

# Razvoj modela dubokog učenja za uklanjanje gipsa na pedijatrijskim medicinskim rendgenskim slikama

---

**Vičević, Dominik**

**Master's thesis / Diplomski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:190:362752>

*Rights / Prava:* [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

*Download date / Datum preuzimanja:* **2025-02-17**



*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Engineering](#)



SVEUČILIŠTE U RIJECI  
**TEHNIČKI FAKULTET**  
Sveučilišni diplomski studij računarstva

Diplomski rad

**Razvoj modela dubokog učenja za  
uklanjanje gipsa na pedijatrijskim  
medicinskim rendgenskim slikama**

Rijeka, srpanj 2024.

Dominik Vičević  
0069088197

SVEUČILIŠTE U RIJECI  
**TEHNIČKI FAKULTET**  
Sveučilišni diplomski studij računarstva

Diplomski rad

**Razvoj modela dubokog učenja za  
uklanjanje gipsa na pedijatrijskim  
medicinskim rendgenskim slikama**

Mentor: prof. dr. sc. Ivan Štajduhar

Rijeka, srpanj 2024.

Dominik Vičević  
0069088197

Rijeka, 12.03.2024.

Zavod:                   Zavod za računarstvo  
Predmet:                Strojno učenje

## ZADATAK ZA DIPLOMSKI RAD

Pristupnik:           **Dominik Vičević (0069088197)**  
Studij:                Sveučilišni diplomski studij računarstva (1400)  
Modul:                 Programsko inženjerstvo (1441)

Zadatak:              **Razvoj modela dubokog učenja za uklanjanje gipsa na pedijatrijskim medicinskim rendgenskim slikama / Development of a deep learning model for cast removal on pediatric medical X-ray images**

Opis zadatka:

Proučiti i opisati metode dubokog učenja za translaciju ne-uparenih slika. Selektirane modele potrebno je istrenirati nad pedijatrijskim medicinskim rendgenskim slikama s fokusom na više različitih dijelova tijela u svrhu uklanjanja/supresije gipsa. Modele je potrebno evaluirati i, ako je moguće, verificirati s radiolozima. Prilikom selekcije modela potrebno je uzeti u obzir i metodu stabilne difuzije te ispitati mogućnosti korištenja iste u navedenom problemu.

Rad mora biti napisan prema Uputama za pisanja diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.

Zadatak uručen pristupniku:    20.03.2024.

Mentor:  
prof. dr. sc. Ivan Štajduhar

Komentor:  
dr. sc. Franko Hržić

Predsjednik povjerenstva za  
diplomski ispit:  
prof. dr. sc. Miroslav Joler



## Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, srpanj 2024.

Domink Včević

Ime Prezime



# Sadržaj

Popis slika	viii
Popis tablica	ix
<b>1 Uvod</b>	<b>1</b>
<b>2 Modeli</b>	<b>4</b>
2.1 Generative Adversarial Networks . . . . .	4
2.1.1 CycleGAN . . . . .	5
2.1.2 HarmonicGAN . . . . .	7
2.2 Difuzijski modeli . . . . .	10
2.2.1 CycleDiffusion . . . . .	13
<b>3 Podaci</b>	<b>15</b>
3.1 Skup podataka . . . . .	15
3.2 Uravnoteženje podataka . . . . .	18
<b>4 Postavljanje eksperimenta</b>	<b>22</b>
4.1 Eksperiment . . . . .	22
4.2 Mjerenje performansi . . . . .	24
<b>5 Rezultati</b>	<b>27</b>

*Sadržaj*

<b>6</b>	<b>Diskusija</b>	<b>31</b>
<b>7</b>	<b>Zaključak</b>	<b>44</b>
	<b>Bibliografija</b>	<b>46</b>
	<b>Pojmovnik</b>	<b>49</b>
	<b>Sažetak</b>	<b>49</b>

# Popis slika

1.1	Usporedba rendgenske snimke bez gipsa sa snimkom s gipsom. . . .	3
2.1	Topologija CycleGAN metode . . . . .	7
2.2	Primjer nedostatka dosljednosti ciklusa . . . . .	8
2.3	Topologija HarmonicGAN modela . . . . .	9
2.4	Prikaz procesa difuzije prema naprijed i procesa obrnute difuzije . .	11
2.5	Primjer ubrzanog generiranja difuzijskog modela . . . . .	13
3.1	Distribucija anatomskih regija skupa podataka . . . . .	16
6.1	Prikaz prosječne FID metrike po regijama . . . . .	32
6.2	Primjer slike regije potkoljenice . . . . .	33
6.3	Prikaz FID metrike za regije šaka i noga . . . . .	34
6.4	Primjer gipsa regije šaka . . . . .	35
6.5	Primjer slike regije noga . . . . .	36
6.6	Primjer projekcija slika regije petne kosti . . . . .	37
6.7	Primjer uklanjanja fraktura difuzijskog modela . . . . .	38
6.8	Primjer mijenjanja strukture kosti difuzijskog modela . . . . .	39
6.9	Usporedba generiranih slika difuzijskih modela . . . . .	40
6.10	Mozaik generiranih slika (prvi dio) . . . . .	42
6.11	Mozaik generiranih slika (drugi dio) . . . . .	43

# Popis tablica

3.1	Sažetak statistike skupa podataka . . . . .	15
3.2	Količine podataka po regijama. . . . .	17
3.3	Pregled konfiguracije treninga i validacije . . . . .	21
5.1	Metrike modela računane na svim regijama . . . . .	28
5.2	Metrike modela računane isključivo na regiji zapešća . . . . .	29
5.3	Prosječne metrike modela izračunate zasebno za svaku regiju . . . .	30

# Poglavlje 1

## Uvod

Gips je potporno sredstvo koje okružuje ozlijeđeni dio tijela kako bi ga zaštitilo, imobiliziralo i potaknulo brže zacjeljivanje [1]. U modernoj ortopediji, postoperativna skrb vezana uz ortopedске traume klasično se tretira koristeći gipseve ili udlage. Imobilizacija potiče zacjeljivanje ograničavanjem kretanja, čime se sprječava oštećenje okolnog mekog tkiva [2].

Na rendgenskoj snimci kosti, gips se prikazuje kao homogena bijela ili svjetlosiva struktura koja obavija ozlijeđenu kost ili dio kosti. Primjer slike s gipsom apliciranim na kosti zapešća prikazan je na slici 1.1. Zbog svoje gustoće, gips otežava prolazak rendgenskih zraka, što smanjuje vidljivost detalja i unutarnjih struktura ispod gipsa. Neke od poteškoća koje se mogu javiti uključuju teškoće uočavanja zarastanja kostiju ili oticanja tkiva. Dodatno, nije uvijek moguće ukloniti gips prije snimanja rendgenskih snimaka, što otežava postupak prikupljanja podataka fraktura bez prisutnosti gipsa na kostima. U ovim slučajevima bilo bi od izrazite koristi smanjiti ili potpuno ukloniti utjecaj gipsa sa slike.

U istraživanju [3], Hržić et al. su koristili CycleGAN [4] model za uklanjanje gipsa s rendgenskih snimki zapešća. Navedeno istraživanje predstavlja jedinu poznatu metodu uklanjanja gipsa sa slika primjenom tehnika strojnog učenja.

Zadatak uklanjanja gipsa sa slika je zadatak translacije slike u sliku (engl. *image to image translation*). Budući da je zahtjevno i skupo prikupiti skup parova slika za treniranje modela, koriste se dvije različite domene neuparenih slika. Cilj je da

## Poglavlje 1. Uvod

model sam nauči značajke tih dviju domena, te za danu sliku iz jedne domene, translatira tu sliku u drugu domenu, generirajući vjerodostojnu i reprezentativnu inačicu slike. U navedenom slučaju, korištenje generativnih modela je neophodno. Neki od najpoznatijih generativnih modela koji se koriste u translaciji slike u sliku su: GAN (engl. *generative adversarial network*) modeli [5], VAE (engl. *variational autoencoders*) modeli [6], te difuzijski modeli [7]. VAE modeli se relativno lako treniraju, ali često rezultiraju mutnim slikama [8], te su zato iz ovog konkretnog istraživanja izostavljeni.

U ovom radu će se proširiti metodologija istraživanja [3] dodavanjem dodatnog GAN modela, zvanog HarmonicGAN [9]. Oba modela koriste neuparene parove slika za translaciju. Također će se istražiti difuzijski modeli, koji trenutno predstavljaju najnapredniju tehnologiju u sintezi slika [10]. Konkretno, bit će implementirana metoda CycleDiffusion [11], koja uključuje dva difuzijska modela trenirana na različitim domenama s ciljem postizanja translacije slike u sliku. Sve navedeno će biti napravljeno na proširenom skupu slika od 17 dijelova (regija) tijela. Također cilj je istaknuti prednosti i mane svih modela, te procijeniti iskoristivost modela u kliničkoj praksi.





(a) Rendgenska snimka bez gipsa.



(b) Rendgenska snimka s gipsom.

Slika 1.1 Usporedba rendgenske snimke bez gipsa sa snimkom s gipsom.

# Poglavlje 2

## Modeli

U ovome poglavlju nalazi se detaljan opis korištenih generativnih modela.

### 2.1 Generative Adversarial Networks

"Generative Adversarial Network" (GAN u daljnjem tekstu) su duboke neuronske mreže koje se koriste za generiranje novih podataka sličnih skupu podataka na kojem su trenirane. GAN-ovi su bazirani na igri nulte sume (engl. *zero-sum game*), gdje se dva modela, generator i diskriminator, natječu. Generator pokušava stvoriti podatke koji pripadaju distribuciji stvarnih podataka, dok diskriminator pokušava dokučiti jesu li generirani podaci stvarni ili ne. Kroz iteracije ovog natjecanja, oba modela se postepeno poboljšavaju dok generirani podaci ne postanu nerazlučivi od stvarnih [5]. Da bi se GAN mogao trenirati, koristi se suparnički gubitak (engl. *adversarial loss*), koji se računa na sljedeći način.

Za skup podataka  $X$  kojemu su uzorci  $\{x_i\}_{i=1}^N$  gdje je  $x_i \in X$  i distribucijom podataka  $x \sim p_{data}(x)$ , generator  $G(z; \theta_g)$  uči distribuciju  $p_g$  tako da  $p_g = p_{data}$ . Generator  $G(z; \theta_g)$  je derivabilna funkcija koja preslikava ulazni šum  $z \sim p_z(z)$  u prostor podataka, gdje su  $\theta_g$  parametri generatora. U daljnjem tekstu će se  $G(z; \theta_g)$  skratiti kao  $G$ . Isto tako se definira diskriminator  $D(x; \theta_d)$  kojem je izlaz realni broj koji predstavlja vjerojatnost da je  $x$  uzorkovan iz  $p_{data}$  umjesto iz  $p_g$ . U daljnjem tekstu će se  $D(x; \theta_d)$  skratiti kao  $D$ .

## Poglavlje 2. Modeli

Model se trenira tako što se za  $D$  maksimizira  $\log D(x)$ , dok se istovremeno za  $G$  minimizira  $\log(1 - D(G(z)))$ . Suparnički gubitak je onda definiran kao:

$$\mathcal{L}_{supar.}(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))], \quad (2.1)$$

gdje se, tijekom treninga, odvija "minimax" igra između generatora  $G$  i diskriminatora  $D$ :  $\min_G \max_D \mathcal{L}_{supar.}(G, D)$  [5].

Navedeni izvorni GAN model generira slike iz ulaznog šuma, što se ne može primijeniti u translaciji slike u sliku. Potrebno je dodatno prilagoditi mrežu.

### 2.1.1 CycleGAN

CycleGAN [4] je nadogradnja na istraživanje Isola et al. [12], u kojemu je korišten uvjetni GAN (engl. *conditional GAN*) u svrhu okvira za translaciju slike iz jedne domene u sliku druge domene zvanom "pix2pix". Problem navedenog pristupa je što su potrebni upareni podaci za trening modela. CycleGAN se rješava potrebe za uparenim podacima koristeći dosljednost ciklusa (engl. *cycle consistency*).

Cilj CycleGAN-a je naučiti funkcije preslikavanja između dvije dane domene  $X$  i  $Y$ , gdje su uzorci  $X$  domene:  $\{x_i\}_{i=1}^N$  gdje je  $x_i \in X$ , a uzorci  $Y$  domene:  $\{y_j\}_{j=1}^M$  gdje je  $y_j \in Y$ . Njihove distribucije su  $x \sim p_{data}(x)$  i  $y \sim p_{data}(y)$ . Definirane su dvije funkcije preslikavanja: generator  $G : X \rightarrow Y$  (koji uzima slike iz  $X$  domene i pretvara ih u  $Y$  domenu) i generator  $F : Y \rightarrow X$  (koji uzima slike iz  $Y$  domene i pretvara ih u  $X$  domenu), te dva diskriminatora  $D_x$  i  $D_y$ . Diskriminator  $D_x$  razaznaje je li slika iz domene  $X$  ili je to generirana slika  $F(y)$ , dok diskriminator  $D_y$  razaznaje dali je slika iz domene  $Y$  ili je to generirana slika  $G(x)$  [4].

