio već istrenirani model Stable Diffusion v1.5. od RunawayML [23]. Scheduler u uvjetnom modelu definira difuzijski proces sa 1000 koraka odšumljavanja, a varijance se povećavaju od  $\beta_0 = 0.00085$  do  $\beta_{1000} = 0.012$  linearno.

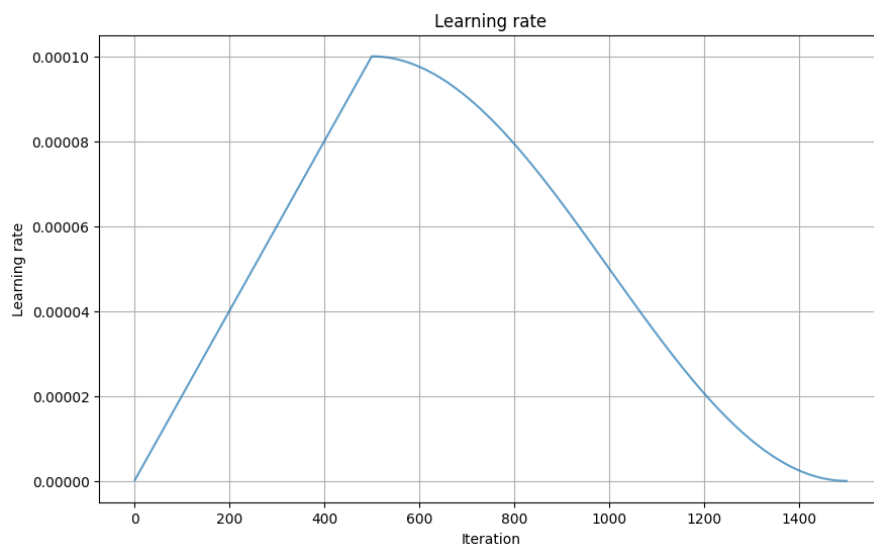


## 5 Treniranje modela

Modeli su trenirani na računalnom klasteru Supek [24] koristeći jednu Nvidia A100 40GB grafičku karticu.

### 5.1 Bezuvjetni model

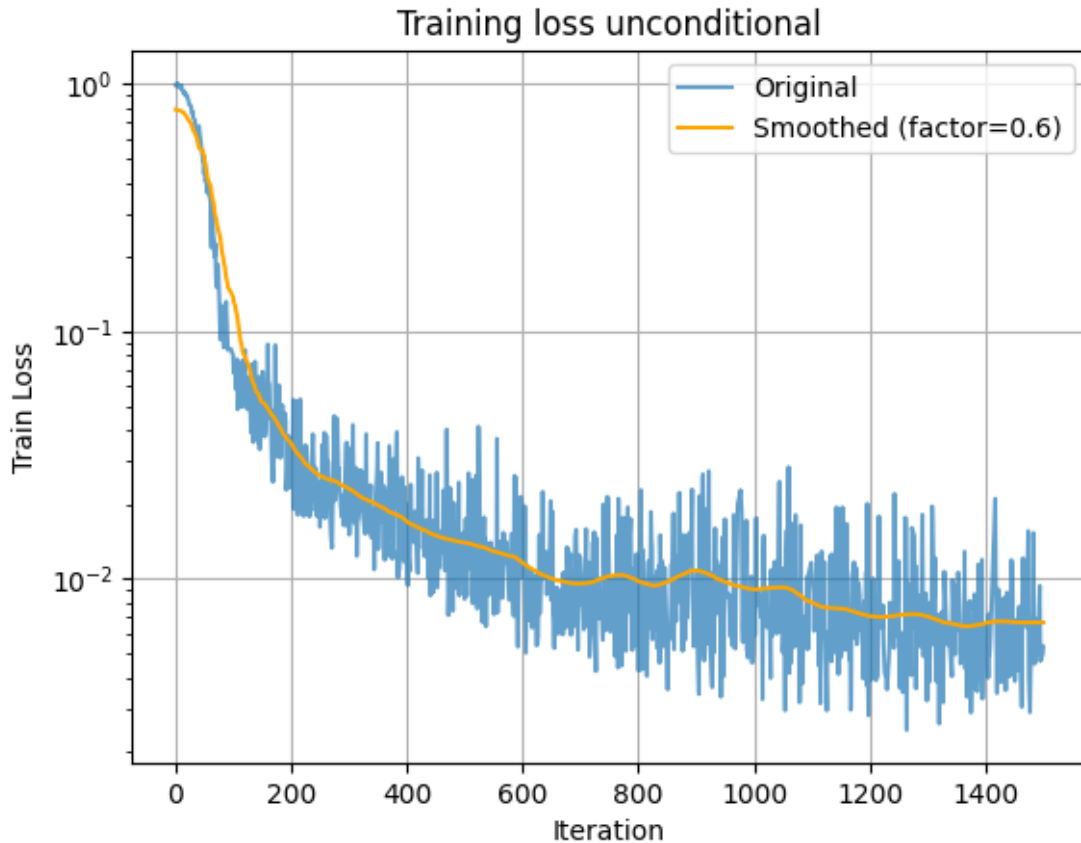
Skripta za treniranje bezuvjetnog modela je prilagođena iz već dostupne skripte na HuggingFace Diffusers GitHub stranici <sup>1</sup>. Bezuvjetni model je treniran 50 epoha sa stopom učenja  $10^{-4}$  uz 500 iteracija warmupa, te se nakon  $10^{-4}$  smanjuje prateći cosine scheduler Slika 5. Za računanje funkcije gubitka koristi



