Zahvaljujem mentoru prof. dr. sc. Siniši Šegviću na prenesenom znanju te brojnim inženjerskim i životnim mudrostima. Zahvaljujem prijateljici i asistentici Anji Delić na velikoj pomoći pri izradi ovog rada.

Hvala mojoj obitelji na tome što su moj vječni vjetar u leđa i oslonac. Hvala Klari na podršci, strpljenju i ohrabrivanju. Hvala svim prijateljima na tome što su mi uljepšali i obilježili studentske dane.

SADRŽAJ

1. Uvod

U mnogim sustavima, najzanimljivije pojave su nam one koje znatno odskaču od svih ostalih. Takve pojave nazivamo anomalijama i prisutne su u raznim domenama. U nekima od njih su relativno česte i zanemarive, poput greške na proizvodu u velikim tvorničkim postrojenjima, dok u ostalim sadrže iznimnu vrijednost, poput pronalaska anomalije u krvnim nalazima pacijenta. Anomalije u znanosti su često posljedica greške prilikom mjerenja, ali mogu predstavljati i neki novi, dosad neobjašnjeni fenomen, te dovesti do novih otkrića. Gomilanje velike količine podataka kroz vrijeme dovelo je do potrebe za razvojem kvalitetnih i efikasnih algoritama za pronalaženje anomalija.

Pronalaženje anomalija obično provodimo tehnikama strojnog učenja. Tradicionalne tehnike strojnog učenja se oslanjaju na kvalitetne ručno odabrane značajke, no ponekad je teško unaprijed odrediti što točno čini anomaliju. Prikupljanje velikih skupova podataka i rast računalne moći omogućio je razvoj dubokih modela. Njihova prednost je u tome što automatski uče značajne hijerarhijske reprezentacije podataka i time umanjuju potrebu za ručno dizajniranje značajki.

Jedan od najvećih izazova u detekciji anomalija je njihova rijetkost, što dovodi do nebalansiranih skupova podataka. Ovo otežava razvoj diskriminativnih modela koji će moći pouzdano detektirati raznolike vrste anomalija. S druge strane, generativni modeli uče distribuciju normalnih podataka i anomalijama proglašavaju one podatke koji znatno odstupaju od naučene distribucije. To ih čini primamljivim izborom u polunadziranom i nenadziranom okruženju, slučajevima kad su anomalni primjeri rijetko ili nimalo zastupljeni u skupu za učenje.

Primjer generativnih modela koji se uspješno primjenjuju u zadatku detekcije anomalija su difuzijski modeli. Njihova mana je velika računalna složenost. Zbog toga, proučavamo varijantu difuzijskih modela specijaliziranu za detekciju anomalija koja se zasniva na procjeni difuzijskog vremena (engl. *Diffusion Time Estimation, DTE*).

U ovom radu, bavimo se pronalaženjem anomalnih ljudskih radnji u videozapisima. Tipična primjena je detekcija opasnih i nesvakidašnjih događaja u snimkama nadzornih kamera. Model ne učimo izravno na okvirima videozapisa, već na segmentima

skeletonima, tj. reprezentacijama ljudske poze skupom ključnih točaka na ljudskom tijelu kroz vrijeme. Segmenti skeletona su semantički bogati i znatno manjih dimenzija u odnosu na slike. Uz to, imaju neka povoljna svojstva poput anonimizacije osoba, smanjenja pristranosti na temelju izgleda te manje osjetljivosti na vremenske uvjete.

Učimo dvije varijante DTE modela na skupovima UBnormal i ShanghaiTech. Demonstriramo kompetitivne rezultate na oba skupa podataka. Promatrajući segmente skeletona kao prostorno-vremenske grafove, dodajemo slojeve konvolucije nad grafovima u arhitekturu našeg modela za procjenu difuzijskog vremena. Ovim pristupom postižemo novo stanje tehnike na skupu UBnormal.

2. Teorijska podloga

Ovaj rad tematski spaja dva važna zadatka strojnog učenja: generativno modeliranje i pronalaženje anomalija. U ovom poglavlju predstavit ćemo osnovne teorijske koncepte vezane uz ta područja koji su nužni za shvaćanje korištenih metoda i provedenih eksperimenata.

Najprije, u potpoglavlju 2.1 upoznat ćemo se općenitim pojmom generativnih modela u strojnom učenju. Fokusirat ćemo se zatim na specifični razred generativnih modela koji se zasnivaju na optimiranju donje granice izglednosti podataka. Po uzoru na [47] krenut ćemo od varijacijskih autoenkodera, pa preko hijerarhijskih varijacijskih autoenkodera doći do varijacijskih difuzijskih modela, koji su nam bitni za ovaj rad.

U potpoglavlju 2.2 objasnit ćemo glavne probleme i ciljeve zadatka pronalaženja anomalija. Osvrnut ćemo se na neke od klasičnih pristupa. Posebnu pozornost pridat ćemo jednom modernom pristupu pronalaženja anomalija koji se zasniva na procjeni difuzijskog vremena.

2.1. Generativno modeliranje

Generativno modeliranje zadatak je strojnog učenja koji ima za cilj na temelju slučajnog uzorka $\mathcal{D} = \{x^{(1)}, x^{(2)}, ..., x^{(N)}\}$ aproksimirati stvarnu distribuciju tog uzorka $p^*(\mathbf{x})$ modelom $p_{\boldsymbol{\theta}}(\mathbf{x})$ [35]. Uobičajeno, u strojnom učenju, proces učenja modela se svodi na pronalazak parametara θ takvih da model funkcije gustoće distribucije aproksimira pravu funkciju gustoće, što simbolički možemo zapisati kao

$$p_{\boldsymbol{\theta}}(\mathbf{x}) \approx p^*(\mathbf{x}). \tag{2.1}$$

Jedna od glavnih karakteristika generativnih modela je sposobnost generiranja novih podataka na temelju procijenjene gustoće vjerojatnosti $p_{\boldsymbol{\theta}}(\mathbf{x})$. Osim toga, naučeni generativni model se često može koristiti i za procjenu izglednosti podataka, odnosno može odgovoriti koliko je vjerojatno da promatrani podaci dolaze iz naučene distribucije.

Postoji više pristupa generativnom modeliranju. Generativne suparničke mreže (engl. *Generative Adversarial Networks, GAN*) [25] modeliraju podatke koristeći dvije mreže. Zadatak generatorske mreže je generiranje podataka koji dolaze iz ciljne distribucije. Zadatak diskriminatorske mreže je naučiti razlikovati podatke generirane generatorskom mrežom od stvarnih podataka iz skupa za učenje. Optimizacija ovih mreža se se provodi kao igra u kojoj se mreže suprotstavljaju jedna drugoj: generator želi "prevariti" diskriminator tako što će izgenerirati podatke koji su dovoljno slični ciljanoj distribuciji, a diskriminator nastoji prepoznati uzorke u podacima koji pomažu razlikovati umjetno stvorene podatke od stvarnih. GANovi spadaju u kategoriju *implicitnih generativnih modela* [52] što znači da ne modeliraju eksplicitno funkciju izglednosti podataka. Takvi modeli obično ciljano izbjegavaju računsku intraktabilnost koja se pojavljuje pri eksplicitnom modeliranju funkcije izglednosti te najveći fokus stavljaju na proces generiranja novih uzoraka.

Drugi pristup predstavljaju vjerojatnosni generativni modeli koji se zasnivaju na maksimizaciji funkcije izglednosti parametara modela. Drugim riječima, nastoje naučiti model koji će pridavati veliku gustoću vjerojatnosti podacima iz skupa za učenje. Primjeri takvih generativnih modela su autoregresijski modeli, normalizirajući tokovi i varijacijski autoenkoderi. U tu kategoriju spadaju i modeli zasnovani na energijskoj funkciji, koji definiraju fleksibilnu energijsku funkciju parametriziranu neuronskom mrežom, čijom normalizacijom dolaze do željene distribucije. Alternativno, umjesto izravnog modeliranja, moguće je učiti funkciju mjere (gradijente energijske funkcije) (engl. *score-based models*) neuronskom mrežom te tako naučiti distribuciju podataka. Difuzijski modeli su jedan od suvremenih pristupa generativnom modeliranju koji se pokazao iznimno uspješnim. Oni se mogu interpretirati na dva načina: kao modeli zasnovani na izglednosti te kao modeli zasnovani na procjeni funkcije mjere [30].

2.1.1. Usmjereni vjerojatnosni grafički modeli

Usmjereni vjerojatnosni grafički modeli ili Bayesovske mreže su tip vjerojatnosnih modela kod kojih su sve varijable topološki organizirane u usmjereni aciklički graf [35]. Združena distribucija svih varijabli može se faktorizirati kao

$$p_{\boldsymbol{\theta}}(\mathbf{x}_1, ..., \mathbf{x}_M) = \prod_{j=1}^{M} p_{\boldsymbol{\theta}}(\mathbf{x}_j | Pa(\mathbf{x}_j)) \tag{2.2}$$

$Pa(\mathbf{x}_j)$ pritom označava skup svih roditeljskih varijabli vrha j u pridruženom usmjerenom grafu.