Isti suparnički gubitak iz originalnog GAN modela se u CycleGAN-u primjenjuje dva puta. Za  $G$  i diskriminator  $D_y$ :

$$\mathcal{L}_{GAN}(G, D_y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_y(G(x)))], \quad (2.2)$$

gdje generator  $G$  pokušava stvoriti slike slične domeni  $Y$ , tako da diskriminator

## Poglavlje 2. Modeli

$D_y$  ne zna ako je slika  $G(x)$  stvarna ili generirana. Slično vrijedi za generator  $F$  i diskriminator  $D_x$ .

Suparnički gubitak nije dovoljan za efektivnu translaciju slika. Naime, s dovoljnim kapacitetom, mreža može pretvoriti bilo koju sliku u bilo koju drugu, bez konkretne relacije između generirane slike i izvorne. Zato je potrebna dosljednost ciklusa. Dosljednost ciklusa znači da za bilo koju sliku  $x$  koja se translira u  $Y$  domenu, rekonstrukcija te slike u  $X$  domeni mora vratiti sliku sličnu originalnoj ( $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ ). Vrijedi i obrnuti slučaj, gdje se slika iz  $Y$  domene translira u  $X$  domenu, te nazad u  $Y$ .

Dosljednost ciklusa se potiče funkcijom gubitka ciklusa (engl. *cycle loss*) i definirana je ovako:

$$\mathcal{L}_{\text{ciklus}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1] \quad (2.3)$$

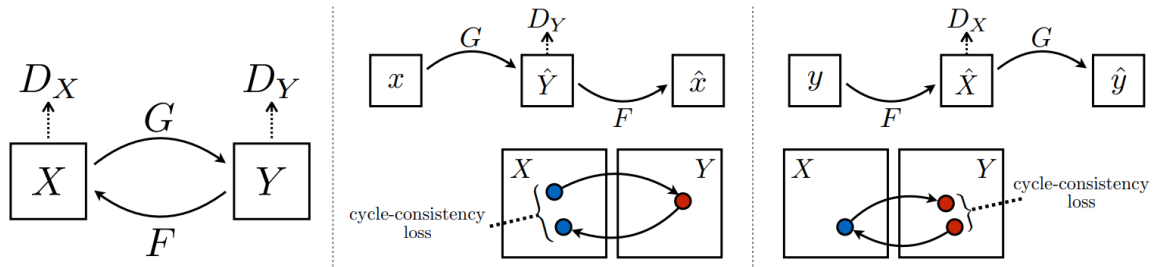
Jednadžba 2.3 prikazuje da je gubitak ciklusa zbroj L1 udaljenosti između rekonstruiranih slika i njihovih originala. Cilj generatora je minimizirati tu funkciju.

Spajajući suparnički gubitak i gubitak ciklusa, dobivamo punu funkciju gubitka CycleGAN-a:

$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) &= \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda_{\text{ciklus}} \times \mathcal{L}_{\text{ciklus}}(G, F), \end{aligned} \quad (2.4)$$

gdje je  $\lambda_{\text{ciklus}}$  hiperparametar koji kontrolira utjecaj gubitka ciklusa.

Prilikom treniranja generatori  $G$  i  $F$  pokušavaju minimizirati funkciju gubitka 2.4, dok je diskriminatori  $D_x$  i  $D_y$  pokušavaju maksimizirati; srodno s originalnim GAN modelom, odvija se "minimax" igra:  $\min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y)$  [4]. Topologija CycleGAN modela može se vidjeti na slici 2.1



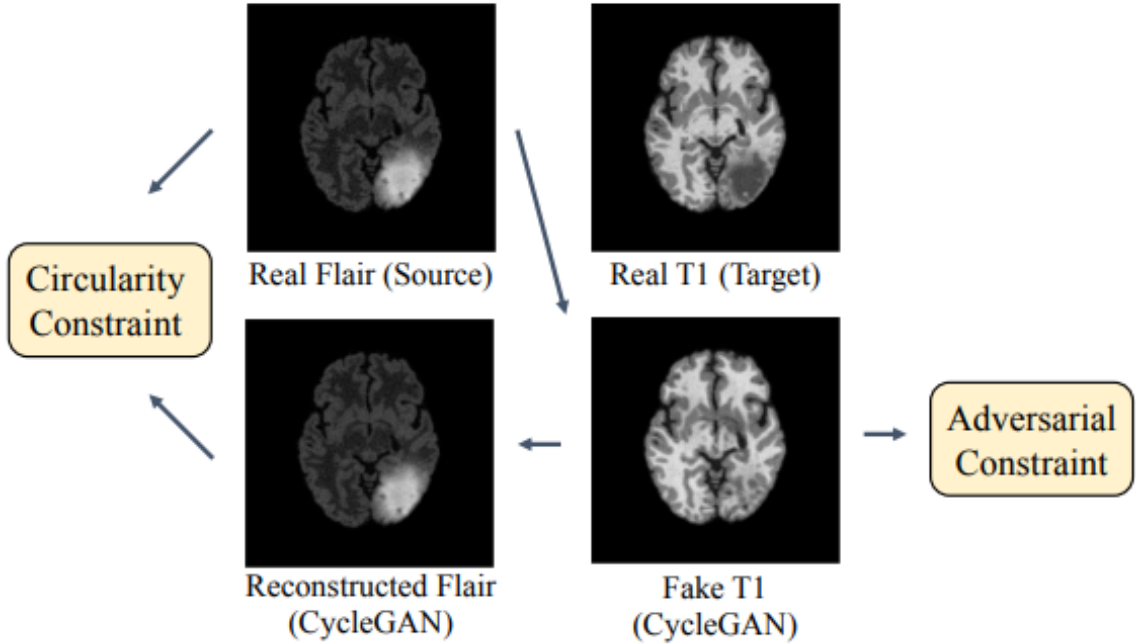
Slika 2.1 Topologija CycleGAN metode [4]. Na lijevom dijagramu prikazane su dvije funkcije preslikavanja (generatori  $G : X \rightarrow Y$  i  $F : Y \rightarrow X$ ), te njihovi odgovarajući diskriminatori. Na srednjem i desnom dijagramu prikazan je način računanja gubitka ciklusa (na slici "cycle-consistency loss") za oba generatora.

### 2.1.2 HarmonicGAN

HarmonicGAN je nadogradnja na CycleGAN. Naime, korištenjem dosljednosti ciklusa garantira se samo sličnost rekonstruirane slike i originala, ali ne potiče se da translatairana slika zadržava značajke originalne slike [9]. Primjer navedenog se može vidjeti na slici 2.2.

HarmonicGAN uklanja taj nedostatak korištenjem funkcije gubitka glatkoće (engl. *smoothness loss*). Neka se svaka slika iz originalnog skupa  $X$  podijeli na  $M$  dijelova slika (engl. *image patches*), te se za svih  $M$  dijelova, izračuna vektor značajki; taj skup je definiran kao  $S = \{\vec{x}(i), i = 1, 2, \dots, M\}$ , gdje  $\vec{x}(i)$  označava vektor značajki svakog  $i$ -tog dijela slike  $x \in X$ . Značajke  $i$ -tog dijela slike translatairane generatorima se označuju  $G(\vec{x})(i)$  i  $F(\vec{y})(i)$ . Zatim, neka je  $w_{ij}(X)$  (skalarna vrijednost) sličnost između vektora značajki  $\vec{x}(i)$  i vektora značajki  $\vec{x}(j)$ . Možemo definirati graf  $\mathcal{G}$ , čiji su čvorovi  $S$ , a čiji su rubovi sličnosti tih čvorova  $w_{ij}(X)$ . Gubitak glatkoće djeluje kao Laplaceov operator grafa nametnut na sve parove dijelova slike i definira se kao:

$$\begin{aligned} \mathcal{L}_{\text{glatko.}}(G, F, X) = \mathbb{E}_{x \in X} \left[ \sum_{i,j} w_{ij}(X) \times \text{Dist}[G(\vec{x})(i), G(\vec{x})(j)] \right. \\ \left. + \sum_{i,j} w_{ij}(G(X)) \times \text{Dist}[F(G(\vec{x})(i)), F(G(\vec{x})(j))] \right], \end{aligned} \quad (2.5)$$



Slika 2.2 Primjer nedostatka dosljednosti ciklusa [9]. Rekonstruirana slika može savršeno odgovarati izvornoj slici (dosljednost ciklusa), dok translaticirana slika gubi svojstvo izvorne slike. U ovom primjeru tumor iz izvorne slike ("Real Flair (Source)") se uklanja pri translaciji u ciljanu domenu ("Fake T1").

gdje je  $w_{ij}(X) = \exp\{-\text{Dist}[\vec{x}(i), \vec{x}(j)]/\sigma^2\}$  [13].  $\text{Dist}[\vec{x}(i), \vec{x}(j)]$  je funkcija udaljenosti između dva vektora značajki dijela slike, a  $\sigma$  služi kao regularizacijski hiperparametar. Jednadžba 2.5 definira gubitak glatkoće za translaciju slike iz X domene u Y domenu; slično se definira i za suprotan slučaj:

$$\begin{aligned} \mathcal{L}_{\text{glatko.}}(F, G, Y) = \mathbb{E}_{y \in Y} \left[ \sum_{i,j} w_{ij}(Y) \times \text{Dist}[F(\vec{y})(i), F(\vec{y})(j)] \right. \\ \left. + \sum_{i,j} w_{ij}(F(Y)) \times \text{Dist}[G(F(\vec{y})(i)), G(F(\vec{y})(j))] \right] \end{aligned} \quad (2.6)$$

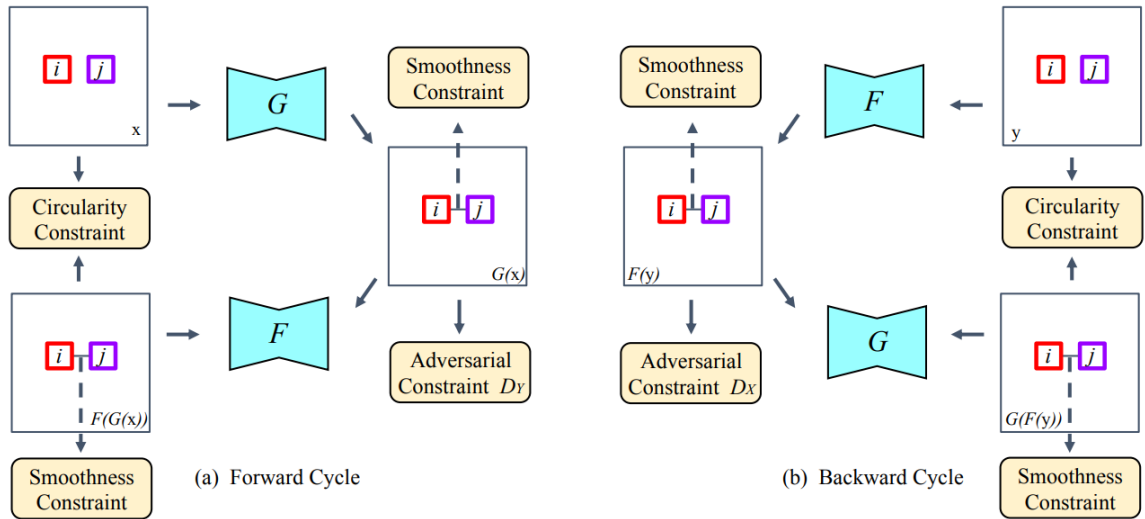
Udružena funkcija gubitka glatkoće je onda definirana kao  $\mathcal{L}_{\text{glatko.}}(G, F) = \mathcal{L}_{\text{glatko.}}(G, F, X) + \mathcal{L}_{\text{glatko.}}(F, G, Y)$ . Zbrajanjem navedenog gubitka glatkoće skupa sa funkcijama gubitka CycleGAN-a dobivamo:

Poglavlje 2. Modeli

$$\mathcal{L}_{\text{HarmonicGAN}}(G, F, D_X, D_Y) = \mathcal{L}_{\text{CycleGAN}}(G, F, D_X, D_Y) + \lambda_{\text{glatko.}} \times \mathcal{L}_{\text{glatko.}}(G, F), \quad (2.7)$$

gdje je  $\lambda_{\text{glatko.}}$  hiperparametar koji kontrolira utjecaj gubitka glatkoće [9].

Intuitivno, gubitak glatkoće uspoređuje sličnost originalnih dijelova slika i translahiranih dijelova. Ako je udaljenost između originalnih dijelova 0, sličnost će biti  $w_{ij}(X) = 1$ , te će gubitak ovisiti o sličnosti translahiranih slika. Što je udaljenost originalnih dijelova slika veća, to će se sličnost  $w_{ij}(X)$  približavati nuli, pa će onda udaljenost translahiranih dijelova slika biti manje bitna. Na slici 2.3 nalazi se ilustracija topologije HarmonicGAN modela.