Slika 5: Stopa učenja za bezuvjetni model.

se min-SNR weighting gdje se minimizira signal-to-noise ratio (SNR) i time je konvergencija učenja nekoliko puta brža [25].

<sup>1</sup>[https://github.com/huggingface/diffusers/tree/main/examples/unconditional\\_image\\_generation](https://github.com/huggingface/diffusers/tree/main/examples/unconditional_image_generation)



Slika 6: Gubitak bezuvjetnog modela na logaritamskoj skali.

Nakon 12 minuta treniranja gubitak se približava prema 0.005. Na Slici 6 vidimo da se model ustabilio već u 600-toj iteraciji.

## 5.2 Uvjetni model

Uvjetni model temeljen na Stable Diffusion je istreniran kroz 150 epoha, ukupno 4500 iteracija. Stopa učenja je slična bezuvjetnom modelu. Povećava se prvih 500 epoha do  $10^{-4}$  te se smanjuje prateći cosine scheduler. Nakon jednog sata uvjetni model se nije ustabilio kao bezuvjetni što se vidi iz trenda na Slici 7.

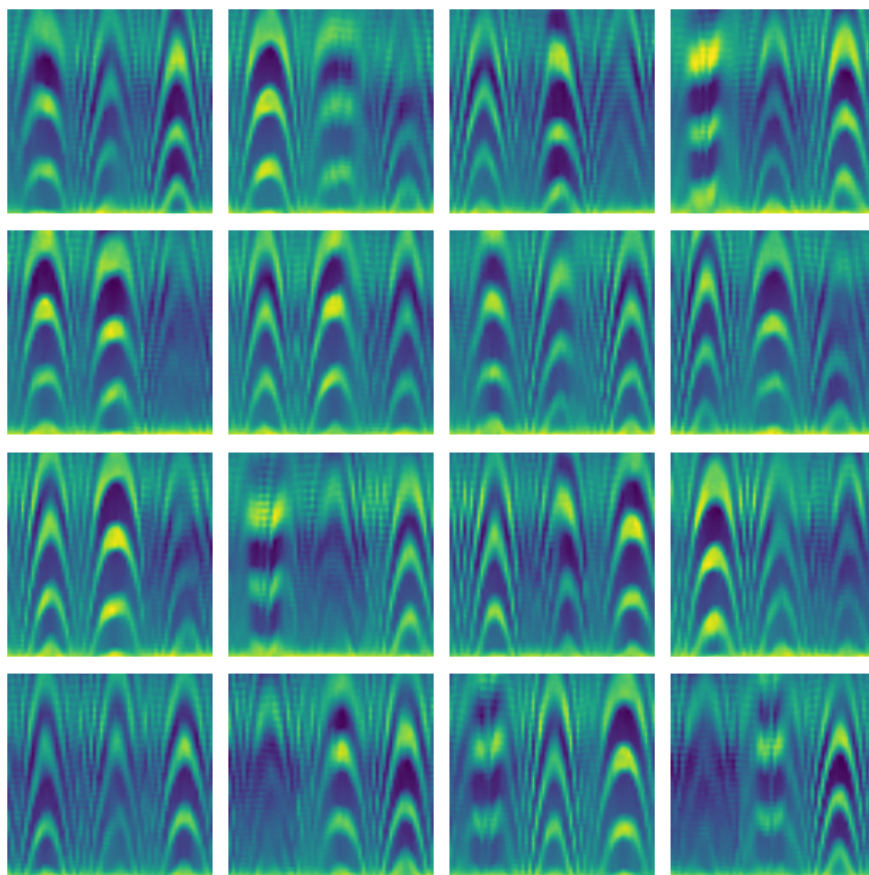


Slika 7: Gubitak uvjetnog modela na logaritamskoj skali.