Često složene gustoće vjerojatnosti uvjetnih distribucija želimo parametrizirati nekom jednostavnijom funkcijom $f_{\boldsymbol{\theta}}(Pa(\mathbf{x}_j))$. Unaprijedne neuronske mreže su jedan tip fleksibilnih funkcijskih aproksimatora koji se često koriste za ovu svrhu. U tom slučaju, na ulaz neuronske mreže dovodimo roditeljske varijable $Pa(\mathbf{x}_j)$, a izlaz koristimo kao parametre uvjetne distribucije.

$$\boldsymbol{\pi} = f_{\boldsymbol{\theta}}(Pa(\mathbf{x})) \tag{2.3}$$

$$p_{\boldsymbol{\theta}}(\mathbf{x}|Pa(\mathbf{x})) = p_{\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\pi}) \tag{2.4}$$

Tipičan primjer korištenja neuronskih mreža za procjenu parametara uvjetne distribucije je klasifikacija slika. Distribucija koju pritom procjenjujemo je kategorička distribucija $p_{\boldsymbol{\theta}}(y|\mathbf{x})$ pri čemu je y oznaka razreda, uvjetovana slikom $\mathbf{x}$.

$$\mathbf{p} = f_{\boldsymbol{\theta}}(\mathbf{x}) \tag{2.5}$$

$$p_{\boldsymbol{\theta}}(y|\mathbf{x}) = \text{Kategorička}(y; \mathbf{p}) \tag{2.6}$$

Pritom obično na posljednji sloj neuronske mreže $f_{\boldsymbol{\theta}}(\mathbf{x})$ stavljamo operaciju $\mathrm{softmax}$, da se osiguramo da će se izlazi zbrajati u 1, odnosno da na izlazu imamo ispravne parametre kategoričke distribucije.

2.1.2. Potpuno osmotrivi modeli

Potpuno osmotrivi modeli su usmjereni grafički modeli koji sadrže isključivo varijable koje su dostupne u skupu podataka $\mathcal{D}$. Kod takvih modela, možemo jednostavno izračunati log-vjerojatnost podataka i optimirati model algoritmom gradijentnog spusta. Pod da su pretpostavkom da su primjeri iz skupa $\mathcal{D}$ nezavisno i identično distribuirani (engl. *independently and identically distributed, i.i.d.*), izglednost parametara modela za cijeli skup faktorizira se kao produkt pojedinačnih izglednosti za svaki podatak. Log-izglednost parametara možemo onda računati jednadžbom 2.8.

$$\mathcal{D} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}\} \equiv \{\mathbf{x}^{(i)}\}_{i=1}^{N} \tag{2.7}$$

$$\log p_{\boldsymbol{\theta}}(\mathcal{D}) = \sum_{\mathbf{x} \in \mathcal{D}} \log p_{\boldsymbol{\theta}}(\mathbf{x}) \tag{2.8}$$

2.1.3. Modeli s latentnim varijablama

Potpuno osmotrive modele možemo proširiti u modele s latentnim varijablama. Latentne (ili skrivene) varijable su varijable koje nisu dio skupa podataka što znači da ih

ne možemo osmotriti. Intuicija iza latentnih varijabli je da one predstavljaju skrivene strukture i koncepte koji generiraju promatrane podatke. Zadaća latentnih varijabli je modeliranje složenih interakcija između promatranih (osmotrivih) varijabli. U generativnom modeliranju, često učimo latentne reprezentacije niske dimenzionalnosti. Na to se može gledati kao na jedan oblik kompresije informacija što može pomoći u otkrivanju skrivenih veza u podacima. Latentne varijable obično označavamo sa $\mathbf{z}$, a združenu gustoću vjerojatnosti modela preko osmotrivih i latentnih varijabli s $p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})$. Do gustoće vjerojatnosti preko isključivo osmotrivih varijabli možemo doći marginalizacijom latentnih varijabli $\mathbf{z}$:

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \int p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})\, d\mathbf{z} \tag{2.9}$$

Alternativno, možemo faktorizirati združenu vjerojatnost kao:

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})} \tag{2.10}$$

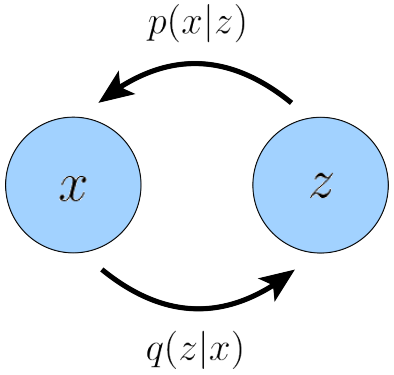
Obe formulacije pokazuju se problematičnima za izračun i maksimizaciju izglednosti parametara $\boldsymbol{\theta}$. Integral u jednadžbi 2.9 integrira preko svih latentnih varijabli $\mathbf{z}$. Ne postoji formula u zatvorenom obliku ili nekakav procjenitelj kojim bismo ga mogli efikasno izračunati kao ni njegov gradijent, za razliku od potpuno osmotrivih modela. Zbog toga kažemo da je njegov izračun *netraktabilan*. U jednadžbi 2.10 je problem to što zapravo ne poznajemo distribuciju $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ u nazivniku.

Razni tipovi generativnih modela s latentnim varijablama koriste različite ideje kako bi zaobišli problem netraktabilnosti. U nastavku ćemo proučiti modele koji to ostvaruju optimiranjem zamjenskog gubitka, odnosno optimiranjem donje ograde logaritma funkcije izglednosti $p_{\boldsymbol{\theta}}(\mathbf{x})$.

2.1.4. Varijacijski autoenkoder i varijacijsko zaključivanje

Varijacijski autoenkoder (engl. *Variational Autoencoder, VAE*) [36] je generativni model s latentnim varijablama koji zaobilazi problem netraktabilnosti aproksimacijom aposteriorne distribucije $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ zamjenskom distribucijom $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$. Parametri $\boldsymbol{\phi}$ nazivaju se *varijacijski* parametri, a zamjenska distribucija $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ se još naziva i *enkoderskom* distribucijom. Optimizacijskim postupkom tražimo varijacijske parametre $\boldsymbol{\phi}$ takve da zamjenska distribucija što bolje aproksimira stvarnu aposteriornu distribuciju:

$$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \approx p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}) \tag{2.11}$$



Slika 2.1: Ilustracija varijacijskog autoenkodera. Enkoder $q(\boldsymbol{z}|\boldsymbol{x})$ definira distribuciju latentnih varijabli $\boldsymbol{z}$ za promatrane podatke $\boldsymbol{x}$. $p(\boldsymbol{x}|\boldsymbol{z})$ dekodira latentne reprezentacije iz latentnog prostora u ulazni prostor. Ilustracija preuzeta iz [47].

"Varijacijski" u imenu varijacijskog autoenkodera dolazi od toga što optimizacijom tražimo najbolji $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ iz cijele obitelji distribucija parametriziranih s $\boldsymbol{\phi}$. "Autoenkoder" dolazi od sličnosti arhitekture varijacijskog autoenkodera s tradicionalnim autoenkoderima iz dubokog učenja, kod kojih neuronska mreža enkoder projicira ulazne podatke u nižedimenzionalni latentni prostor, a zatim neuronska mreža dekoder latentne reprezentacije vraća u originalni prostor.

Optimizacijski cilj varijacijskog enkodera je varijacijska donja ograda, odnosno donja ograda na log-izglednost promatranih podataka (engl. *Evidence Lower Bound, ELBO*). Odnos log-izglednosti i varijacijske donje ograde može se prikazati kao:

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) \geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right] \tag{2.12}$$

Da taj odnos bude očitiji, možemo pratiti sljedeći izvod:

$$\begin{aligned}
\log p_{\boldsymbol{\theta}}(\mathbf{x}) &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x})\right] && (2.13)\\
&= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})}\right] && (2.14)\\
&= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\frac{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})}\right] && (2.15)\\
&= \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right]}_{\substack{=\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x})\\ \text{(ELBO)}}} + \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})}\right]}_{=D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}))} && (2.16)
\end{aligned}$$

Iz jednadžbe 2.16 vidimo da je logaritam izglednosti jednak zbroju varijacijske donje ograde i KL divergencije između aproksimacijske aposteriorne distribucije $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ i stvarne aposteriorne distribucije $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$. KL divergencija je mjera udaljenosti dviju

distribucija koja je po definiciji uvijek nenegativna, čime je dokazano da ELBO nikad ne može premašiti logaritam izglednosti, odnosno da je ELBO doista njegova donja ograda.

U idealnom slučaju želimo minimizirati navedenu KL divergenciju, ali to ne možemo raditi izravno zbog toga što ne znamo stvarnu aposteriornu distribuciju $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$. Međutim, primijetimo da je lijeva strana jednadžbe ($\log p_{\boldsymbol{\theta}}(\mathbf{x})$) konstantna s obzirom na varijacijske parametre $\boldsymbol{\phi}$. To znači da maksimizacijom varijacijske donje ograde (lijevi član) s obzirom na $\boldsymbol{\phi}$ također minimiziramo KL divergenciju između $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ i $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ (desni član). Minimizacijom KL divergencije približavamo ove dvije distribucije jednu drugoj te učimo latentne strukture koje će dobro opisivati naše podatke. S druge strane, maksimizacijom varijacijske donje ograde s obzirom na parametre $\boldsymbol{\theta}$ istovremeno maksimiziramo i $\log p_{\boldsymbol{\theta}}(\mathbf{x})$ s lijeve strane jednadžbe, odnosno vjerojatnost promatranog skupa podataka postaje sve veća pod modelom što je ujedno naš izvorni cilj.