Slika 2.3 Topologija HarmonicGAN modela [9]. Ilustriran je način računanja gubitka sličnosti (na slici "Smoothness Constraint") dijelova slika  $i$  i  $j$  za translahirane slike  $G(x)$  i  $F(y)$ , te za rekonstruirane slike  $F(G(x))$  i  $G(F(y))$ . "Forward Cycle" označava translahiranje i rekonstrukciju slika domene  $X$ , dok "Backward Cycle" označava translahiranje i rekonstrukciju slika domene  $Y$

Uz navedene funkcije gubitka, koristi se još i gubitak identiteta (engl. *identity loss*). Gubitak identiteta tjera generatore modela da ne mijenjaju sliku ako je slika već slična ciljanoj domeni generatora [14]. Navedeni gubitak identiteta koristi se u

implementaciji CycleGAN i HarmonicGAN modela u ovom radu.

Za funkciju računanja vektora značajki u HarmonicGAN istraživanju korišteno je: (1) niskorazinski "soft" histogrami i (2) predtrenirane značajke konvolucijske neuronske mreže [9]. U istom, pokazano je da su oba pristupa valjana, te postižu slične rezultate. U ovom radu će se koristiti prvi pristup zbog lakoće implementacije.

Korištena metoda računanja "soft" histograma je predložena u Wang et al. [15]. Jedina promjena je što histogram ima konstantne sredine i širine "bin"-ova. Funkcija za računanje histograma se sastoji od linearnih funkcija  $\psi_b$ ,  $b = 1, 2, \dots, B$ , gdje je  $B$  broj "bin"-ova u histogramu. Funkcija  $\psi_b$  predstavlja doprinos pixela  $p$  za  $b$ -ti "bin" i definirana je ovako:

$$\psi_b(x_p(i)) = \max\{0, 1 - \frac{|x_p(i) - \mu_b|}{w_b}\}, \quad (2.8)$$

gdje su  $\mu_b$  i  $w_b$  sredina i širina  $b$ -tog "bin"-a, a  $x_p$  predstavlja  $p$ -ti piksel dijela slike  $x(i)$ .

Linearnom kombinacijom  $\psi_b$  za sve piksele  $x(i)$  dobivamo:

$$\phi_h(X, i, b) = \phi_h(x(i), b) = \sum_p \psi_b(x_p(i)), \quad (2.9)$$

gdje je  $\phi_h$  značajka RGB histograma, a  $b$  indeks dimenzije reprezentacije RGB histograma. "Soft" histogram reprezentacija  $\phi_h(X, i)$  dijela slike  $x(i)$  je  $B$ -dimenzionalni vektor [9].

## 2.2 Difuzijski modeli

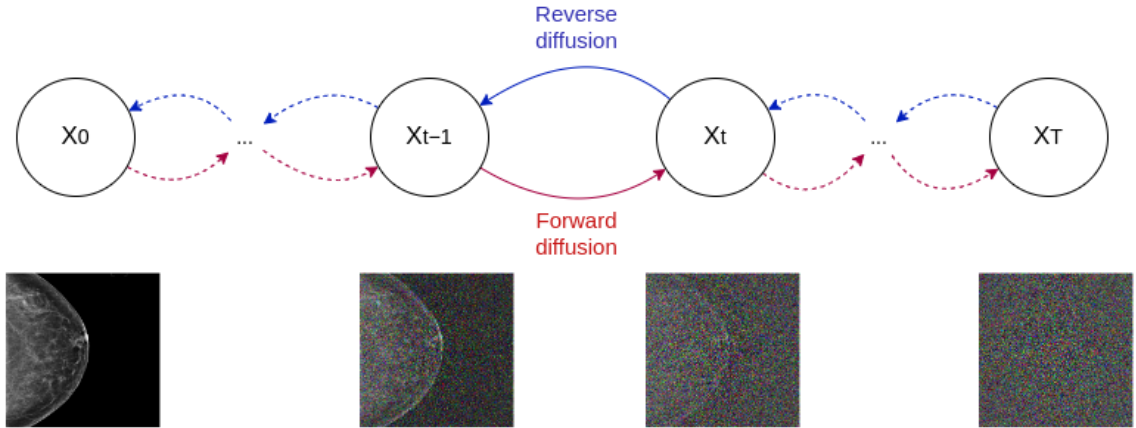
Trenutačni najsvremeniji generativni modeli slika su difuzijski modeli. Iako GAN modeli generiraju visokokvalitetne uzorke, pokazalo se da su ti uzorci manje raznoliki u usporedbi s modelima temeljenim na vjerojatnosti [16]. Difuzijski modeli mogu generirati slike visoke kvalitete i istovremeno bolje modeliraju distribuciju raznolikih podataka [10].

Difuzijski modeli stvaraju nove uzorke pomoću difuzijskog procesa. Difuzijski proces sastoji se od dva dijela: proces difuzije prema naprijed (engl. *forward diffusion*



## Poglavlje 2. Modeli

*process*) i proces obrnute difuzije (engl. *reverse diffusion process*). Vizualni prikaz oba procesa može se vidjeti na slici 2.4. Proces difuzije prema naprijed se koristi za treniranje modela, dok se proces obrnute difuzije koristi za generiranje novih uzoraka.



Slika 2.4 Prikaz procesa difuzije prema naprijed i procesa obrnute difuzije [17]. Prilikom procesa difuzije prema naprijed, slici se postepeno dodaje šum. Prilikom procesa obrnute difuzije, šum se postepeno uklanja sa slike da bi se generirali novi uzorci.

Tijekom procesa difuzije prema naprijed, slici se dodaje Gaussov šum. Količina Gaussovog šuma ovisi o parametrima  $\beta_t$  koji definiraju količinu dodanog šuma za svaki korak  $t \leq T$ , gdje je  $T$  maksimalan broj koraka dodavanja i uklanjanja šuma. U jednadžbama se često koristi i  $\alpha_t$ , gdje je  $\alpha_t = 1 - \beta_t$ . Proces difuzije prema naprijed definiran je na sljedeći način:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad (2.10)$$

gdje je  $x_0$  originalna slika,  $x_t$  zašumljena slika u koraku  $t$ , a  $\epsilon$  je šum koji se dodaje slici ( $\epsilon \sim \mathcal{N}(0, I)$ ). Parametar  $\bar{\alpha}_t$  kontrolira količinu dodanog šuma i dobiven je kumulativnim produktom:

$$\bar{\alpha}_t = \prod_{s=1}^t \alpha_s. \quad (2.11)$$

## Poglavlje 2. Modeli

$\bar{\alpha}_t$  se koristi za direktno uzorkovanje  $x_t$  iz  $x_0$  bez potrebe za  $t$  koraka računanja.

Tijekom treniranja model uči predvidjeti šum  $\epsilon$  sa zašumljene slike  $x_t$ , pa je funkcija gubitka:

$$\mathcal{L}_{DDPM} = \mathbb{E}_{t, x_0, \epsilon} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2] \quad (2.12)$$

gdje je  $\epsilon$  stvarni šum, a  $\epsilon_\theta$  predviđeni šum za zašumljenu sliku  $x_t$  i korak  $t$  [18].

Proces obrnute difuzije počinje sa slikom čistog šuma  $x_T$ , te se tijekom procesa šum postepeno uklanja sa slike kako bi se generirao novi uzorak. Generiranje uzoraka originalnog DDPM (engl. *Denoising Diffusion Probabilistic Models*) [18] istraživanja je definirano kao obrnuti Markovljev difuzijski proces, što proces generiranja čini izrazito sporim. Naime, potrebno je uklanjati šum za svih  $T$  koraka. Kako bi se ubrzao proces obrnute difuzije, koristi se metoda predložena od strane Song et al. Metoda redefinira obrnuti proces difuzije, tako da isti više nije Markovljev difuzijski proces, čime se može smanjiti broj koraka uklanjanja šuma i samim time ubrzati proces generiranja [19]. Proces obrnute difuzije istraživanja [19] je definiran ovako:

$$x_{t-1} = \underbrace{\sqrt{\bar{\alpha}_{t-1}} \left( \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \right)}_{\text{"predviđeni } x_0"} + \underbrace{\sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \epsilon_\theta(x_t, t)}_{\text{"smjer prema } x_t"} + \underbrace{\sigma_t \epsilon_t}_{\text{nasumični šum}}, \quad (2.13)$$

gdje je  $\epsilon_t \sim \mathcal{N}(0, I)$  Gaussov šum, a  $t$  korak procesa obrnute difuzije.  $\sigma_t$  je definirana kao:

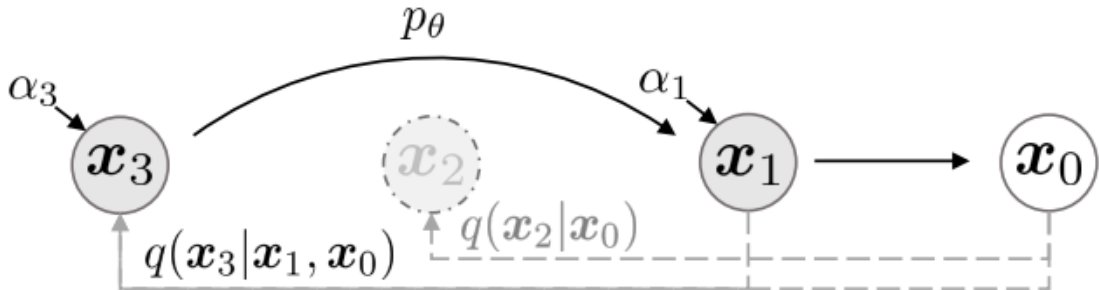
$$\sigma_t = \eta \sqrt{\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}} \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}}, \quad (2.14)$$

gdje je  $\eta$  hiperparametar koji kontrolira količinu Gaussovog šuma  $\epsilon_t$  dodanog tijekom generiranja uzoraka. U slučaju kada je  $\eta = 0$ , generiranje uzoraka postaje deterministički proces [19].

Proces obrnute difuzije definiran jednadžbom 2.13 ovisi o predviđenoj izvornoj slici  $x_0$ , te o zašumljenoj slici  $x_t$ . Ta činjenica omogućava smanjenje broja koraka obrnute difuzije. Koristeći smanjeni broj koraka  $\tau$ , gdje je  $\tau$  povećavajući podslijed

$[1, \dots, T]$  duljine  $S$ , u jednadžbi 2.13, uklanjanje šuma se provodi samo  $S$  puta. Za dovoljno mali  $S$  brzina generiranja se znatno povećava [19].

Intuitivno, zato što proces obrnute difuzije više nije Markovljev proces, moguće je "preskakati" korake tijekom njegovog računanja. To znači da se difuzijski modeli mogu trenirati na velikom broju koraka  $T$ , ali da se tijekom procesa generiranja računaju samo određeni koraci. Time se dobiva na brzini generiranja. Ilustracija navedenog je prikazana na slici 2.5. Broj smanjenih koraka  $S$  ne smije biti premalen, jer to rezultira generiranjem zašumljenih slika zbog nedovoljnog broja koraka uklanjanja šuma.



Slika 2.5 Primjer ubrzanog generiranja difuzijskog modela, gdje  $\tau = [1, 3]$ , a  $t = [1, 2, 3]$  [19]. Korištenjem metode iz [19], tijekom procesa obrnute difuzije korak 2 se preskače radi ubrzanja.

### 2.2.1 CycleDiffusion

CycleDiffusion je metoda za neuparenu translaciju slika koja zahtijeva korištenje dva stohastička difuzijska modela (hiperparametar  $\eta > 0$ ). Jedan model je treniran na izvornoj domeni, dok je drugi treniran na ciljanoj domeni. Koristeći DPM-Enkoder (engl. *Denoising Probabilistic Models - Encoder*) definiran u istom, slike izvorne domene se enkodiraju koristeći pripadajući difuzijski model, te se zatim dekodiraju u ciljanoj domenu pomoću drugog difuzijskog modela [11]. U slučaju ovog rada, slike s gipsom se enkodiraju pomoću difuzijskog modela treniranom na slikama s gipsom, te zatim dekodiraju u slike bez gipsa koristeći difuzijski model treniran na slikama

## Poglavlje 2. Modeli

bez gipsa. DMP-Enkoder enkodira sliku  $x_0$  u  $z$  ovim izrazom:

$$z := (x_{T_{ENC}} \oplus \epsilon_{T_{ENC}} \oplus \dots \oplus \epsilon_2 \oplus \epsilon_1) \quad (2.15)$$

gdje je  $T_{ENC}$  broj koraka enkodiranja,  $x_{T_{ENC}}$  slika zašumljena za korak  $T_{ENC}$  izvornim difuzijskim modelom, a  $\epsilon_t$  se dobiva izrazom:

$$\epsilon_t = \frac{(x_{t-1} - \mu(x_t, t))}{\sigma_t}, \quad (2.16)$$

za  $t = T_{ENC}, \dots, 1$ , gdje je  $\mu(x_t, t)$  estimator dobiven jednadžbom 2.13, a  $\sigma_t$  dobiven jednadžbom 2.14 [11].