## 6 Rezultati

### 6.1 Generiranje bezuvjetnim modelom

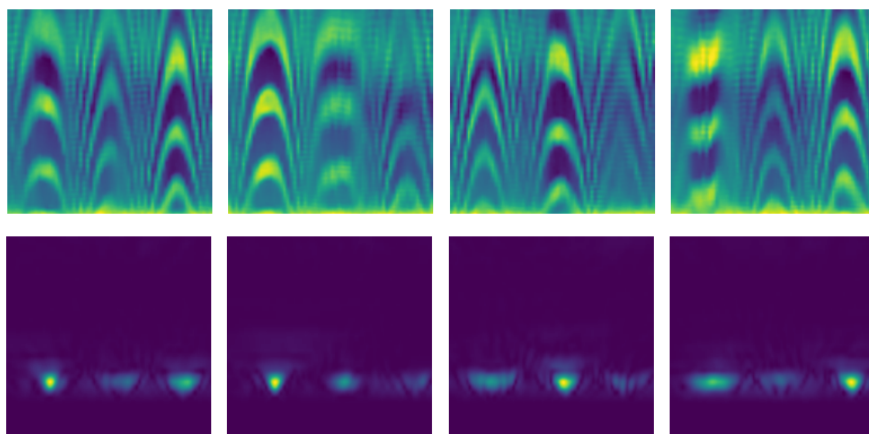
Na Slici 8 možemo vidjeti 16 generiranih radarskih slika. Generiranje ovih 16 slika traje oko 5 minuta. Svaka slika izgleda kao što izgledaju originalne. Model ne generira slike koje sadrže npr. 3 ista objekta jer takvih slika nije bilo u skupu za treniranje. Na slici 9 možemo vidjeti rekonstrukciju prvih 4 generiranih radarskih slika. Rekonstrukcije radarskih slika odgovaraju njihovim intenzitetima za pojedine objekte i možemo reći da su generirane slike realistične i po izgledu sirovih radarskih snimki i njihovih rekonstrukcija.



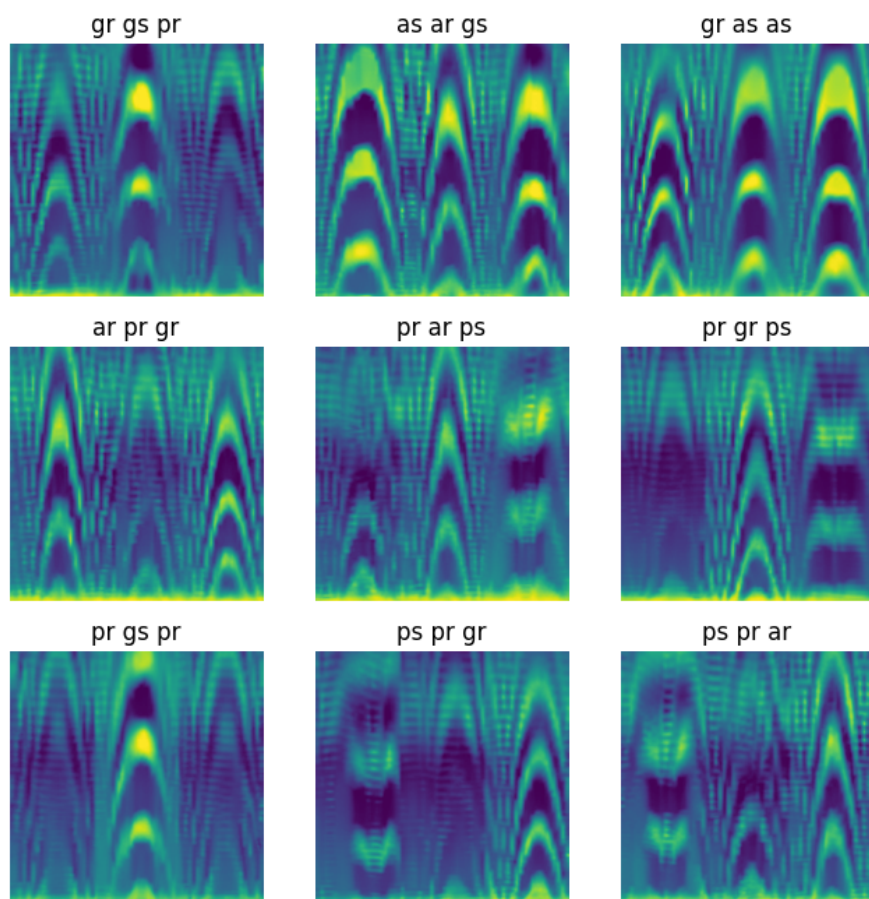
Slika 8: Primjeri radarskih snimki generiranih bezuvjetnim modelom.