Varijacijska donja ograda se dalje može raspisati kao:

$$\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right] = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right] \tag{2.17}$$

$$= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right] + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right] \tag{2.18}$$

$$= \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})\right]}_{\text{rekonstrukcijski član}} - \underbrace{D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))}_{\text{član podudaranja s apriornom dist.}} \tag{2.19}$$

U jednadžbi 2.19 jasno možemo prepoznati izraz $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ koji predstavlja *enkoder*, odnosno za dane ulaze definira distribuciju latentnih varijabli. Izraz $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ predstavlja *dekoder* jer latentne reprezentacije preslikava u distribuciju u originalnom prostoru podataka. Vizualno taj odnos možemo prikazati ilustracijom 2.1. U jednadžbi 2.19 također rastavljamo varijacijsku donju ogradu na dva člana. Prvi je rekonstrukcijski član jer osigurava da naučena združena distribucija modelira latentne varijable iz kojih se mogu ponovo generirati originalni podaci. Drugi član mjeri sličnost naučene aposteriorne distribucije $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ s apriornom distribucijom latentnih varijabli $p(\mathbf{z})$. Minimizacija drugog člana tjera model da zaista nauči nešto što nalikuje distribuciji umjesto da se svede na neku trivijalnu funkciju gustoće vjerojatnosti, npr. Diracovu delta funkciju.

Česti izbor za apriornu distribuciju kod varijacijskog autoenkodera je standardna jedinična multivarijatna Gaussova distribucija, a za enkoder multivarijatna Gaussova

razdioba s dijagonalnom matricom kovarijance.

$$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathbf{x}), \boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\mathbf{x})\mathbf{I}) \tag{2.20}$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I}) \tag{2.21}$$

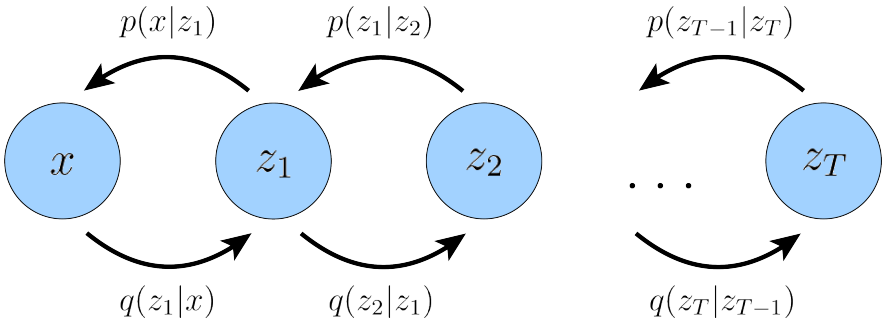
Pod tim se uvjetima KL divergencija u varijacijskoj donjoj ogradi može izračunati u zatvorenoj analitičkoj formi, a rekonstrukcijski član se može aproksimirati Monte Carlo procjeniteljem. Parametre $\boldsymbol{\theta}$ i $\boldsymbol{\phi}$ učimo združeno stohastičnim gradijentnim spustom.

Generiranje novih podataka naučenim varijacijskim autoenkoderom je jednostavno i provodi se kroz dva koraka. Najprije uzorkujemo latentni vektor iz apriorne distribucije $p(\mathbf{z})$. Zatim taj vektor postavljamo na ulaz dekoderske neuronske mreže koja na izlazu vraća parametre dekoderske distribucije $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$. Uzorkovanjem iz dekoderske distribucije generiramo novi primjer. Ako je naučeni latentni prostor semantički smislen, mijenjanjem latentnog vektora $\mathbf{z}$ možemo donekle i kontrolirati kakve podatke želimo generirati.

2.1.5. Hijerarhijski varijacijski autoenkoder

Hijerarhijski varijacijski autoenkoder (engl. *Hierarchical Variational Autoencoder, HVAE*) je generalizacija varijacijskog autoenkodera na latentne varijable višeg reda. Kod klasičnih varijacijskih autoenkodera interpretiramo promatrane podatke kao generirane od strane latentnih varijabli koje reprezentiraju apstraktne koncepte i složene strukture prisutne u osmotrivim varijablama. Kod hijerarhijskih varijacijskih autoenkodera tu ideju generaliziramo na način da latentne varijable također interpretiramo kao generirane od strane nekih latentnih varijabli višeg reda. U općenitom slučaju, imamo T hijerarhijskih razina i svaka od njih može biti uvjetovana bilo kojom kombinacijom preostalih. Ako se ograničimo na uvjet da je svaka latentna varijabla $\mathbf{z}_t$ uvjetovana isključivo neposredno prethodnom latentnom varijablom $\mathbf{z}_{t+1}$, tada takav model zovemo Markovljev hijerarhijski varijacijski autoenkoder (engl. *Markovian Hierarchical Variational Autoencoder, MHVAE*). MHVAE možemo vizualno prikazati ilustracijom 2.2.

Združena distribucija preko osmotrivih i svih latentnih varijabli Markovljevog va-



Slika 2.2: Ilustracija Markovljevog hijerahijskog varijacijskog autoenkodera s T latentnih varijabli. I dalje interpretiramo varijablu $\mathbf{x}$ kao generiranu latentnom varijablom $\mathbf{z}_1$. Generalizacijom, latentna varijabla $\mathbf{z}_t$ generirana je latentnom varijablom $\mathbf{z}_{t+1}$. Ilustracija preuzeta iz [47].

rijacijskog autoenkodera te aposteriorna distribucija mogu se faktorizirati kao:

$$p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}_{1:T}) = p(\mathbf{z}_T) p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}_1) \prod_{t=2}^{T} p_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}|\mathbf{z}_t) \tag{2.22}$$

$$q_{\boldsymbol{\phi}}(\mathbf{z}_{1:T}|\mathbf{x}) = q_{\boldsymbol{\phi}}(\mathbf{z}_1|\mathbf{x}) \prod_{t=2}^{T} q_{\boldsymbol{\phi}}(\mathbf{z}_t|\mathbf{z}_{t-1}) \tag{2.23}$$

Varijacijska donja ograda se također jednostavno generalizira na slučaj s više latentnih varijabli:

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) \geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}_{1:T}|\mathbf{x})} \left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}_{1:T})}{q_{\boldsymbol{\phi}}(\mathbf{z}_{1:T}|\mathbf{x})} \right] \tag{2.24}$$

2.1.6. Varijacijski difuzijski modeli

Varijacijski difuzijski modeli su klasa probabilističkih generativnih modela koja je stekla veliku popularnost u zadnjem vremenu. Pokazali su se kao vrlo uspješan pristup modeliranju i generiranju novih podataka, osobito u domenama slike, zvuka i videa. Difuzijski modeli su izvorno predstavljeni u [68]. Autori navode ideje iz neravnotežne termodinamike kao inspiraciju za dizajn modela. U [30] autori povezuju difuzijske modele s modelima zasnovanim na učenju gradijenta funkcije izglednosti (engl. *score matching*). Također, prvi pokazuju da difuzijski modeli mogu generirati slike visoke kvalitete koje se mogu mjeriti s ostalim popularnim generativnim modelima. To potvrđuju postizanjem stanja tehnike za FID metriku na skupu CIFAR-10 [39]. Iako su pokazali da su difuzijski modeli usporedivi s ostalim generativnim modelima po pitanju kvalitete generiranih slika, još uvijek su zaostajali po pitanju ostalih metrika, poput log-izglednosti. U [54] istraživači iz OpenAI-a uvode niz jednostavnih prijedloga i poboljšanja s pomoću kojih znatno povećavaju log-izglednost bez gubljenja na

kvaliteti generiranih slika. Svoj rad nastavljaju u [14], u kojem pokazuju da difuzijski modeli mogu ostvariti bolju vjerodostojnost (engl. *fidelity*) kao i bolju raznolikost (engl. *diversity*) generiranih slika od generativnih suparničkih mreža.

Formalno, varijacijski difuzijski modeli su zapravo Markovljevi hijerarhijski varijacijski autoenkoderi s 3 ograničenja:

- Dimenzionalnost svih latentnih varijabli $\mathbf{z}_t$ je jednaka dimenzionalnosti stvarnih (osmotrivih) podataka $\mathbf{x}$.
- Distribucija enkodera u svakom vremenskom koraku (razini hijerarhije) je fiksirana na Gaussovu distribuciju koja je centrirana oko izlaza prethodnog vremenskog koraka.
- Parametri Gaussove distribucije svih enkodera su odabrani tako da se zadnja latentna varijabla $\mathbf{z}_T$ ravna prema jediničnoj Gaussovoj distribuciji.