Generiranje translahiranih slika se odvija u dva dijela. Tijekom prvog dijela koristi se jednadžba 2.13, ali umjesto Gaussovog šuma, koristi se  $\epsilon_t$  dobiven enkodiranjem. Izračun se ponavlja za svaki broj koraka enkodiranja, tako da se prvi dio generiranja CycleDiffusion metode može smatrati dekodiranjem  $z$ . Drugi dio generiranja uključuje običan proces obrnute difuzije definiran jednadžbom 2.13, koji se odvija  $T - T_{ENC}$  koraka.

Intuitivno, broj koraka enkodiranja  $T_{ENC}$  je hiperparametar koji utječe na sličnost originalne slike s translahiranom. U slučaju uklanjanja gipseva, potrebno je držati  $T_{ENC}$  relativno malenim, tako da struktura kostiju i značajke kostiju (npr. frakture) ostanu iste na translahiranoj slici.

# Poglavlje 3

## Podaci

Ovo poglavlje posvećeno je analizi korištenog skupa podataka. Prikazan je njegov sastav i relevantne statistike, te se detaljno opisuje metoda uravnotežavanja regija unutar istog.

### 3.1 Skup podataka

Korišteni skup podataka ustupljen je strane Medicinskog fakulteta u Grazu i označen od strane radiologa. Sastoji se od 93 064 rendgenskih snimki kostiju različitih dijelova tijela. Od toga, 50% snimki prikazuje kost unutar gipsa, dok preostalih 50% prikazuje kost bez gipsa. Sažetak statistike podataka prikazan je u tablici 3.1.

Tablica 3.1 Sažetak statistike skupa podataka. Oznaka AP označava anteroposteriornu projekciju.

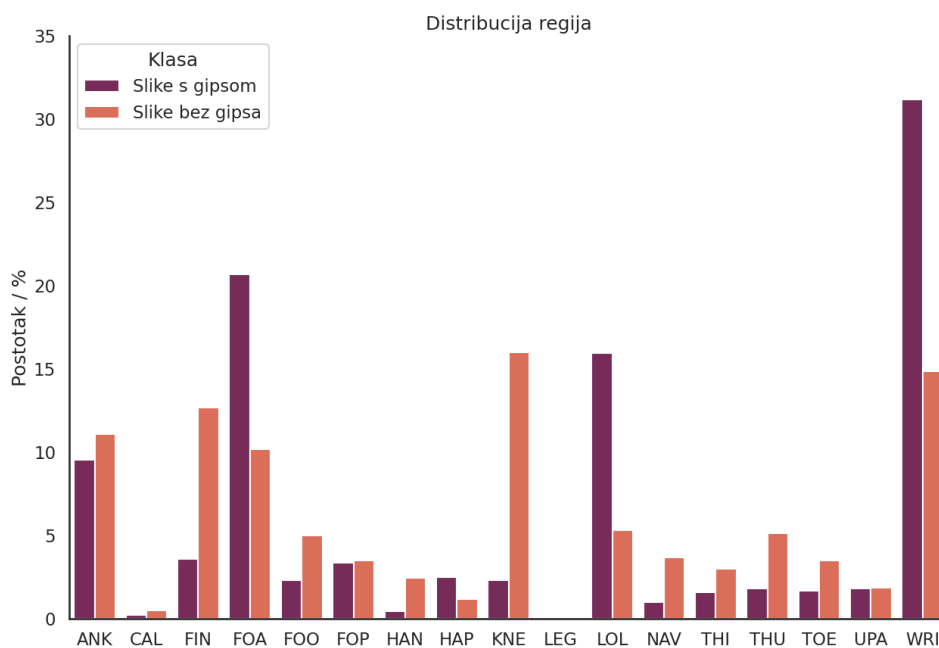
Svojstvo	Vrijednost			
Gips	Prisutan: 46 532	Odsutan: 46 532	\	\
Spol	Muški: 51 995	Ženski: 41 022	Nepoznato: 47	\
Dob	Muški: $14.17 \pm 13.42$	Ženski: $20.60 \pm 23.45$	Nepoznato: $10.74 \pm 4.80$	Svi: $17.00 \pm 18.79$
Lateralnost	Lijeva: 26 236	Desna: 24 720	Nepoznato: 42 108	\
Projekcija	AP: 41 715	Lateralna: 35 300	Kosa: 1 392	Nepoznato: 14 657

Uz statistike navedene u tablici 3.1, rendgenske snimke u skupu podataka se mogu podijeliti na 17 različitih regija. Regija rendgenske snimke je jako bitna zbog uparivanja podataka. Iako CycleGAN i HarmonicGAN mreže koriste nesparene parove

### Poglavlje 3. Podaci

slika za translaciju, u ovom slučaju je važno da svi parovi slika pripadaju istoj regiji. Primjerice, regija zapešća s gipsom se mora translirati u regiju zapešća bez gipsa. Korištenje parova slika iz različitih regija otežava trening modela jer model ne može znati u koju regiju treba translirati originalnu sliku. Suprotno tome, transliranje parova slika iz istih regija olakšava učenje relevantnih značajki za svaku regiju.

Zbog navedenog, potrebno je detaljno proučiti količine i distribucije regija u skupu podataka. Na slici 3.1 se nalazi distribucija regija skupa podataka, dok se u tablici 3.2 nalaze količine slika pojedinih regija.



Slika 3.1 Distribucija anatomskih regija skupa podataka. Značenja kratica osi x su napisane u tablici 3.2

Tablica 3.2 Količine podataka po regijama.

Regija	Oznaka	Broj slika s gipsom	Broj slika bez gipsa
Gležanj	ANK	4438	5151
Petna kost	CAL	98	234
Prst	FIN	1661	5904
Podlaktica	FOA	9634	4742
Tarzalna kost	FOO	1072	2320
Metatarzalna kost	FOP	1575	1622
Šaka	HAN	204	1148
Karpalna kost	HAP	1156	542
Koljeno	KNE	1073	7445
Noga	LEG	20	35
Potkoljenica	LOL	7423	2487
Metakarpalna kost	NAV	463	1703
Bedro	THI	744	1387
Palac	THU	851	2393
Nožni prst	TOE	772	1632
Nadlaktica	UPA	837	867
Zapešće	WRI	14511	6920

Iz analize navedene tablice i distribucije podataka, uočena je značajna neuravnoteženost u broju slika između različitih regija tijela. Razlog neuravnoteženosti može biti različita frekvencija ozljeda regija tijela ili češća primjena drugačijih terapija za liječenje manje zastupljenih regija skupa. Regije zapešća i podlaktice predstavljaju dominantne dijelove skupa podataka (najviše slika), što može pozitivno utjecati na preciznost modela za ove regije. Nasuprot tome, regije poput petne kosti i noge najmanje su zastupljene, što predstavlja izazov za modeliranje tih regija zbog nedostatka reprezentativnih podataka.

Pored varijacija među regijama, podaci unutar samih regija također su neuravnoteženi iz perspektive prisutnosti gipsa. Najizraženiji primjer neuravnoteženosti je regija koljena i šake. Regija šake je osobito problematična zbog malog broja slika s gipsom. Nasuprot tome, regije poput gležnja, metatarzalne kosti, i nadlaktice pokazuju relativno uravnotežene brojeve slika sa i bez gipsa.

Uravnoteženje klasa izuzetno je važno za smanjenje pristranosti modela i poboljšanje njegovih performansi. Kako bi se postigla optimalna učinkovitost modela, ključno je osigurati da je skup podataka uravnotežen i da su sve regije jednako zastupljene.

## 3.2 Uravnoteženje podataka

Kao što je već spomenuto, GAN mreže korištene u ovom radu se treniraju pomoću nesparenih parova slika, zbog čega je ključno osigurati kvalitetno uparivanje slika bez gipsa sa slikama s gipsom.

Uparivanje se obavlja nasumično, što osigurava da će model tijekom treninga susresti različite parove slika, čime se potiče fokus na dodavanje i uklanjanje gipsa i smanjuje pretreniranost. S obzirom na veliku neujednačenost podataka unutar regija, nije moguće postići idealno uparivanje slika 1:1, a odbacivanje viška podataka ograničava robusnost modela. Ovaj problem je riješen prekomjernim uzorkovanjem slika. Prekomjerno uzorkovanje omogućuje generiranje dodatnih slika manje zastupljenih kategorija unutar regija, čime se poboljšava uravnoteženost skupa podataka. Slike manje kategorije unutar regije se uzorkuju dok uparivanje svih slika nije moguće (algoritam 1).

Algoritam 1 osigurava ravnomjerno uzorkovanje slika. Svaka slika unutar skupa se iskoristi isti broj puta (ovisno o broju kopiranja). Nasumično miješanje skupa je implementirano za slučaj gdje novo kopirani skup slika dodan u ulazne podatke premašuje ciljanu duljinu. U tom slučaju, udio slika kopiranog skupa bit će izrezan. Nasumičnim miješanjem osigurava se da će navedeni udio slika uvijek biti različit, što će kroz epohe treniranja omogućiti ravnomjernu raspodjelu skupa.

Dok korištenje prekomjernog uzorkovanja pomaže u supresiji problema neujed-



---

**Algorithm 1** Prekomjerno uzorkovanje skupa slika

---

**Require:** ulazni\_podaci, ciljana\_duljina

- 1: **while** duljina(ulazni\_podaci) < ciljana\_duljina **do**
  - 2:   kopija\_podataka  $\leftarrow$  Kopija(ulazni\_podaci)
  - 3:   promiješani\_podaci  $\leftarrow$  Promiješaj(kopija\_podataka)
  - 4:   ulazni\_podaci  $\leftarrow$  Konkatencija(ulazni\_podaci, promiješani\_podaci)
  - 5: **end while**
  - 6: ulazni\_podaci  $\leftarrow$  Izreži(ulazni\_podaci, ciljana\_duljina)
  - 7: **return** ulazni\_podaci
- 

načenosti kategorija slika unutar regija, istovremeno stvara novi problem vezan uz varijancu podataka. Dodavanje previše kopija postojećih podataka može rezultirati pretreniranjem, jer se model previše prilagođava najzastupljenijim slikama. Kako bi se izbjegao ovaj problem, potrebno je ograničiti prekomjerno uzorkovanje kako bi se minimizirao gubitak varijance, dok se i dalje postiže uravnoteženost podataka.

Ograničavanje prekomjernog uzorkovanja podataka usklađeno je s drugim problemom uravnotežavanja podataka, što je uravnotežavanje regija. Kao što je već spomenuto, na slici 3.1 i u tablici 3.2 se može vidjeti da postoje velike razlike između količina slika između regija. Potrebno je regije uravnotežiti da bi se smanjila pristranost modela. Rješenje tog problema (korišteno u ovom radu) je ograničavanje količine podataka po regijama po epohi treninga.

Tijekom svake epohe treninga, nasumično se odabire  $n$  parova slika iz svake regije koje će model koristiti za učenje. Korištenjem nasumičnog uparivanja, zajedno s prekomjernim uzorkovanjem podataka koji su manji od  $n$ , uspješno se rješava problem neuravnoteženosti među regijama. Istovremeno se minimizira gubitak varijance podataka jer se osigurava da model uči iz svih dostupnih podataka (iako u različitim epohama).

Broj  $n$  parova varira između dva pristupa treniranja modela korištena u ovom radu. Prvi pristup uključuje ravnomjerno korištenje svih regija tijekom treninga modela, dok se drugi pristup temelji na treniranju modela na najzastupljenijoj regiji skupa podataka: zapešće. U oba pristupa, regije petna kost i noga nisu korištene za trening modela zbog njihove manje zastupljenosti u skupu podataka. Umjesto toga,

### *Poglavlje 3. Podaci*

te regije su rezervirane za testiranje i evaluaciju modela.

S obzirom na sadržaj tablice 3.2, za prvi pristup odabrano je  $n = 500$  za trening modela. Odabir 500 parova slika predstavlja minimalno prekomjerno uzorkovanje i rezultira u ukupno 7 500 parova slika za trening modela. Za drugi pristup, odabrano je  $n = 5\ 000$ , zbog visoke zastupljenosti podataka zapešća.

Broj parova korištenih u validacijskom skupu iznosi 10% broja trening slika, što je  $n = 50$  za prvi pristup i  $n = 500$  za drugi pristup. Tijekom validacije modela nije primijenjeno prekomjerno uzorkovanje. Kao rezultat toga, 50 parova slika korištenih u validaciji modela treniranih na svim regijama rezultira s ukupno 716 slika u validacijskom skupu, umjesto očekivanih 750.

Navedeni postupak uravnoteženja podataka primijenjen je na GAN mreže, što predstavlja veći izazov u usporedbi s uravnoteženjem skupa kod difuzijskih modela. Kao što je već spomenuto u uvodu, CycleDiffusion koristi dva različita difuzijska modela za translaciju slike u sliku, što znači da uparivanje slika nije potrebno. U ovom slučaju, jedan difuzijski model treniran je na slikama s gipsom, dok je drugi treniran na slikama bez gipsa.