## 6.2 Generiranje uvjetnim modelom

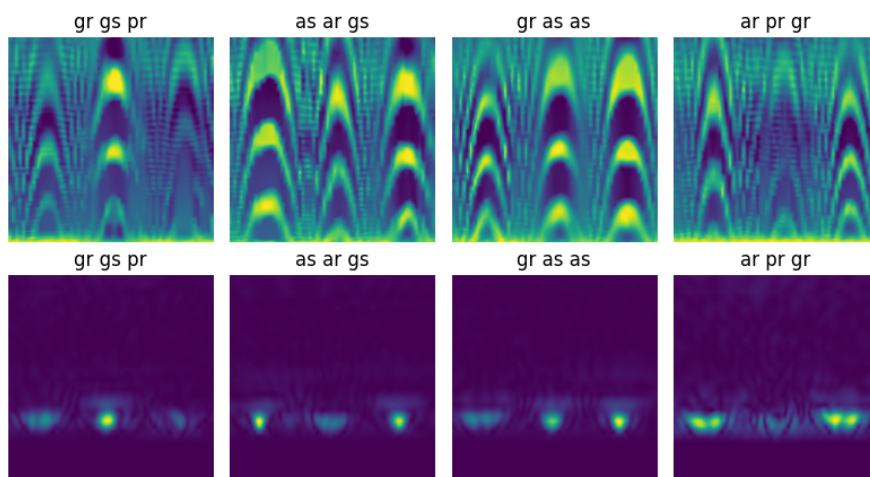
Na Slici 10 prikazano je 9 radarskih snimki generiranih nasumičnim odabirom teksta. Potrebna je oko 1 minuta za generiranje ovih 9 slika. Uvjetni model može stvarati radarske slike scene koja nije u skupu za treniranje npr. pr gs pr. Tranzicije između pojedinih isječaka na snimci nisu prirodne kao kod bezuvjetnog modela npr. ps pr gr. Intenziteti odgovaraju objektima. Rekonstrukcija prvih 4 slika sa Slike 10 je na Slici 11. Sve rekonstrukcije odgovaraju intenzitetima objekata kao što je prikazano na Slici 2.



Slika 9: Radarske snimke bezuvjetnog modela i njihove rekonstrukcije.



Slika 10: Primjeri radarskih snimki generiranih uvjetnim modelom



Slika 11: Rekonstrukcija prvih 4 generiranih slika

## 7 Zaključak

U ovom radu kreirana su dva generativna modela za generiranje sirovih SAR snimki, uvjetni i bezuvjetni model. Bezuvjetni model generira slike s 3 nasumična objekta iz skupa za učenje. Kvaliteta radarske snimke generirane bezuvjetnim modelom i njene rekonstrukcije je u skladu sa originalnim snimkama i njihovim rekonstrukcijama što vidimo na Slikama 8 i 9. Najveći nedostatak bezuvjetnog modela jest to što ne znamo koji objekti su na slici. Također, brzina izvođenja modela, tj. generiranja snimke bezuvjetnim modelom je nekoliko puta dulja od generiranja snimke uvjetnim modelom. Uvjetni model generira radarske snimke sa 3 zadana objekta. Na Slici 10 možemo vidjeti 9 različitih slika generiranih uvjetnim modelom. Na Slici 11 možemo vidjeti da uvjetni model generira realistične radarske snimke u smislu realistične rekonstrukcije. Uvjetni model je puno brži od bezuvjetnog modela jer se difuzijski proces odvija u manjoj dimenziji. Iako uvjetni model može generirati snimku bilo koje zadane scene granice između objekata nisu glatke kao u originalnim snimkama i snimkama bezuvjetnog modela. Unatoč tome, rekonstrukcije uvjetnog modela su dobre. Korištenje difuzijskih modela za generiranje radarskih snimki omogućuje brzo stvaranje realističnih slika različitih scena s visokom kvalitetom rekonstrukcije za daljne potrebe. Zaključno, bezuvjetni model generira kvalitetnije slike sa nepoznatim klasama, ali je sporiji i računalno zahtjevniji. Uvjetni model omogućuje brže generiranje specifičnih scena, iako ponekad generirani objekti nisu savršeno precizni.