Koristeći prvo ograničenje, možemo izmijeniti oznake tako da i stvarne i sve latentne varijable označimo s $\mathbf{x}_t$, pri čemu $t = 0$ označava stvarne podatke, a $t \in [1, T]$ latentne varijable. Faktorizacija aposteriorne distribucije može se onda kompaktno zapisati kao:

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{2.25}$$

Drugo ograničenje fiksira distribuciju latentne varijable na Gaussovu distribuciju uvjetovanu latentnom varijablom iz prethodnog vremenskog koraka. Srednju vrijednost i varijancu Gaussove distribucije u svakom koraku najčešće odabiremo kao hiperparametre parametrizirane na sljedeći način:

$$\boldsymbol{\mu}_t(\mathbf{x}_t) = \sqrt{\alpha_t}\mathbf{x}_{t-1} \tag{2.26}$$

$$\boldsymbol{\Sigma}_t(\mathbf{x}_t) = (1 - \alpha_t)\mathbf{I} \tag{2.27}$$

Dakle, enkoderske prijelazne distribucije možemo matematički izraziti kao:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1 - \alpha_t)\mathbf{I}) \tag{2.28}$$

Treće ograničenje koje smo naveli postavlja uvjet na raspored hiperparametara α_t. Raspored mora osigurati da će se latentna varijabla u posljednjem vremenskom koraku ravnati prema jediničnoj Gaussovoj distribuciji, što možemo pisati kao $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$. Združena distribucija svih varijabli varijacijskog difuzijskog modela može se onda zapisati kao:

$$p(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t) \tag{2.29}$$

Ulančano uzorkovanje latentnih varijabli iz Gaussove distribucije uvjetovane latentnom varijablom iz prethodnog koraka možemo interpretirati kao postepeno zašumljivanje ulaznog podatka $\mathbf{x}_0$. U svakom vremenskom koraku dodajemo malu količinu Gaussovog šuma i tako postepeno uništavamo informaciju iz ulaznog podatka. Ovaj postupak koji kreće od podatka $\mathbf{x}_0$ i završava s latentnom varijablom $\mathbf{x}_T$ koja izgleda kao jedinični Gaussov šum naziva se *unaprijedni difuzijski Markovljev proces*. U ovom kontekstu, raspored hiperparametara α_t kroz vremenske korake difuzije t nazivamo još i raspored šuma. Uz fiksan raspored šuma, unaprijedni proces je potpuno određen i nema parametara. Primijetimo da je to drugačije nego kao kod varijacijskog autoenkodera gdje je unaprijedni proces (enkoder) imao parametre $\boldsymbol{\phi}$ koje smo morali učiti.

Niz prijelaza koji kreće od jediničnog Gaussovog šuma u vremenskom koraku $t = T$ i završava s novim podatkom u koraku $t = 0$ naziva se *unatražni difuzijski Markovljev proces*. Prijelazi unatražnog procesa također se modeliraju Gaussovom distribucijom. [17] su pokazali da uz dovoljno male varijance prijelaznih Gaussovih distribucija, unatražni proces ima isti funkcionalni oblik kao i unaprijedni. Prijelaze u unatražnom procesu označavamo s $p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)$. Želimo naučiti unatražne prijelaze koji će znati ukloniti šum u vremenskom koraku t kako bi proizveli podatak u koraku $t-1$. Srednju vrijednost prijelazne Gaussove distribucije u unatražnom prolazu računamo neuronskom mrežom. Što se tiče varijance, prvi radovi poput [30] su je fiksirali da odgovara poznatoj varijanci stvarnog unatražnog procesa dodatno uvjetovanog s $\mathbf{x}_0$. Kasniji radovi poput [54] pokazali su da učenje varijanci može ubrzati kasnije uzorkovanje naučenog modela uz neznatnu razliku u generiranim primjerima.

Optimizacija se provodi maksimizacijom varijacijske donje ograde koja se u slučaju varijacijskih difuzijskih modela može raspisati kao:

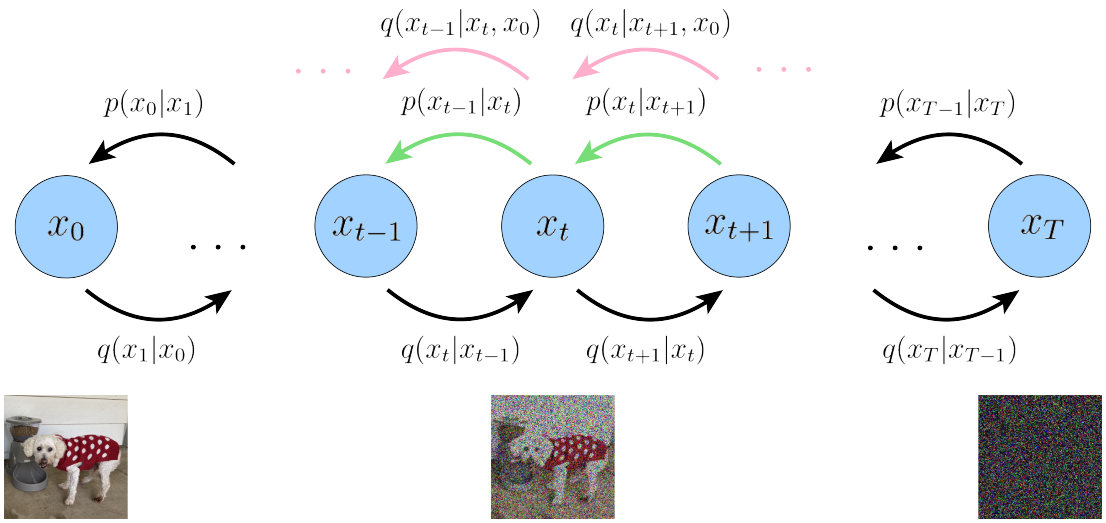
$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[\log \frac{p_{\boldsymbol{\theta}}(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] \tag{2.30}$$

$$= \underbrace{\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}_0|\mathbf{x}_1)\right]}_{\text{rekonstrukcijski član}} - \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T))}_{\text{član podudaranja s apriornom dist.}} \tag{2.31}$$

$$- \sum_{t=1}^{T-1} \underbrace{\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t))\right]}_{\text{član podudaranja unatražnih prijelaza}}$$

Članove možemo interpretirati zasebno:

1. *Rekonstrukcijski član* je analogan rekonstrukcijskom članu iz varijacijske donje ograde varijacijskog autoenkodera. Jednostavno se računa Monte Carlo procjeniteljem.



Slika 2.3: Ilustracija varijacijskog difuzijskog modela s T vremenskih koraka. Korak x_0 predstavlja ulazni podatak, npr. sliku psa. Kroz idućih T koraka unaprijednog procesa, slici dodajemo male količine Gaussovog šuma. Posljednji korak x_T izgleda kao potpuni šum. Zatim učimo model koji će moći krenuti od potpunog šuma i kroz T koraka ga postepeno uklanjati te tako generirati novi podatak. To ostvarujemo minimizacijom očekivane KL divergencije između stvarnog unatražnog prijelaza (roza strelica) i njegove aproksimacije (zelena strelica). Ilustracija preuzeta iz [47].

2. *Član podudaranja s apriornom distribucijom* se brine da zašumljeni ulaz u posljednjem vremenskom koraku unaprijednog procesa odgovara distribuciji koja je apriorno odabrana kao jedinična Gaussova. Ovaj član nema parametara za učenje. Ako smo ispravno odabrali hiperparametre unaprijednog procesa, obe distribucije će biti približno jednake pa će član iznositi 0.

3. *Član podudaranja unatražnih prijelaza* dominira izrazom. On tjera model da nauči unatražne prijelaze $p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)$ koji će odgovarati stvarnim prijelazima (dodatno uvjetovanima na $\mathbf{x}_0$). Drugim riječima, tjera difuzijski model da nauči uklanjati šum iz podatka u pripadajućem vremenskom koraku.

Unaprijedni i unatražni proces varijacijskih difuzijskih modela cjelovito možemo prikazati kao na ilustraciji 2.3.