Isto kao i za GAN mreže, dva pristupa korištena su prilikom treniranja modela: treniranje na svim regijama i treniranje samo na regiji zapešća. Radi vjerodostojnosti usporedbe, odabran je isti način uravnoteženja regija kao i kod GAN mreža (uravnotežavanje kategorija slika unutar regija nije potrebno jer se slike ne trebaju uparivati). Kao i za GAN mreže, odabran je  $n = 500$  za prvi pristup i  $n = 5000$  za drugi pristup, pri čemu  $n$  odgovara broju slika po regiji, a ne broju parova slika. Broj slika korištenih u validacijskom skupu također je 10% trening skupa. Prekomjerno uzorkovanje korišteno je tijekom treninga, ali ne i tijekom validacije.

U tablici 3.3 može se vidjeti pregled konfiguracije treninga i validacije za svaki model i pristup.

Tablica 3.3 Pregled konfiguracije treninga i validacije.

Model	Pristup	Trening (n/regija)	Validacija (n/regija)	Trening (sve)	Validacija (sve)
GAN	Sve regije	500 parova	50 parova	7500 parova	716 parova
	Samo zapešće	5000 parova	500 parova	5000 parova	500 parova
CycleDiffusion	Sve regije	500 slika	50 slika	7500 slika	716 slika
	Samo zapešće	5000 slika	500 slika	5000 slika	500 slika

# Poglavlje 4

## Postavljanje eksperimenta

U ovom poglavlju je detaljno opisan postavljeni eksperiment, što uključuje arhitekture korištenih modela, te postavke hiperparametara. Također, definiraju se metrike evaluacije koje se koriste za procjenu eksperimenta.

### 4.1 Eksperiment

U radu su korištena dva pristupa treniranja modela: pristup gdje je model treniran na svim regijama i pristup gdje je model treniran samo na regiji zapešća. Kod testiranja modela korišten je isti testni skup za svaki model. Testni skup se sastoji od sveukupno 2965 slika s prisutnim gipsom s kojim se modeli nisu prije susreli. Slike su veličine  $128 \times 128$  zbog složenosti i resursa potrebnih za treniranje modela na većim slikama, što značajno povećava zahtjeve za računalnom memorijom. Cilj eksperimenta je evaluirati mogućnost uklanjanja gipsa modela za sve regije i za svaku pojedinu regiju.

Arhitektura generatora korištena u CycleGAN modelu temelji se na ResNet [20] arhitekturi sa 64 značajke (engl. *features*). Sastoji se od dva sloja za smanjivanje rezolucije (engl. *downsampling layer*), devet rezidualnih blokova (engl. *residual blocks*), te dva sloja za povećavanje rezolucije (engl. *upsampling layer*). Za regularizaciju se koristila metoda nasumičnog izostavljanja neurona (engl. *dropout*) vrijednosti 0.5. Hiperparametar funkcije gubitka ciklusa je  $\lambda_{cyc} = 10$  kao i u original-

## Poglavlje 4. Postavljanje eksperimenta

nom istraživanju [4]. Koristi se L1 udaljenost za računanje funkcije gubitka ciklusa i gubitka identiteta.

Arhitektura generatora HarmonicGAN modela je ista kao i u CycleGAN modelu. Kao i kod CycleGAN modela, koristi L1 udaljenost za funkciju gubitka ciklusa i funkciju gubitka identiteta. Tijekom računanja gubitka glatkoće, slika se podijeli na četiri dijela veličine 64 x 64, gdje hiperparametri iznose:  $\sigma = 0.22$  i  $\lambda_{glatko.} = 1$ . Kao funkciju udaljenosti ( $\text{Dist}[\vec{x}(i), \vec{x}(j)]$ ) koristi se L1 udaljenost normaliziranih "soft" histograma kao i u originalnom istraživanju [9]. Hiperparametar  $\sigma$  je dobiven pretragom, dok je vrijednost  $\lambda_{glatko.}$  uzeta iz originalnog istraživanja [9].

Za diskriminatore oba modela korištena je PatchGAN arhitektura [12] sa 64 značajke i pet konvolucijskih blokova, od kojih je prvi ulazni blok, a zadnji izlazni blok kojemu je izlaz jedan neuron. PatchGAN klasificira lažne i stvarne podatke pomoću preklapajućih 70 x 70 piksela slike. Funkcija gubitka korištena za treniranje diskriminatora je MSE gubitak (engl. *Mean Squared Error loss*).

Radi učinkovitijeg treniranja diskriminatora, koristio se spremnik slika (engl. *image pool*) definiran u istraživanju Shrivastava et al. [21]. Spremnik slika koristi se za pohranu slika generiranih u prethodnim iteracijama treninga. Diskriminator se potom trenira na slikama iz spremnika umjesto na najnovije generiranim slikama, što pomaže u smanjenju oscilacija tijekom treninga [4]. Korištena veličina spremnika slika je 50 za obje domene slika, s 50% vjerojatnošću da se slika izabere iz spremnika.

Za oba GAN modela korišten je ADAMW [22] optimizator sa stopom učenja  $\alpha = 2 \times 10^{-4}$  i hiperparametrima  $\beta_1 = 0.05$  i  $\beta_2 = 0.999$ . GAN modeli su trenirani kroz 200 epoha, pri čemu je stopa učenja prvih 100 epoha bila konstantna, dok se tijekom preostalih 100 epoha postepeno smanjivala do nule.

Arhitektura difuzijskog modela je temeljena na UNet [23] arhitekturi s četiri bloka za smanjivanje i povećavanje rezolucije, te srednjeg bloka. Blok za smanjivanje rezolucije se sastojao od dva rezidualna bloka i sloja za smanjivanje rezolucije. Blok za povećavanje rezolucije je isti, samo što se na kraju bloka nalazi sloj za povećavanje rezolucije. Blok pažnje (engl. *attention block*) je korišten samo u "najnižim" blokovima za povećavanje i smanjivanje rezolucije, te se ti slojevi sastoje od dva rezidualna bloka i bloka pažnje. Srednji blok se sastoji od rezidualnog bloka, bloka pažnje i

## Poglavlje 4. Postavljanje eksperimenta

još jednog rezidualnog bloka. Dodatno, koristila metoda nasumičnog izostavljanja neurona od 0.1.

CycleDiffusion metodi su potrebna dva difuzijska modela (jedan treniran na skupu podataka bez gipsa i drugi treniran na skupu podataka s gipsom), što znači da je bilo potrebno trenirati četiri modela zbog dva pristupa koja će biti ispitana. Svi modeli su koristili istu već navedenu arhitekturu difuzijskih modela. Koristilo se  $T = 1000$  vremenskih koraka za treniranje s parametrima  $\beta_t$  ravnomjerno raspoređenim u rasponu  $[0.0001, 0.02]$ . Slično kao i za GAN modele korišten je ADAMW optimizator s hiperparametrima  $\beta_1 = 0.05$  i  $\beta_2 = 0.999$  i stopom učenja  $\alpha = 10^{-5}$ .

Prilikom generacije slika broj koraka uklanjanja šuma je  $T = 500$ . Idealno, hiperparametar  $\eta$  bi trebao biti jednak nuli, tako da je proces translacije deterministički proces, ali CycleDiffusion metoda dopušta samo stohastičke difuzijske modele. Stoga, prema rezultatima rada [19], odabrana vrijednost navedenog hiperparametra je  $\eta = 0.1$ . Korišten je mali broj koraka enkodiranja  $T_{ENC} = 50$  i  $T_{ENC} = 100$ , tako da se što više značajki originalne slike zadrži u transliranoj, bez da se utječe na sposobnost uklanjanja gipseva. Implementacija CycleDiffusion metode ponekad generira zašumljene uzorke, što se rješava procesom prerađivanja nakon što je slika već generirana [11]. Proces prerađivanja uključuje dodatnih  $T_{SDEDIT}$  koraka uklanjanja šuma sa slike. Eksperimentalno je utvrđeno da broj koraka  $T_{SDEDIT} = 20$  dovoljno koraka da generirane slike ne budu zašumljene.

## 4.2 Mjerenje performansi

Kao što je spomenuto u uvodu i prikazano na slici 1.1b, postoji znatna razlika između slika bez gipsa i slika s gipsom. Zbog usporedbe i sličnosti problema, u ovom radu se koriste usporedbe histograma slika izračunatih na testnom skupu isto kao i u [3]. Uspoređuje se prosječni histogram stvarnih slika bez gipsa s prosječnim histogramom generiranih slika bez gipsa.

Prvi način usporedbe je korištenje korelacije između histograma. Korelacija dva histograma je definirana kao:

Poglavlje 4. Postavljanje eksperimenta

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}}, \quad (4.1)$$

gdje je

$$\bar{H}_k = \frac{1}{N} \sum_J H_k(J), \quad (4.2)$$

i  $N$  broj "bin"-ova histograma [24]. Korelacija je skalarna vrijednost između  $-1$  i  $1$ , gdje  $-1$  označava savršenu negativnu korelaciju, a  $1$  označava savršenu pozitivnu korelaciju. Idealno, model bi trebao ostvariti savršenu pozitivnu korelaciju.

Drugi način usporedbe je koristeći hi-kvadrat (engl. *chi-squared*) udaljenost. Manje vrijednosti predstavljaju sličnije histograme i time bolji model. Hi-kvadrat udaljenost je definirana ovako:

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}. \quad (4.3)$$

Treći način je presjek histograma, gdje je veća vrijednost bolja. Presjek histograma mjeri količinu zajedničkog intenziteta između dva histograma i definiran je ovako:

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I)). \quad (4.4)$$

Posljednji način usporedbe je Hellinger (Bhattacharyya) udaljenost. Hellinger udaljenost mjeri sličnost dvije distribucije, gdje je manja udaljenost bolja i definirana je ovako:

$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H}_1 \bar{H}_2 N^2}} \sum_I \sqrt{H_1(I) \cdot H_2(I)}}. \quad (4.5)$$

IS (engl. *Inception Score*) [25] je često korištena metrika za automatsku evaluaciju generativnih modela. Problem IS-a, i razlog zašto se neće koristiti u ovom radu, je što IS metrika nije pouzdana za skupove podataka koji su značajno različiti od "ImageNet" skupa (na kojem je model koji računa IS treniran) [26].

#### *Poglavlje 4. Postavljanje eksperimenta*

Umjesto IS-a, koristit će se FID (engl. *Fréchet Inception Distance*) metrika [27], koja se isto često koristi za ocjenu performansi generativnih modela. Kao i IS, FID isto koristi model treniran na "ImageNet" skupu slika, ali za razliku od IS metrike, FID uzima u obzir i stvarne slike i generirane slike, što ga čini prikladnim za korištenje za korišteni skup podataka. FID metrika uspoređuje distribuciju stvarnih podataka s distribucijom generiranih podataka, di manja vrijednost označava sličnije distribucije.



# Poglavlje 5

## Rezultati

Prilikom prikaza rezultata u tablicama koriste se kratice radi preglednosti. "C. GAN" označava CycleGAN, a "H. GAN" HarmonicGAN. CycleDiffusion metoda je označena DDPM kraticom zato što su metodi korišteni DDPM modeli. Oznaku "SVI" imaju modeli koji su trenirani na svim regijama, dok oznaku "WRI" imaju modeli trenirani samo na zapešću. Brojevi 50 i 100 kod DDPM modela označavaju broj koraka enkodiranja  $T_{ENC}$ .

U tablici 5.1 prikazane su metrike izračunate za sve regije. U tablici 5.2 prikazane su metrike izračunate isključivo za regiju zapešća. Tablica 5.3 prikazuje prosječne vrijednosti metrika izračunate zasebno za svaku regiju. Razlika između metrika tablice 5.1 i tablice 5.3 leži u tome što se korištene metrike računaju uspoređujući distribucije generiranih i stvarnih slika (detaljnije opisano u 4.2). To znači da postoji razlika ako se uspoređuju slike regije zapešća ili ako se sve regije uspoređuju zajedno. Stoga, tablica 5.1 prikazuje performanse modela za cijelu distribuciju testnih podataka, dok tablica 5.3 prikazuje prosječnu performansu modela za sve regije zasebno, pružajući uvid u prosjek individualnih performansi svake regije.

Tablica 5.1 Metrike modela računata na svim regijama.

Model	FID	Korelacija	Presjek	Hellinger	Hi-kvadrat
C. GAN SVI	52.972	0.685	27.957	0.502	122.959
C. GAN WRI	55.018	0.997	49.965	0.288	31.265
DDPM SVI 50	28.958	0.996	47.379	0.283	134.907
DDPM SVI 100	<b>28.929</b>	0.992	43.785	0.324	117.172
DDPM WRI 50	32.586	0.456	21.809	0.642	27823.1
DDPM WRI 100	32.847	<b>0.999</b>	<b>51.336</b>	<b>0.200</b>	<b>16.404</b>
H. GAN SVI	47.713	0.896	24.035	0.623	274.021
H. GAN WRI	53.097	0.791	34.699	0.488	7917.57

Tablica 5.2 Metrike modela računate isključivo na regiji zapješća.