## 8 Literatura

- [1] Filip Turčinović. Ground based sar data obtained with different polarizations, 2024.
- [2] Filip Turčinović. *Sustav za snimanje radarom sa sintetičkom aperturom i integriranom klasifikacijom objekata*. PhD thesis, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2024. Disertacija.
- [3] F. Turčinović, M. Kačan, D. Bojanjac, and M. Bosiljevac. Impact of ground based sar parameters on radar data based object classification. In *24th International Conference on Applied Electromagnetics and Communications (ICECOM)*, 2023.
- [4] Marin Kačan, Filip Turčinović, Dario Bojanjac, and Marko Bosiljevac. Deep learning approach for object classification on raw and reconstructed gbsar data. *Remote Sensing*, 14(22), 2022.
- [5] F. Turčinović, M. Kačan, D. Bojanjac, and M. Bosiljevac. Ground-based sar system for object classification with parameter optimization based on deep learning feedback algorithm. In *Proc. SPIE 12733, Image and Signal Processing for Remote Sensing XXIX*, page 127330R, 2023.
- [6] D. Tognon, F. Turčinović, and M. Bosiljevac. Deep learning approaches for object detection in short-range gbsar images. In *2023 International Symposium ELMAR*, 2023.
- [7] F. Turčinović, M. Kačan, D. Bojanjac, and M. Bosiljevac. Deep learning approach based on gbsar data for detection of defects in packed objects. In *7th European Conference on Antennas and Propagation (EuCAP)*, 2023.



- [8] F. Turčinović. Ground based sar data for classification - 9 objects in the near distance, 2023.
- [9] F. Turčinović. Near-distance raw and reconstructed ground based sar data, 2023.
- [10] F. Turčinović. Ground based sar data of cardboard box with aluminium bottles for anomaly detection, 2022.
- [11] Filip Turčinović. Omegak. <https://github.com/filt27/OmegaK>, 2024. Accessed: 2024-08-20.
- [12] Lilian Weng. What are diffusion models? *lilianweng.github.io*, Jul 2021.
- [13] Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023.
- [14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.
- [15] Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient langevin dynamics. In *International Conference on Machine Learning*, 2011.
- [16] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021.
- [17] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [19] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack

- Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.
- [20] Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.
- [21] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, William Berman, Yiyi Xu, Steven Liu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models. <https://github.com/huggingface/diffusers>, 2022.
- [22] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.
- [23] runawayml. runwayml/stable-diffusion-v1-5. <https://huggingface.co/runwayml/stable-diffusion-v1-5>. Accessed: 2024-08-27.
- [24] Srce. Supek supercomputer. [wiki.srce.hr/display/NR/Supek](http://wiki.srce.hr/display/NR/Supek).
- [25] Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Dong Chen, Han Hu, Xin Geng, and Baining Guo. Efficient diffusion training via min-snr weighting strategy, 2024.

# Sažetak

**Autor:** Luka Mucko

**Naslov:** Generiranje radarskih slika difuzijskim modelima

Snimanje radarskih slika sintetičkom aperturom je vremenski zahtjevno i za treniranje klasifikacijskoga modela potrebno je stotina takvih slika. U ovom radu upotrebljavamo dva difuzijska modela, uvjetni temeljen na Stable Diffusion koji generira snimke uz dani tekstualni opis scene i bezuvjetni koji generira nasumične scene. Oba modela generiraju kvalitetne radarske slike sa realističnim rekonstrukcijama.

**Ključne riječi:** Ground Based Synthetic Aperture Radar (GBSAR), difuzijski modeli, Stable Diffusion, Denoising Diffusion Probabilistic Models (DDPM)

# Summary

**Author:** Luka Mucko

**Title:** Generating radar images using diffusion models

Capturing synthetic aperture radar images is time-consuming, and training a classification model requires hundreds of such images. In this paper, we use two diffusion models: a conditional model based on Stable Diffusion, which generates images based on a given textual description of the scene, and an unconditional model that generates random scenes. Both models produce high-quality images with realistic reconstructions.

**Key words:** Ground Based Synthetic Aperture Radar (GBSAR), diffusion models, Stable Diffusion, Denoising Diffusion Probabilistic Models (DDPM)