Zbroj nezavisnih Gaussovih distribucija je i dalje Gaussova distribucija. Ta činjenica daje zgodno svojstvo unaprijednom procesu. Da bismo došli do zašumljenog podatka u koraku t, ne trebamo pratiti sve korake 0 do $T-1$ unaprijednog Markovljevog procesa. Umjesto toga, možemo doći do proizvoljnog koraka difuzije t u jednom

koraku, rekurzivno koristeći izraz 2.28 i sljedeću parametrizaciju:

$$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i \tag{2.32}$$

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{2.33}$$

U jednadžbi 2.32 rastavili smo ELBO na tri člana i zaključili da je treći član najbitniji jer dominira cijelim izrazom. On se, podsjetimo se, računa kao KL divergencija između stvarnog unatražnog prijelaza $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ i njegove aproksimacije $p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)$. Podužim izvodom, koristeći Bayesovo pravilo te izraze 2.28 i 2.33, može se pokazati da se stvarni unatražni prijelaz također ravna prema Gaussovoj distribuciji. Uzimajući u obzir da pretpostavljamo da se aproksimacijski unatražni prijelaz također ravna prema Gaussovoj distribuciji, to znači da se KL divergencija u trećem članu gubitka može izračunati analitički u zatvorenoj formi. Polazeći od te činjenice, možemo izvesti 3 ekvivalentna konačna gubitka za difuzijske modele:

$$\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2} \left[\|\hat{\mathbf{x}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \mathbf{x}_0\|_2^2\right] && (2.34) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \frac{(1-\alpha_t)^2}{(1-\bar{\alpha}_t)\alpha_t} \left[\|\boldsymbol{\epsilon}_0 - \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\|_2^2\right] && (2.35) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \frac{(1-\alpha_t)^2}{\alpha_t} \left[\|\boldsymbol{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \nabla \log p(\mathbf{x}_t)\|_2^2\right] && (2.36)
\end{aligned}$$

Dakle, da bismo minimizirali KL divergenciju između stvarnog unatražnog prijelaza i njegove aproksimacije, možemo učiti neuronsku mrežu koja će predviđati originalnu sliku (jednadžba 2.34), šum na originalnoj slici (jednadžba 2.35) ili gradijent izglednosti, koji se još naziva mjera (engl. *score*) (jednadžba 2.36). U [30] su empirijski pokazali da najbolje rezultate postižu predviđanjem šuma. Štoviše, tvrde da u izbacivanjem skalara u jednadžbi 2.35 generiraju primjere bolje kvalitete, čime gubitak poprima vrlo jednostavan oblik:

$$\mathcal{L}_{jednostavni} = \mathbb{E}_{t,\mathbf{x}_0,\epsilon_0}\left[\|\boldsymbol{\epsilon}_0 - \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\|_2^2\right] \tag{2.37}$$

Učenje se može opisati sljedećim algoritmom:

Algoritam 1 Učenje modela

Ponavljaj
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim Uniformna(\{1, ..., T\})$
  $\boldsymbol{\epsilon}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Odradi korak gradijentnog spusta na temelju
    $$\nabla_{\boldsymbol{\theta}} ||\boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)||_2^2$$
Do konvergencije

Generiranje/uzorkovanje novog podatka kreće od zadnjeg koraka i prolazi kroz sve korake difuzije. U posljednjem unatražnom koraku $t = 1$ ne dodajemo varijancu (šum) jer očekujemo da bi on samo mogao "pokvariti" konačno generirani podatak u koraku $t = 0$.

Algoritam 2 Uzorkovanje iz naučene distribucije

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Za $t = T, \ldots, 1$ **radi**
  $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ako $t > 1$, inače $\mathbf{z} = \mathbf{0}$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)) + \sigma_t \mathbf{z}$
Vrati $\mathbf{x}_0$

Već smo spomenuli kako je [54] uveo niz poboljšanja nad klasičnim varijacijskim difuzijskim modelima koje smo predstavili u ovom potpoglavlju. Ovdje ćemo ih samo navesti:

- ne koriste fiksnu varijancu u unatražnom procesu, već ju također uče
- uvode hibridni gubitak, kao kombinaciju jednostavnog gubitka $\mathcal{L}_{jednostavni}$ i kompletnog gubitka varijacijske donje ograde $\mathcal{L}_{vdo}$
- ne koriste linearni nego kosinusni raspoređivač šuma
- koriste veći broj koraka (4000 umjesto 1000)
- pokazuju da model generira dobre slike i sa smanjenim brojem koraka prilikom uzorkovanja

2.2. Pronalaženje anomalija

Definicija anomalije je opažanje koje značajno odskače od nekog koncepta normalnosti [63]. Anomalije nazivamo još i novinama (engl. *novelty*) ili stršećim vrijednos-

tima (engl. *outlier*). Pronalaženje (ili detekcija) anomalija zadatak je koji proučava pronalaženje anomalnih opažanja koristeći metode, modele i algoritme zasnovane na podacima. Važnost pronalaženja anomalija se proteže kroz razne domene, poput kibernetičke sigurnosti [68], detekcije prevara u području financija [4], detekcije grešaka u industrijskim postrojenjima [45], detekcije zabrinjavajućeg stanja pacijenta u medicini [71], neobičnih transakcija na burzi dionica i brojnih drugih [24]. Pronalazak anomalija također može pomoći raznim znanstvenim disciplinama [26, 55] u širenju granice poznatoga te pronalasku novih otkrića i saznanja.

Razlikovanje anomalnih događaja od normalnih je poprilično zahtjevan zadatak. Inherentna svojstva anomalija su nepoznatost i raznolikost. Čak i u slučajevima kada imamo dostupne primjere anomalnih događaja u podacima, to ne garantira da se u primjeni neće pojaviti neki nepredviđeni oblik anomalije kojeg nismo obuhvatili ograničenim skupom podataka. Želimo graditi robusne modele koji će se znati nositi s tim. Osim toga, veliki izazov predstavlja varijabilnost unutar normalnih primjera, koja može dovesti do toga da krivo klasificiramo normalne primjere kao anomalije. Uz to, anomalije su općenito jako rijetke, što dovodi do nebalansiranih skupova podataka. U mnogim slučajevima prikupljanje anomalija je preskupo ili čak nemoguće. Tada se oslanjamo na nenadzirane ili polunadzirane tehnike strojnog učenja.

Anomalije možemo svrstati u nekoliko kategorija [63], pri čemu svaka potencijalno zahtjeva zaseban pristup pri pronalaženju:

- *Točkaste anomalije*: Individualni podaci koji značajno odstupaju od ostalih podataka, poput lažne financijske transakcije.
- *Kontekstualne anomalije*: Podaci koji su anomalni u određenom kontekstu, poput očitanja temperature koja je inače normalna, ali anomalna za neku određenu regiju ili vrijeme u godini.
- *Kolektivne anomalije*: Skup podataka koji zajedno čini anomaliju, poput skupine mrežnih paketa koje zajedno upućuju na kibernetički napad.
- *Anomalije niske razine*: Anomalije u osnovnim značajkama podataka, poput šuma ili artefakata u slikama.
- *Anomalije visoke razine*: Anomalije u apstraktnim značajkama visoke razine, poput anomalnih objekata ili pojava u videu.

Kroz prošlost, pronalaženje anomalija najčešće se provodilo *klasičnim metodama* strojnog učenja. Klasične metode otkrivanja anomalija često su nenadzirane, što znači da nemaju pristup oznakama primjera prilikom učenja. Neki primjeri takvih metoda su:

analiza glavnih komponenti (engl. *Principal Component Analysis, PCA*) [31], jednorazredni stroj potpornih vektora (engl. *One-Class Support Vector Machine, OC-SVM*) [65], algoritmi zasnovani na najbližim susjedima (engl. *Nearest Neighbor Algorithms*) [37], jezgrena procjena gustoće (engl. *Kernel Density Estimation, KDE*) [57]. Ono što je zajedničko ovim metodama je modeliranje isključivo normalnih podataka. Anomalije se onda pronalaze kao odstupanja od naučenog modela normalnosti. Ovakav pristup može biti problematičan u slučajevima kada je distribucija normalnih primjera promjenjiva.

Pojava dubokog učenja revolucionirala je pronalaženje anomalija u složenim, visokodimenzionalnim podacima. Duboki modeli automatski uče hijerarhijske reprezentacije značajki bez potrebe ručnog dizajniranja značajki kao kod klasičnih metoda. Primjeri dubokih modela koji su uspješno iskorišteni u svrhu pronalaženja anomalija su: autoenkoderi (engl. *Autoencoder, AE*) [10], generativne suparničke mreže (engl. *Generative Adversarial Network, GAN*) [64], normalizirajući tokovi (engl. *Normalizing Flow*) [53], duboki jednorazredni klasifikatori (engl. *Deep One-Class Classification*) [56], duboki samonadzirani modeli (engl. *Deep Self-Supervised Models*) [70].

Ovisno o vrsti podataka kojih imamo na raspolaganju prilikom učenja, postoji nekoliko različitih pristupa pronalaženju anomalija:

- *Nenadzirani pristup*: Prilikom učenja dostupni su samo neoznačeni podaci $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$. Pretpostavka je da su podaci identično i nezavisno distribuirani te da su generirani nekom distribucijom $\mathbb{P}$ koja odgovara distribuciji normalnih podataka. U praksi, ta distribucija može sadržavati šum ili može biti "zagađena" anomalnim podacima koje nismo predvidjeli.
- *Polunadzirani pristup*: Prilikom učenja dostupni su neoznačeni podaci $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$ i označeni podaci $(\tilde{\mathbf{x}}_1, \tilde{\mathbf{y}}_1), \ldots, (\tilde{\mathbf{x}}_m, \tilde{\mathbf{y}}_m) \in \mathcal{X} \times \mathcal{Y}$. Obično vrijedi $m \ll n$, odnosno puno je manje označenih nego neoznačenih podataka, jer je označavanje skup i vremenski zahtjevan posao. Uključivanje anomalnih primjera u proces učenja može značajno poboljšati performanse modela. U ponekoj literaturi poput [9] se polunadziranim pristupom naziva i slučaj u kojem model uči isključivo na normalnim primjerima.
- *Nadzirani pristup*: Svi podaci za učenje su označeni $(\tilde{\mathbf{x}}_1, \tilde{\mathbf{y}}_1), \ldots, (\tilde{\mathbf{x}}_m, \tilde{\mathbf{y}}_m) \in \mathcal{X} \times \mathcal{Y}$. Podaci se sastoje od normalnih i anomalnih primjera. Ne postoji jedinstvena konvencija za oznake u literaturi. Mi označavamo anomalne primjere s $\tilde{\mathbf{y}} = 1$, a normalne s $\tilde{\mathbf{y}} = 0$. U ovom pristupu pronalaženje anomalija se svodi na zadatak nadzirane binarne klasifikacije.