Model	FID	Korelacija	Presjek	Hellinger	Hi-kvadrat
C. GAN SVI	90.114	0.99755	49.559	0.265	28.239
C. GAN WRI	78.4	0.99061	43.18	0.318	32.435
DDPM SVI 50	56.167	0.9982	49.219	0.243	32.244
DDPM SVI 100	58.739	0.99116	43.855	0.354	143.13
DDPM WRI 50	<b>52.474</b>	<b>0.99833</b>	<b>52.492</b>	<b>0.214</b>	<b>113.52</b>
DDPM WRI 100	52.825	0.99811	51.531	0.226	32.385
H. GAN SVI	85.226	0.89026	26.773	0.596	661.508
H. GAN WRI	69.143	0.96926	37.715	0.455	99.595

Tablica 5.3 Prosječne metrike modela izračunate zasebno za svaku regiju.

Model	FID	Korelacija	Presjek	Hellinger	Hi-kvadrat
C. GAN SVI	141.59 ± 41.11	<b>0.969 ± 0.077</b>	44.295 ± 8.381	0.360 ± 0.107	146.1 ± 283.8
C. GAN WRI	145.34 ± 42.79	0.930 ± 0.114	42.349 ± 9.486	0.343 ± 0.112	675.9 ± 2188.0
DDPM SVI 50	117.08 ± 45.96	0.933 ± 0.139	<b>45.255 ± 7.512</b>	<b>0.317 ± 0.106</b>	603.0 ± 1648.1
DDPM SVI 100	<b>116.65 ± 46.78</b>	0.857 ± 0.220	41.758 ± 9.795	0.377 ± 0.120	2680.1 ± 8924.1
DDPM WRI 50	124.80 ± 42.51	0.926 ± 0.183	44.089 ± 11.054	0.335 ± 0.150	3286.9 ± 8762.7
DDPM WRI 100	126.33 ± 41.88	0.862 ± 0.290	43.264 ± 11.982	0.349 ± 0.155	2168.2 ± 7624.3
H. GAN SVI	135.76 ± 40.07	0.930 ± 0.136	43.794 ± 9.289	0.357 ± 0.136	7121.5 ± 28004.0
H. GAN WRI	146.81 ± 45.01	<b>0.926 ± 0.153</b>	45.193 ± 8.694	0.329 ± 0.104	<b>142.3 ± 140.2</b>

# Poglavlje 6

## Diskusija

Tablica 5.1 prikazuje metrike modela računane na svim regijama. U kontekstu FID metrika, CycleDiffusion metoda s modelima treniranim na svim regijama pokazuje najbolje performanse, pri čemu generiranje s  $T_{ENC} = 100$  postiže najniži FID od 28.929. Sveukupno, CycleDiffusion metoda s modelima treniranim samo na zapešću s brojem koraka enkodiranja  $T_{ENC} = 100$  pokazuje najbolje rezultate, s najvišom korelacijom od 0.999 i najvišim presjekom od 51.336. Hellingerova i hi-kvadrat udaljenost također ukazuju na bolje performanse istog, s najnižom Hellingerovom vrijednosti od 0.200 i hi-kvadrat udaljenošću od 16.404.

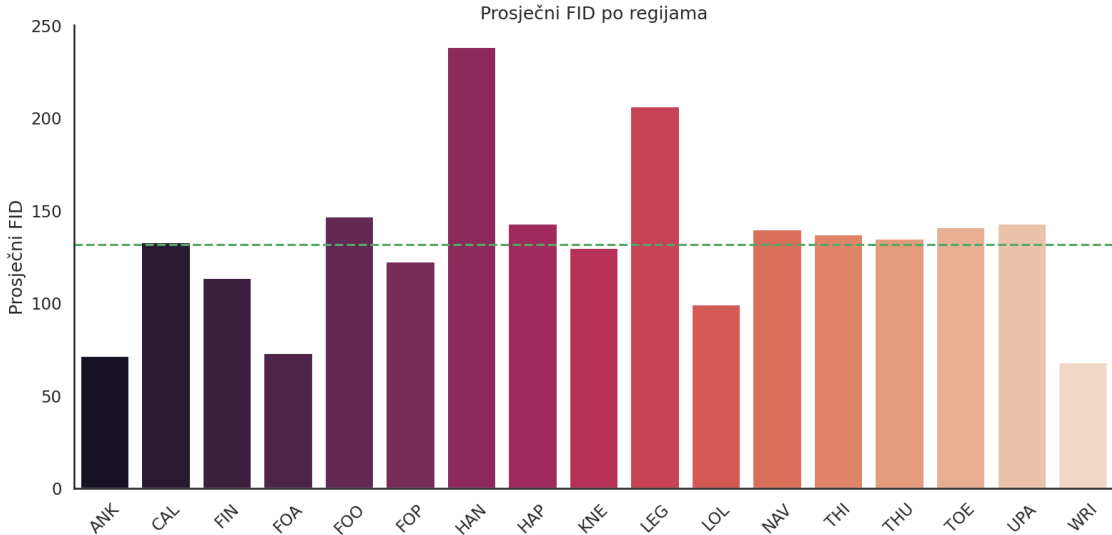
U tablici 5.2 su prikazane metrike modela računane isključivo na regiji zapešća. Najmanju hi-kvadrat udaljenosti ima CycleGAN treniran na svim regijama s vrijednošću od 28.239. Za sve ostale metrike CycleDiffusion metoda s modelima treniranim samo na zapešću i brojem koraka enkodiranja  $T_{ENC} = 50$  postiže najbolje rezultate, s najnižom FID metrikom od 52.474, najvišom korelacijom od 0.99833, najvišim presjekom od 52.492, i najmanjom Hellingerovom udaljenosti od 0.214. Bitno je istaknuti da ista CycleDiffusion metoda s  $T_{ENC} = 100$  koraka enkodiranja blisko prati metodu s  $T_{ENC} = 50$  koraka enkodiranja. Čak je i znatno bolja po hi-kvadrat udaljenosti.

U tablici 5.3 CycleDiffusion metoda ponovno pokazuje konkurentne rezultate, pri čemu je najbolja CycleDiffusion metoda s modelima trenirana na svim regijama.  $T_{ENC} = 50$  postiže najveći presjek histograma i najnižu prosječnu Hellingerovu uda-

## Poglavlje 6. Diskusija

ljenost od 0.317, dok  $T_{ENC} = 100$  postiže najmanji prosječni FID od 116.65. Korelacija varira među modelima, no CycleGAN treniran na svim regijama pokazuje najvišu prosječnu korelaciju od 0.969. Najmanje hi-kvadrat udaljenost imaju CycleGAN treniran na svim regijama i HarmonicGAN treniran na regiji zapešća, gdje je HarmonicGAN bolji s manjom hi-kvadrat udaljenosti u iznosu od 142.3.

Značajna razlika između vrijednosti metrika tablice 5.1 i tablice 5.3, te konzistentno visoka vrijednost standardne devijacije u tablici 5.3, upućuje na to da modeli bolje aproksimiraju cijelu distribuciju podataka nego pojedinačne regije. Ovo je očekivano, s obzirom na to da je polovina modela trenirana na svim regijama. Kako bi se problematične regije bolje procijenile, može se usporediti prosječna vrijednost metrika za svaki model iz tablice 5.3 s prosječnim vrijednostima metrika modela za svaku regiju. Odabrana je FID metrika zbog dosljednosti kroz eksperimente. Na slici 6.1 prikazan je prosječni FID po regijama, gdje zelena isprekidana linija označava prosječni FID modela dobiven iz tablice 5.3.



Slika 6.1 Prikaz prosječne FID metrike po regijama. Zelena isprekidana linija označava prosječni FID modela dobiven iz tablice 5.3.

Primjećuje se da su regije zapešća, gležnja, podlaktice i potkoljenice najbolje modelirane. Ovo je očekivano, s obzirom na to da su regije zapešća, gležnja i podlak-

## Poglavlje 6. Diskusija

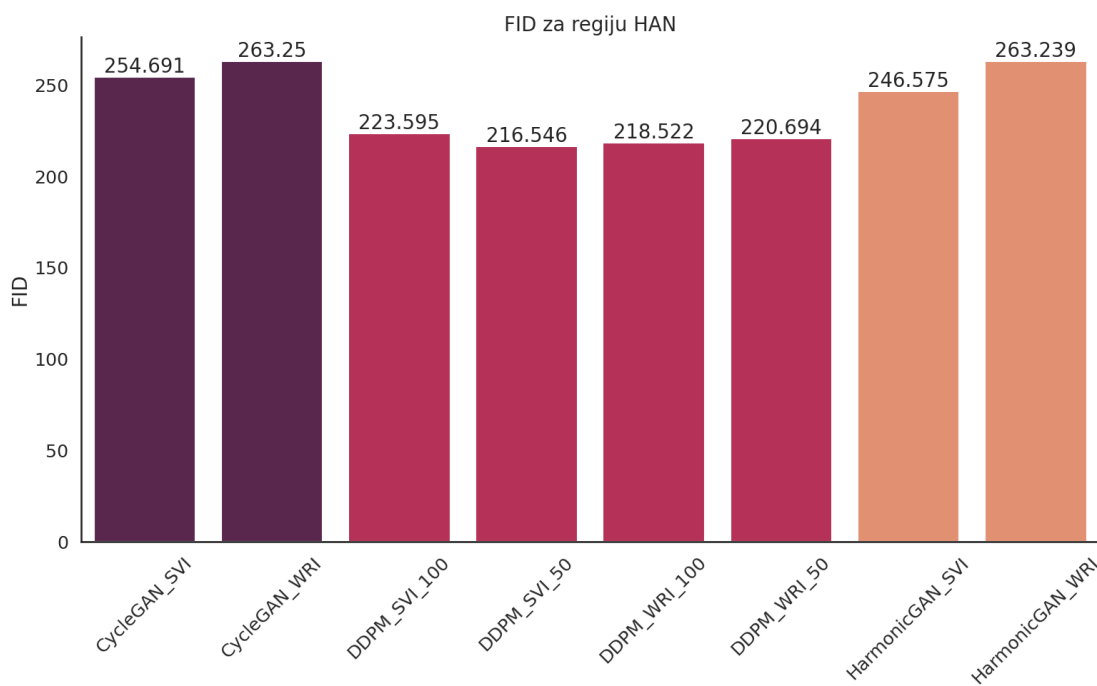
tice najzastupljenije u skupu podataka. Potkoljenica također spada među najbolje modelirane regije, što se može objasniti strukturom slika te regije. Na slici 6.2 prikazan je primjer slike regije potkoljenice, gdje se jasno vidi da, iako je gips prisutan, kosti i obris noge su jasno vidljivi unutar gipsa. Prikazana je bočna projekcija, ali isti princip vrijedi i za ostale projekcije. Osim obilja podataka, važno je napomenuti da struktura slika regija zapešća, gležnja i podlaktice također nalikuje strukturi slika potkoljenice, u smislu jasnog prikaza obrisa i kostiju unutar gipsa, što dodatno pridonosi kvaliteti modeliranja navedenih regija.



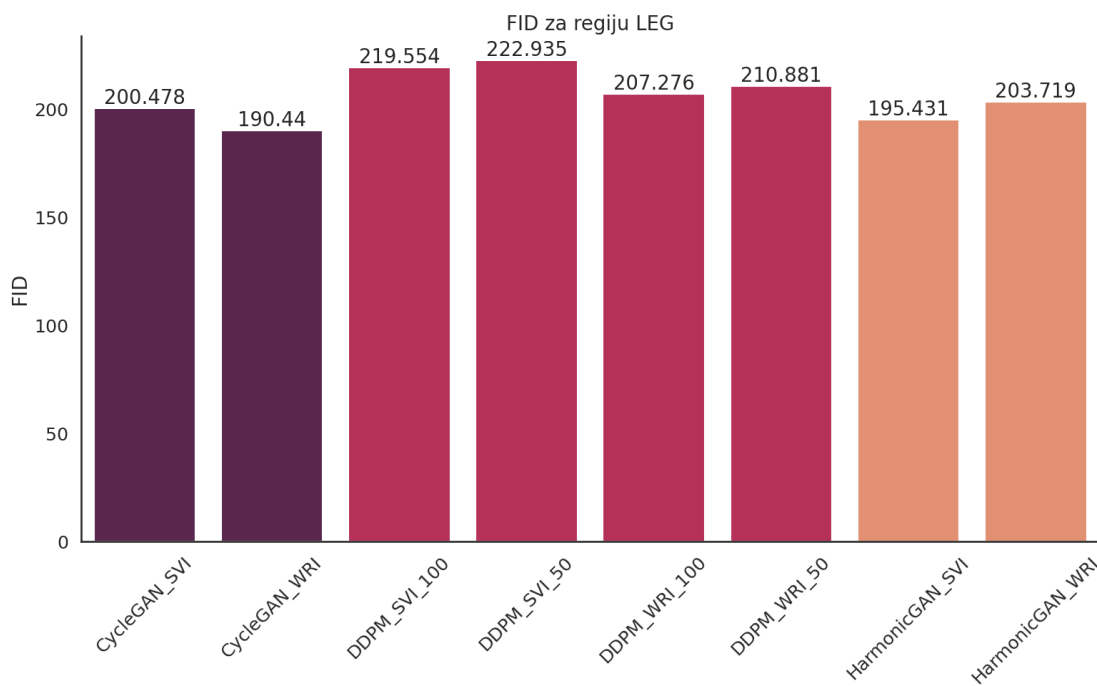
Slika 6.2 Primjer slike regije potkoljenice.

Na slici 6.1, može se vidjeti da je prosječni FID regije šake i noge znatno veći od prosjeka. FID navedenih regija za svaki model je prikazan na slici 6.3.

Poglavlje 6. Diskusija



(a) Prikaz FID metrike za regiju šaka.



(b) Prikaz FID metrike za regiju noga.

Slika 6.3 Prikaz FID metrike za regije šaka i noga.



## Poglavlje 6. Diskusija

Loše modeliranje regija šake i noge je očekivano. Kao što je spomenuto u poglavlju 3, regija šake je vrlo neuravnotežena s obzirom na prisutnost gipsa na slikama u usporedbi sa slikama bez gipsa, što znači da nema dovoljnog broja primjera da model nauči adekvatno modelirati regiju. Također, karakteristike gipsa prisutnog na slikama regije šake znatno se razlikuju od drugih dijelova tijela. Primjer toga može se vidjeti na slici 6.4.

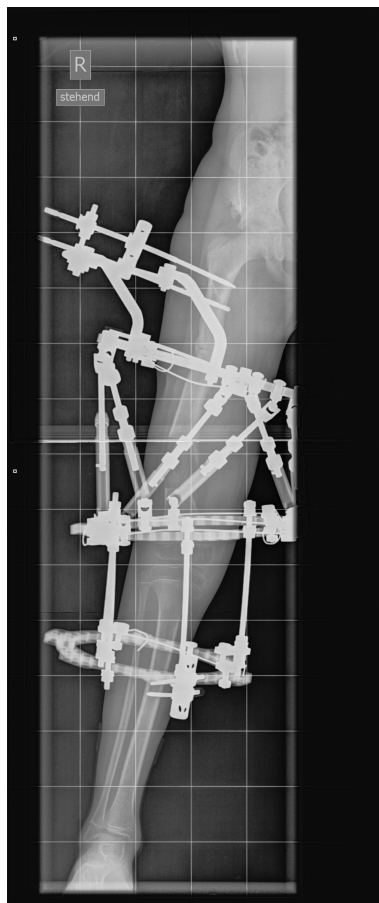


Slika 6.4 Primjer gipsa regije šaka.

Regija noge je jedna od dvije regije koje su isključene iz treninga i validacije modela zbog nedostatka podataka (detaljnije u poglavlju 3). To objašnjava zašto je FID metrika za CycleDiffusion model lošija od FID metrike za GAN modele za tu regiju. GAN modeli bolje modeliraju navedenu regiju zato što GAN modeli uče kako transformirati danu sliku iz izvorne domene u ciljanu domenu. S druge strane, difuzijski modeli generiraju slike iz šuma, te zahtijevaju precizno poznavanje distribucije originalnih podataka kako bi generirali vjerodostojne uzorke. Budući da model nije treniran na podacima regije noge, ne može adekvatno generirati slike te regije. Ova

## Poglavlje 6. Diskusija

činjenica ukazuje na dublji problem u vezi s korištenjem difuzijskih modela u kontekstu generiranja ili translacije medicinskih slika, o čemu će se detaljnije raspravljati kasnije u poglavlju. Osim nedostatka podataka, struktura slika regije noge značajno se razlikuje od ostatka skupa podataka (slika 6.5), što pridonosi lošijoj FID metrici.



Slika 6.5 Primjer slike regije noga.

Druga regija koja je isključena iz treninga modela je regija petne kosti. Na slici 6.1 može se primijetiti da ova regija ima prosječnu vrijednost FID metrike. Razlog prosječnoj vrijednosti može biti taj što se navedena regija sastoji od slika lateralne projekcije koje su slične slikama regije gležnja, dok su slike kose projekcije slike s kojima modeli se modeli nisu susreli. Slike koje su sličnije slikama regije gležnja bolje su modelirane, što smanjuje FID metriku, dok su slike poprečne projekcije

## Poglavlje 6. Diskusija

lošije modelirane, što povećava FID metriku. Primjer iste kosti za obje projekcije može se vidjeti na slici 6.6.



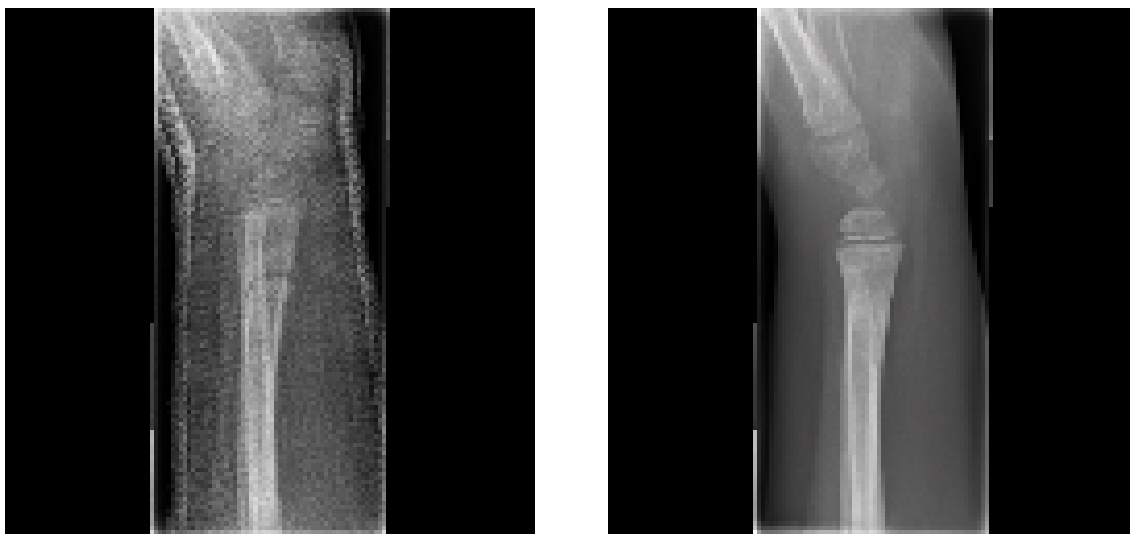
(a) Primjer lateralne projekcije petne kosti.



(b) Primjer kose projekcije petne kosti.

Slika 6.6 Primjer projekcija slika regije petne kosti.

Kao što je ranije spomenuto, difuzijski modeli zahtijevaju dobro poznavanje distribucije originalnih podataka za generiranje novih uzoraka. Budući da generiraju slike iz šuma, model treba "pretpostaviti" što je šum kako bi ga mogao ukloniti i time generirati novi uzorak. Kod translacije slika koristeći CycleDiffusion metodu, količina šuma dodana originalnoj slici utječe na sličnost translirane slike s originalnom. Više šuma omogućava modelu treniranom na ciljanoj domeni veću slobodu u translaciji slike, što može dovesti do "halucinacija" na transliranoj slici. Primjerice, snimka kosti s frakturom i gipsom može se translirati u sliku bez gipsa, ali se u tom procesu izgubi i fraktura. To je zato što je model naučio da distribucija slika bez gipsa uglavnom sadrži snimke kostiju bez fraktura. Primjer ovoga može se vidjeti na slici 6.7.



(a) Primjer kosti s frakturom ispod gipsa.

(b) Translatirana slika koristeći DDPM WRI 100.

Slika 6.7 Primjer uklanjanja fraktura difuzijskog modela. Na podslici (a) se vidi fraktura, koja se tijekom translacije u podsliku (b) uklonila.

Primjer prikazan na slici 6.7 nije ograničen na uklanjanje fraktura. Na slikama s gušćim gipsom, gdje se kosti slabije vide, tijekom procesa obrnute difuzije (gdje difuzijski modeli postepeno uklanjaju šum sa slike) difuzijski modeli prepoznaju taj dio slike kao šum ili kao tkivo. To dovodi do većih "halucinacija," kao što je prikazano na slici 6.8. Iako je model uspješno uklonio gips sa slike zapešća, struktura kosti šake na translatiranoj slici je netočna. Ključni cilj pri ovakvoj translaciji je očuvanje originalne anatomske strukture ispod gipsa, što u ovom slučaju nije postignuto.



(a) Kost zapešća s gipsom.



(b) Slika (a) translahirana koristeći DDPM SVI 50.



(c) Kost zapešća s gipsom.



(d) Slika (b) translahirana koristeći DDPM SVI 50.

Slika 6.8 Primjer mijenjanja strukture kosti difuzijskog modela. U slučaju gornjeg primjera translacije, iako je model dobro uklonio gips, struktura kosti šake je netočna. U slučaju donjeg primjera translacije, model je prepoznao teže vidljive kosti šake stvarne slike (c) kao tkivo, pa translahirana slika na sebi nema kosti šake. Uz to, frakture stvarne slike (c) su se izgubile pri translaciji.

## Poglavlje 6. Diskusija

Slike generirane CycleDiffusion metodom s modelima koji su trenirani isključivo na regiji zapešća sadrže najviše takvih "halucinacija". Razlog tome je što ovi modeli poznaju samo distribuciju slika regije zapešća, pa ne mogu translirati slike drugih regija s istom razinom kvalitete. Primjer ovog može se vidjeti na slici 6.9.



(a) Kost regije potkoljenice s gipsom.



(b) Translatirana slika koristeći DDPM WRI 50.



(c) Translatirana slika koristeći DDPM SVI 50.

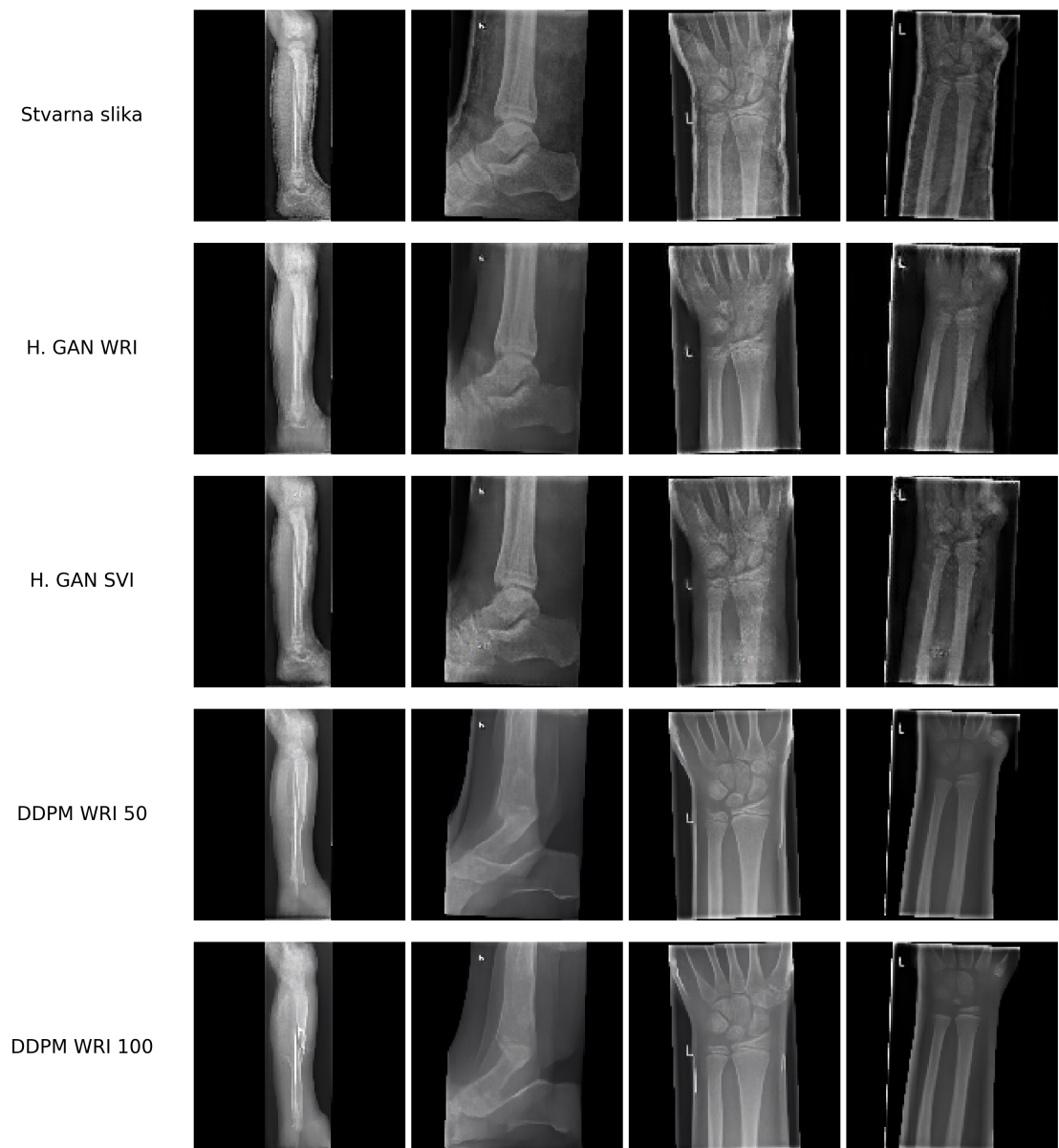
Slika 6.9 Usporedba generiranih slika difuzijskih modela.