Nakon učenja, naučene modele evaluiramo na skupovima za testiranje koji se sastoje od anomalnih i normalnih primjera. Izlaz modela najčešće je *mjera anomalnosti*. Mjera anomalnosti je funkcija koja svaki podatak preslikava u broj koji govori koliko u kojoj mjeri (koliko jako) smatramo taj podatak anomalijom.

$$\mathbf{s}\colon \mathcal{X} \longmapsto \mathbb{R} \tag{2.38}$$

Predikciju o tome je li ulazni podatak anomalija provodimo usporedbom mjere anomalnosti i nekog predefiniranog praga τ.

$$\text{Klasifikacijska odluka} = \begin{cases} \text{anomalija} & \text{ako } \mathbf{s} \geq \tau \\ \text{normalan primjer} & \text{ako } \mathbf{s} < \tau \end{cases} \tag{2.39}$$

Popularan izbor za prag je vrijednost koja postiže 95% na metrici TPR. U općenitom slučaju, prag se podešava ovisno o zahtjevima specifičnog problema koji se pokušava riješiti.

2.2.1. Pronalaženje anomalija metodom k najbližih susjeda

Metoda k najbližih susjeda (engl. *k-Nearest Neighbors, kNN*) [12] je jedna od klasičnih metoda za pronalaženje anomalija. kNN je neparametarski algoritam koji se u različitim varijantama koristi u zadacima klasifikacije i regresije. Osnovna pretpostavka algoritma je da se slični podaci nalaze blizu jedan drugome u prostoru značajki, a različiti daleko.

Algoritam se sažeto može prikazati na sljedeći način:

Algoritam 3 Metoda k najbližih susjeda za pronalaženje anomalija

Ulaz: Skup podataka $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, broj razmatranih susjeda k

Izlaz: Mjere anomalnosti za svaki podatak iz $\mathcal{D}$

1: **Za** svaki podatak $\mathbf{x}_i$ iz $\mathcal{D}$ **radi**

2: Izračunaj udaljenost od $\mathbf{x}_i$ do svih drugih podataka u $\mathcal{D}$

3: Pronađi k najbližih susjeda od $\mathbf{x}_i$

4: Izračunaj mjeru anomalnosti s_i za $\mathbf{x}_i$ kao prosječnu udaljenost do njegovih k najbližih susjeda:

$$\mathbf{s}_i = \frac{1}{k} \sum_{j=1}^{k} d(\mathbf{x}_i, \mathbf{x}_{ij})$$

pri čemu je $\mathbf{x}_{ij}$ j-ti najbliži susjed od $\mathbf{x}_i$, a d neka mjera udaljenosti

5: **Vrati** Mjere anomalnosti $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$

Postoji nekoliko hiperparametara koje mogu značajno poboljšati efikasnost ovog algoritma ako ih ispravno odaberemo. Podatke prije provođenja algoritma možemo projicirati u neki povoljniji prostor koji će biti prikladniji s obzirom na vrstu anomalija koju želimo detektirati. Izbor mjere udaljenosti također je od velike važnosti. Euklidska udaljenost, premda dobro radi s niskodimenzionalnim podacima, može biti problematična kod visokodimenzionalnih podataka. Kako se broj dimenzija povećava, podaci postaju sve više i sve više jednako udaljeni jedni od drugih u Euklidskom prostoru. Ovaj fenomen se popularno naziva prokletstvo dimenzionalnosti [6]. Primjeri nekih alternativnih mjera udaljenosti su Manhattan, Minkowski, Jaccardova, Hammingova i kosinusna udaljenost. Odabir hiperparametra k je također jedna vrijednost koju treba validirati. Ako odaberemo jako malu vrijednost za k, riskiramo da se algoritam "prenauči", odnosno da bude preosjetljiv na šum u podacima. Preveliki odabir k će zagladiti razlike između anomalnih i normalnih primjera, te će algoritam izgubiti mogućnost pronalaska lokaliziranih anomalija zbog gledanja prevelikog susjedstva. Za kraj, osim uprosječivanja udaljenosti do k najbližih susjeda, moguće je i koristiti neku drugu agregaciju, poput minimuma, maksimuma, medijana ili zbroja.

Glavne prednosti metode k najbližih susjeda su jednostavnost algoritma, prilagodljivost na različite tipove podataka dok god možemo definirati mjeru udaljenosti za njih te to što metoda ne pretpostavlja ništa o distribuciji podataka. Glavni nedostatak algoritma je visoka (kvadratna u broju primjera) vremenska složenost, posebno kada radimo s velikim skupovima podataka, jer zahtjeva izračunavanje udaljenosti za sve parove primjera. Pored toga, kNN ima veliku prostornu složenost jer mora spremiti cijeli skup podataka u memoriju.

Kroz vrijeme razvile su se mnoge varijante ovog algoritma. kNN s težinama (engl. *weighted kNN*) [15] ne tretira sve susjede jednako prilikom izračuna prosječne udaljenosti, nego računa težinski prosjek, pri čemu bliži susjedi imaju veću težinu od udaljenijih. Tehnike poput [5, 60] fokusiraju se na smanjivanje vremenske složenosti podrezivanjem (engl. *pruning*) ili particioniranjem prostora. Neki pristupi [58] umjesto isključivo najbližih susjeda za pronalaženje anomalija koriste koncepte reverznih [38] ili dijeljenih [80] zajedničkih susjeda.

2.2.2. Pronalaženje anomalija procjenom difuzijskog vremena

U poglavlju 2.1.6 naveli smo da su se varijacijski difuzijski modeli [30] pokazali kao moćan razred generativnih modela. U posljednje vrijeme su se iskazali i kao dobra opcija za pronalaženje anomalija, osobito u domeni slika [79, 74, 75] i videa [18].

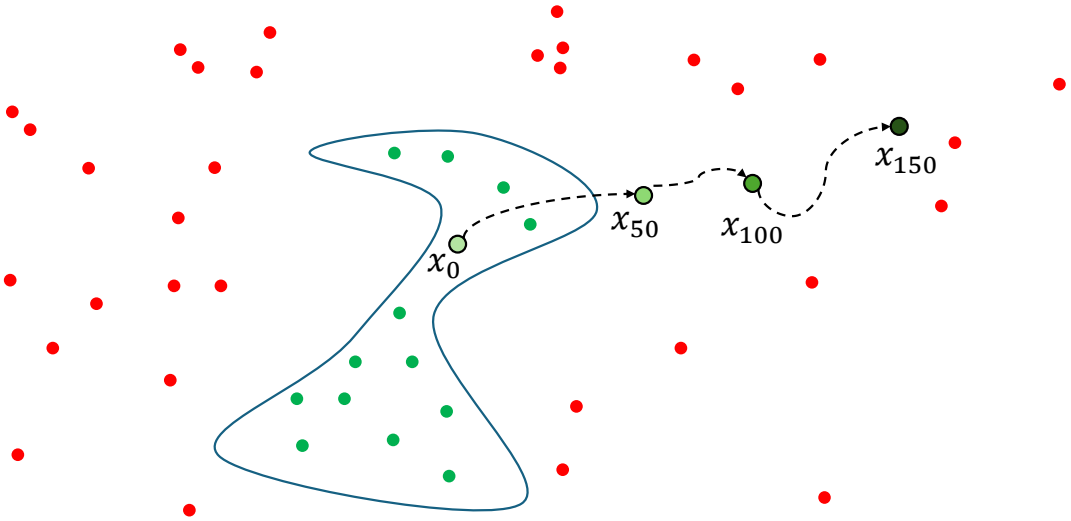
Pronalaženje anomalija se provodi tako da se na ulazni podatak doda određena količina šuma te se zatim podatak provodi kroz unatražni difuzijski proces. Zatim se mjera anomalnosti računa kao rekonstrukcijska udaljenost tako dobivenog podatka od originalnog podataka. Prednost korištenja rekonstrukcijskog gubitka na slikovnim podacima je mogućnost jednostavne lokalizacije anomalija. Anomalije su tamo gdje se originalna i rekonstruirana slika razlikuju. Međutim, nedostatak korištenja difuzijskih modela na ovaj način je velika računalna složenost, jer za svaki podatak moramo proći kroz sve vremenske korake unatražnog procesa.

Rekonstrukcijskim gubitkom zapravo mjerimo koliko je neki podatak udaljen od naučene mnogostrukosti podataka iz skupa za učenje. U [44] tvrde da je učenje cijelog unatražnog procesa nepotrebno ako nam je glavni cilj detekcija anomalija. Iz tog razloga predlažu jednostavniji postupak u kojem modeliramo distribuciju difuzijskog vremena za dani ulazni primjer. Pretpostavka je da se anomalije nalaze daleko od naučene mnogostrukosti, pa očekujemo da će procjena vremenske distribucije za njih imati veću gustoću vjerojatnosti u višim vremenskim koracima. S druge strane, normalni primjeri se nalaze blizu naučene mnogostrukosti pa će procijenjena distribucija difuzijskog vremena imati veću gustoću u početnim vremenskim koracima. Autori tvrde da se ovaj pristup može promatrati kao aproksimacija rekonstrukcijske greške i nazivaju ga procjena difuzijskog vremena (engl. *Diffusion Time Estimation, DTE*).

DTE razmatra nenadzirani i polunadzirani pristup pronalaženju anomalija. Polunadzirani pristup u ovom slučaju pretpostavlja da se skup za učenje sastoji isključivo od normalnih podataka. Problem nedostatka anomalnih primjera prilikom učenja rješava se zašumljivanjem normalnih podataka, odnosno prolaskom kroz unaprijedni difuzijski proces. Zašumljivanje podatka možemo tumačiti kao udaljavanje podataka od mnogostrukosti koja odgovara normalnim podacima, pri čemu će veći vremenski koraci difuzije proizvesti udaljenije podatke. Ako zašumljenim primjerima uspješno pokrijemo čitav prostor značajki, očekujemo da smo onda pokrili i većinu potencijalnih anomalija. Postepeno zašumljivanje normalnih primjera iz mnogostrukosti podataka s ciljem simuliranja potencijalnih anomalija vizualiziramo ilustracijom 2.4.

Uzmimo da je $\mathbf{x}_s \in \mathbb{R}^d$ nastao unaprijednim difuzijskim procesom iz jednadžbe 2.33. Ako pretpostavimo da se cijeli skup podataka sastoji od samo jedne točke koja se nalazi u ishodištu, dolazimo do idućeg izraza za aposteriornu distribuciju preko σ_t^2 uvjetovanu s $\mathbf{x}_s$:

$$p(\sigma_t^2|\mathbf{x}_s) = \frac{p(\sigma_t^2|\mathbf{x}_s)p(\sigma_t^2)}{p(\mathbf{x}_s)} \propto p(\mathbf{x}_s|\sigma_t^2)p(\sigma_t^2) = \mathcal{N}(\mathbf{x}_s; \mathbf{0}, \boldsymbol{\sigma_t^2}) \propto \boldsymbol{\sigma_t^{-d}} \exp\left(-\frac{\|\mathbf{x_s}\|^{\mathbf{2}}}{\mathbf{2}\boldsymbol{\sigma_t^2}}\right) \tag{2.40}$$



Slika 2.4: Konceptualna vizualizacija unaprijednog difuzijskog procesa za simulaciju anomalija. Zelene točke predstavljaju normalne podatke, a crvene anomalne. Plava linija označava granicu mnogostrukosti distribucije normalnih podataka. x_i je primjer x_0 sa šumom koji odgovara i-tom vremenskom koraku difuzije. Zbog jednostavnosti sve je spušteno u dvije dimenzije. Tijekom učenja, postepenim dodavanjem šuma normalni podatak x_0 "izbacujemo" iz mnogostrukosti u kojem leže ostali normalni podaci do prostora u kojem se nalaze anomalije.

Ovaj izraz odgovara gustoći vjerojatnosti inverzne gama distribucije

$$p(\sigma_t^2; a, b) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\sigma_t^2}\right)^{a+1} \exp\left(-\frac{b}{\sigma_t^2}\right) \tag{2.41}$$

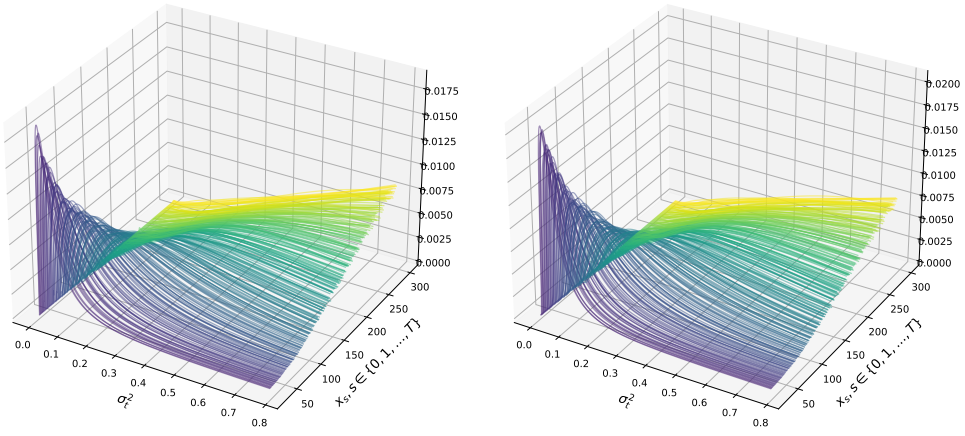
s parametrima $a = \frac{d}{2} - 1$ i $b = \frac{\|\mathbf{x}_s\|^2}{2}$. Ako umjesto jednog podatka imamo cijeli skup podataka $\mathcal{D}$ koji se ravna po distribuciji $p(\mathbf{x})$, onda analitički procjenitelj uvjetne aposteriorne distribucije postaje

$$p(\sigma_t^2|\mathbf{x}_s) \propto p(\mathbf{x}_s|\sigma_t^2)p(\sigma_t^2) = \sum_{\mathbf{x}_0} p(\mathbf{x}_s|\mathbf{x}_0, \sigma_t^2)p(\mathbf{x}_0) = \sum_{\mathbf{x}_0 \in \mathcal{D}} \mathcal{N}(\mathbf{x}_s; \mathbf{x}_0, \sigma_t^2\mathbf{I}) \tag{2.42}$$

Ovaj izraz za uvjetnu aposteriorne distribucije preko σ_t^2 možemo interpretirati kao zbrajanje izglednosti Gaussovih distribucija sa vremenski ovisnim varijancama centriranih u svakoj točki skupa podataka.

Autori u [44] zatim izvode tri različita modela za pronalaženje anomalija procjenom difuzijskog vremena. Najprije, uvrštavanjem izraza za gustoću vjerojatnosti Gaussove distribucije i uz nekoliko aproksimacija dolaze do idućeg izraza:

$$p(\sigma_t^2|\mathbf{x}_s) \propto \sigma_t^{-d} \exp\left(-\frac{1}{\sigma_t^2} \min_{\mathbf{x}_0 \in \mathcal{D}} \frac{\|\mathbf{x}_s - \mathbf{x}_0\|^2}{2}\right) \tag{2.43}$$



(a) Analitički posterior $p(\sigma_t^2|x_s)$ **(b)** Neparametarska procjena $p(\sigma_t^2|x_s)$

Slika 2.5: Aposteriorna distribucija difuzijskog vremena $p(\sigma_t^2|\mathbf{x}_s)$ iz [44]. $\mathbf{x}_s$ je podatak nastao unaprijednim difuzijskim procesom od $\mathbf{x}_0$ prolazeći kroz vremenske korake $s \in \{1, \ldots, T\}$. (a) prikazuje analitički izračunatu aposteriornu distribuciju jednadžbom 2.42. (b) prikazuje neparametarski model aposteriorne distribucije izračunat jednadžbom 2.44 uz $k = 32$. Aposteriorne distribucije su uprosječene preko cijelog skupa podataka. Vidljivo je da točke $\mathbf{x}_s$ koje odgovaraju većem vremenskom koraku u prosjeku imaju veću srednju vrijednost aposteriorne distribucije, što znači da ih možemo identificirati kao točke koje su jako udaljene od mnogostrukosti na kojoj leže normalni podaci. Ilustracija preuzeta iz [44].

To ponovno odgovara inverznoj gama distribuciji, ovaj put s parametrom $a = \frac{d}{2} - 1$ koji ovisi samo o dimenzionalnosti podataka i $b = \min_{\mathbf{x}_0 \in \mathcal{D}} \frac{\|\mathbf{x}_s - \mathbf{x}_0\|^2}{2}$ koji ovisi o udaljenosti zašumljenog podatka $\mathbf{x}_s$ do najbližeg podatka u skupu. Nadalje, tvrde da se empirijski pokazalo da računanje parametra b na temelju prosjeka udaljenosti do k najbližih susjeda radi bolje u praksi. Konačni izraz postaje:

$$p(\sigma_t^2|\mathbf{x}_s) \propto \sigma_t^{-d} \exp\left(-\frac{1}{\sigma_t^2} \cdot \frac{1}{K} \sum_{\mathbf{x}_0 \in \text{kNN}(\mathbf{x}_s)} \frac{\|\mathbf{x}_s - \mathbf{x}_0\|^2}{2}\right) \tag{2.44}$$

Jednadžba 2.44 služi kao *neparametarski model* za pronalaženje anomalija. Mjera anomalnosti se računa kao prosjek distribucije $p(\sigma_t^2|\mathbf{x}_s)$ preko difuzijskog vremena. Neparametarski model funkcionira vrlo slično kao klasična metoda k najbližih susjeda koju smo opisali u potpoglavlju 2.2.1. Štoviše, neparametarski model proizvodi isto rangiranje mjera anomalnosti kao i kNN. Na ilustraciji 2.5 možemo usporediti izgled izravno analitički izračunate inverzne gama distribucije [27] s njenom neparametarskom procjenom na skupu podataka `vertebral`.