## *Poglavlje 6. Diskusija*

Usprkos očitim pogreškama kod transliranja slika difuzijskim modelima, FID metrika je i dalje bolja za sve difuzijske modele. To ukazuje na slabost FID metrike. U kontekstu medicinskih podataka, preciznost na razini detalja slike je izuzetno važna. Stoga, iako difuzijski modeli postižu bolje rezultate prema FID metrici, te metrike mogu prikriti ozbiljne nedostatke u kvaliteti generiranih slika.

U usporedbi s difuzijskim modelima, GAN mreže postižu vizualno bolje rezultate. Translatirane slike nisu toliko "čiste" kao i translirane slike difuzijskih modela, ali kao što je navedeno, difuzijski modeli postižu takve rezultate "pretpostavljanjem" što bi trebalo biti na slici, a ne točnom translacijom slika. Primjeri generiranih slika svih treniranih modela mogu se vidjeti na slikama 6.10 i 6.11.

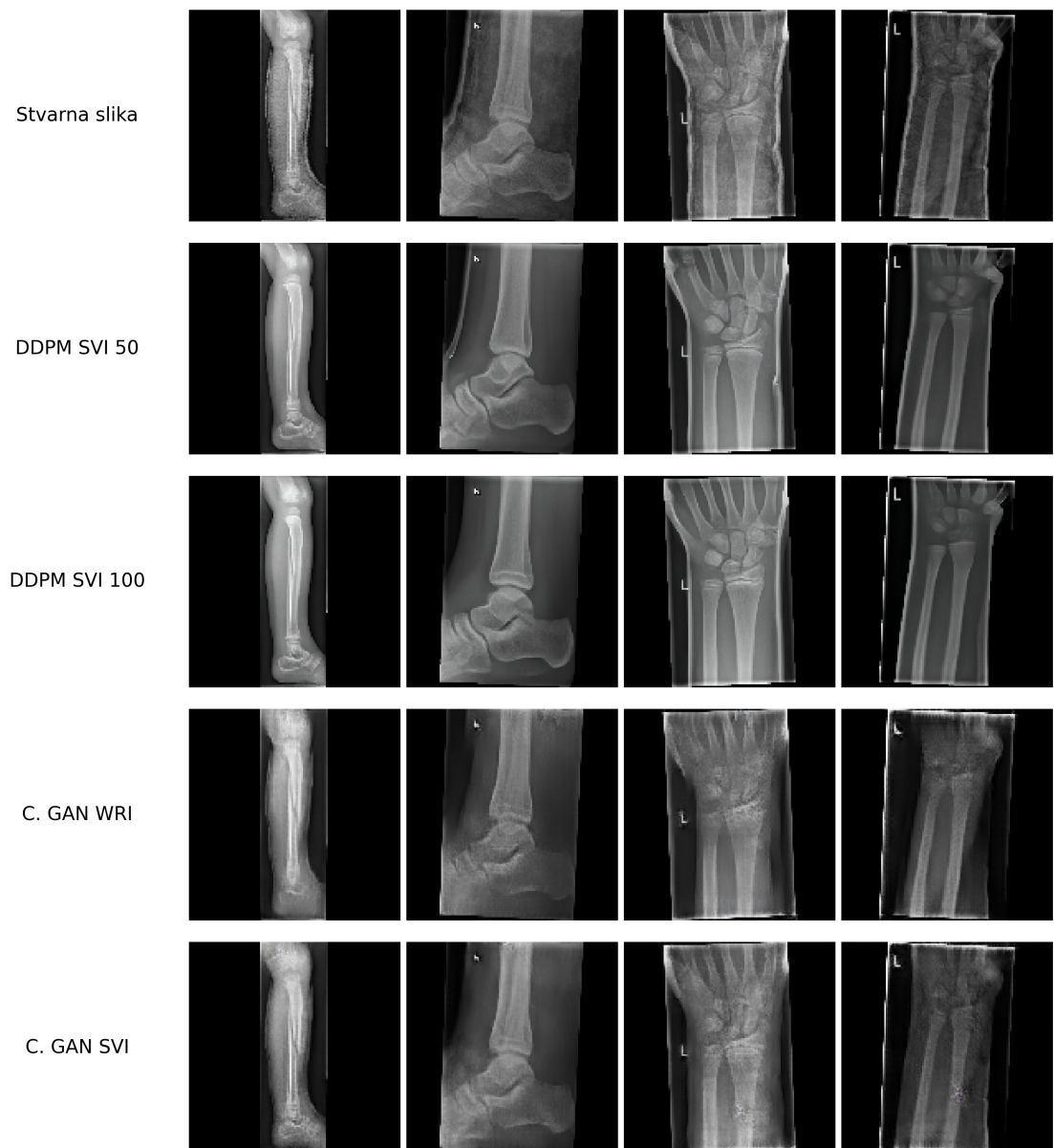
Poglavlje 6. Diskusija



Slika 6.10 Mozaik generiranih slika (prvi dio).



Poglavlje 6. Diskusija



Slika 6.11 Mozaik generiranih slika (drugi dio).

# Poglavlje 7

## Zaključak

U ovom radu implementirane su i testirane tri metode translacije slika u svrhu uklanjanja gipsa s rendgenskih snimki kostiju: CycleGAN, HarmonicGAN modeli te CycleDiffusion metoda.

CycleDiffusion metoda je postigla najbolje rezultate po metrikama, međutim, vizualna inspekcija generiranih slika otkrila je da difuzijski modeli često pate od "halucinacija". Te "halucinacije" uključuju gubitak ili izobličenje anatomskih struktura kostiju, što je nepoželjno u medicinskim primjenama.

S druge strane, GAN modeli su se pokazali superiornima u očuvanju detalja unutar generiranih slika. Iako su njihovi rezultati bili lošiji sudeći po korištenim metrikama, vizualna inspekcija je pokazala da GAN modeli bolje zadržavaju anatomske strukture i pružaju vjerodostojnije slike u odnosu na difuzijske modele. Od korištenih GAN modela, HarmonicGAN treniran na svim regijama ostvario je najbolje rezultate.

Iako difuzijski modeli pokazuju obećavajuće rezultate u smislu distribucijske sličnosti s originalnim podacima, vizualno, GAN modeli pružaju bolju kombinaciju kvalitete i očuvanja detalja unutar transliranih medicinskih slika. Zbog tih karakteristika, GAN modeli se ističu kao superiorniji.

Buduće istraživanje trebalo bi se usmjeriti na unaprjeđenje GAN modela i razvoj novih metoda translacije slika s difuzijskim modelima, koji zadržavaju više detalja izvorne slike. Nadalje, potrebno je provesti daljnja istraživanja kako bi se razvile nove

## *Poglavlje 7. Zaključak*

metrike koje će bolje reflektirati očuvanje detalja translatiranih slika. Te metrike bi trebale omogućiti precizniju evaluaciju kvalitete i vjerodostojnosti translatiranih slika u kontekstu zadržavanja ključnih karakteristika izvornih slika.

# Bibliografija

- [1] S. A. Satryb, T. J. Wilson, and M. M. Patterson, “Casting: all wrapped up,” *Orthopaedic Nursing*, vol. 30, no. 1, pp. 37–41, 2011.
- [2] C. Ekanayake, J. Gamage, P. Mendis, and P. Weerasinghe, “Revolution in orthopedic immobilization materials: A comprehensive review,” *Heliyon*, vol. 9, no. 3, 2023.
- [3] F. Hrzić, I. Žužić, S. Tschauner, and I. Štajduhar, “Cast suppression in radiographs by generative adversarial networks,” *Journal of the American Medical Informatics Association*, vol. 28, no. 12, pp. 2687–2694, 10 2021. , s Interneta, <https://doi.org/10.1093/jamia/ocab192>
- [4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017. , s Interneta, <http://dx.doi.org/10.1109/ICCV.2017.244>
- [5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014. , s Interneta, <https://arxiv.org/abs/1406.2661>
- [6] H. Huang, R. He, Z. Sun, T. Tan *et al.*, “Introvae: Introspective variational autoencoders for photographic image synthesis,” *Advances in neural information processing systems*, vol. 31, 2018.
- [7] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.
- [8] K. Zheng, Y. Cheng, X. Kang, H. Yao, and T. Tian, “Conditional introspective variational autoencoder for image synthesis,” *IEEE Access*, vol. 8, pp. 153 905–153 913, 2020.

## Bibliografija

- [9] R. Zhang, T. Pfister, and J. Li, “Harmonic unpaired image-to-image translation,” *arXiv preprint arXiv:1902.09727*, 2019.
- [10] P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 8780–8794. , s Interneta, [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf)
- [11] C. H. Wu and F. De la Torre, “Unifying diffusion models’ latent space, with applications to cyclediffusion and guidance,” *arXiv preprint arXiv:2210.05559*, 2022.
- [12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [13] X. Zhu, Z. Ghahramani, and J. D. Lafferty, “Semi-supervised learning using gaussian fields and harmonic functions,” in *Proceedings of the 20th International conference on Machine learning (ICML-03)*, 2003, pp. 912–919.
- [14] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 172–189.
- [15] X. Zhu, Z. Ghahramani, and J. D. Lafferty, “Semi-supervised learning using gaussian fields and harmonic functions,” in *Proceedings of the 20th International conference on Machine learning (ICML-03)*, 2003, pp. 912–919.
- [16] A. Razavi, A. Van den Oord, and O. Vinyals, “Generating diverse high-fidelity images with vq-vae-2,” *Advances in neural information processing systems*, vol. 32, 2019.
- [17] R. Montoya-del Angel, K. Sam-Millan, J. C. Vilanova, and R. Martí, “Mam-e: Mammographic synthetic image generation with diffusion models,” *Sensors*, vol. 24, no. 7, p. 2076, 2024.
- [18] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [19] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” *arXiv preprint arXiv:2010.02502*, 2020.

## Bibliografija

- [20] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [21] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2107–2116.
- [22] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*, 2017.
- [23] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [24] P. A. Marin-Reyes, J. Lorenzo-Navarro, and M. Castrillón-Santana, “Comparative study of histogram distance measures for re-identification,” *arXiv preprint arXiv:1611.08134*, 2016.
- [25] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” *Advances in neural information processing systems*, vol. 29, 2016.
- [26] S. Barratt and R. Sharma, “A note on the inception score,” *arXiv preprint arXiv:1801.01973*, 2018.
- [27] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.

# Sažetak

Gips je potporno sredstvo koje obavlja ozlijeđeni dio tijela radi imobilizacije i ubrzanja procesa zacjeljivanja. Na rendgenskim snimkama kostiju, gips se prikazuje kao homogena bijela ili svjetlosiva struktura koja, zbog svoje gustoće, ometa prijenos rendgenskih zraka, čime se smanjuje vidljivost detalja i složenih struktura kostiju. Cilj ovog istraživanja je uklanjanje gipsa s pedijatrijskih medicinskih rendgenskih snimaka kostiju različitih anatomskih regija korištenjem translacije ne-uparenih slika. Tri različite metode su implementirane i evaluirane: CycleGAN, HarmonicGAN i CycleDiffusion. Iako je metoda CycleDiffusion pokazala superiorne rezultate prema kvantitativnim metrikama, vizualna procjena generiranih slika otkrila je česte slučajeve "halucinacija" (gubitak ili izobličenje anatomskih značajki kostiju nakon uklanjanja gipsa). S druge strane, GAN modeli su se pokazali učinkovitijima u očuvanju složenih detalja slika nakon uklanjanja gipsa, iako su prema evaluacijskim metrikama postigli slabije rezultate u usporedbi s CycleDiffusion metodom. Konkretno, HarmonicGAN model treniran na svim anatomskim regijama unutar skupa podataka dao je najbolje rezultate.

***Ključne riječi*** — stabilna difuzija, GAN, uklanjanje gipseva

## Abstract

A cast is a supportive device that encircles the injured part of the body to immobilize it and accelerate the healing process. In radiographic bone imaging, a cast appears as a homogenous white or light gray structure that, due to its density, impedes the transmission of X-rays, thereby diminishing the visibility of details and intricate bone structures. This research aims to suppress casts from pediatric medical X-ray representations of bones in different anatomical regions through unpaired image-to-image translation. Three distinct approaches were employed and evaluated: CycleGAN, HarmonicGAN, and CycleDiffusion. While the CycleDiffusion method demonstrated superior performance based on quantitative metrics, visual assessment of the output images uncovered common instances of "hallucinations" (loss or distortion of bone anatomical features after cast suppression). Conversely, GAN models

proved to be more effective in preserving image intricacies after cast suppression, albeit yielding inferior results compared to the CycleDiffusion method judging by the evaluation metrics. Specifically, the HarmonicGAN model trained on all anatomical regions within the dataset yielded the best results.

***Keywords*** — stable diffusion, GAN, cast suppression