Neparametarski model postaje memorijski i računalno skup kada radimo s velikim

skupovima podataka. Iz tog razloga autori predlažu *inverzni gama model* koji aproksimira parametar b dubokom neuronskom mrežom $f_{\boldsymbol{\theta}}$ koja na ulazu prima zašumljeni primjer $\mathbf{x}_t$. Mreža se uči minimizacijom negativne log-izglednosti:

$$\mathcal{L}(\boldsymbol{\theta}) := -\mathbb{E}_{t,\mathbf{x}_0}\left[a \log f_{\boldsymbol{\theta}}(\mathbf{x}_t) - (a+1)\log \sigma_t^2 - \frac{f_{\boldsymbol{\theta}}(\mathbf{x}_t)}{\sigma_t^2}\right] \tag{2.45}$$

Mjera anomalnosti u inverznom gama modelu se računa kao mod inverzne gama distribucije.

Korištenje neuronske mreže na način kao u inverznom gama modelu donekle ograničava njenu ekspresivnost, jer se koristi isključivo za računanje parametra predefinirane distribucije. Stoga autori predlažu *kategorički model* koji modelira difuzijsko vrijeme kao kategoričku distribuciju preko T razreda, gdje je T duljina unaprijednog difuzijskog Markovljevog procesa. Ovaj pristup je fleksibilniji jer prepušta neuronskoj mreži da modelira cijelu distribuciju vremenskih koraka za dani primjer. Učenje se provodi minimizacijom gubitka unakrsne entropije

$$\mathcal{L}(\boldsymbol{\theta}) := \mathbb{E}_{t,x_0}\left[-\sum_{k=0}^{K} \mathbf{y}_t^{(k)} \log\left(f_{\boldsymbol{\theta}}(\mathbf{x}_t)^{(k)}\right)\right] \tag{2.46}$$

pri čemu $\mathbf{y}_t \in \{0,1\}^T$ predstavlja jednojedinični (engl. *one-hot*) vektor koji ima jedinicu na koordinati t, a $f_\theta : \mathcal{X} \to [0,1]^T$ duboku neuronsku mrežu koja na izlazu daje vjerojatnosti za svaki od T razreda. U praksi, zadatak se pojednostavljuje tako da se umjesto T razreda koristi $B < T$ pretinaca. Pripadajući spremnik za razred t se računa kao $\lfloor \frac{t \cdot B}{T} \rfloor$. Mjera anomalnosti se računa kao prosjek predviđene kategoričke distribucije preko svih pretinaca.

Autori u [44] provode detaljnu analizu ovih modela na skupovima podataka iz natjecanja (engl. *benchmark*) ADBench [27]. ADBench se sastoji od niza tabličnih skupova podataka za pronalaženje anomalija. Pokazuju da DTE postiže kompetitivne performanse u usporedbi s ostalim metodama za pronalaženje anomalija. Štoviše, DTE nadmašuje klasične varijacijske difuzijske modele (DDPM) uz nekoliko redova veličine kraće vrijeme zaključivanja.

3. Skupovi podataka

3.1. Skup podataka UBnormal

UBnormal [2] je skup podataka namijenjen za pronalaženje anomalnih radnji u videozapisima. Sastoji se od 29 virtualnih scena s 236902 video okvira.

Prvi je potpuno nadzirani skup podataka za pronalaženje anomalija u videozapisima. Abnormalne radnje su označene na razini piksela, što omogućuje učenje i evaluiranje metoda koje osim prepoznavanja rade i lokalizaciju anomalnih radnji [22].

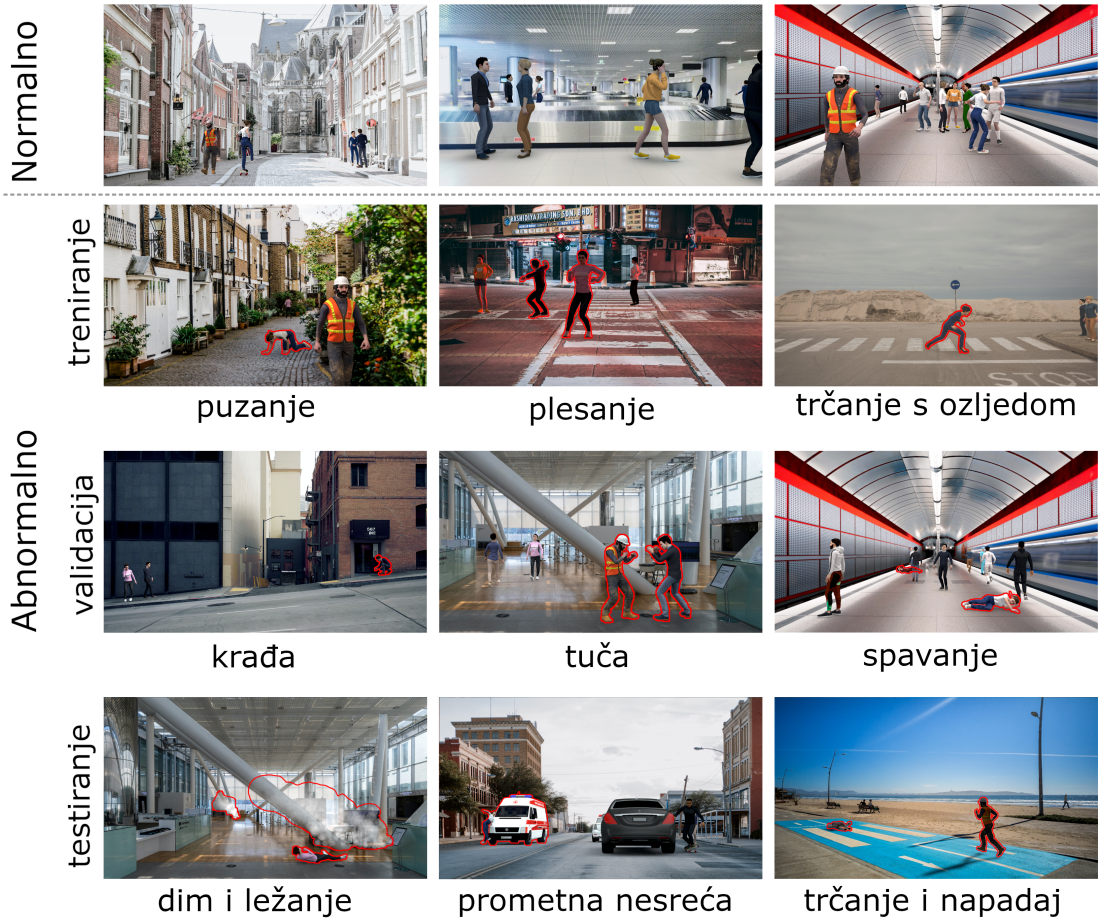
Postoje dva dominantna pristupa pronalaženju anomalija u videozapisima. Prvi formulira pronalaženje anomalija kao polunadzirani ili nenadzirani zadatak, bez oznaka u skupu za učenje. Metode uče model normalnosti, a sve što odstupa od toga naziva se anomalijom. Skup anomalija u ovoj formulaciji je otvoren (engl. *open-set*) jer metoda prilikom učenja ne zna kakve se sve anomalije mogu kasnije pojaviti. Nedostatak ovih metoda to što najčešće imaju lošije performanse od nadziranih, jer nemaju gotovo nikakva saznanja o vrsti anomalnih radnji. Drugi pristup formulira pronalaženje anomalija kao nadzirani zadatak koji ima označene anomalije u skupu za učenje. Pritom skupovi za učenje i testiranje sadrže iste tipove anomalija. Problem pronalaženja anomalija se onda svodi na prepoznavanje zatvorenog skupa radnji (engl. *closed-set*). Nedostatak ovog pristupa je to što ne testira generalizacijske sposobnosti metode na neviđenim anomalijama.

UBnormal se nosi s navedenim problemima tako što koristi disjunktne vrste anomalnih radnji u skupovima za učenje i validaciju u odnosu na skup za testiranje. Drugim riječima, UBnormal po prvi put pristupa pronalaženju anomalija u videozapisima kao nadziranom problemu nad otvorenim skupom anomalija (engl. *supervised open-set setting*). To potiče razvoj nadziranih metoda za pronalaženje anomalija koje će znati iskoristiti informacije o poznatim tipovima anomalija, ali će i dalje pouzdano raditi kad se susretnu s do tada neviđenim vrstama anomalnih radnji. Ovime također otvaraju mogućnost pravednije usporedbe dosadašnjih nenadziranih i polunadziranih metoda nad otvorenim skupom anomalija s nadziranim pristupima nad zatvorenim skupom

anomalija.

Videozapisi iz UBnormal su računalno generirani programom Cinema4D koji omogućuje kreiranje raznih 3D animacija na 2D pozadinama. Korištenje virtualnih scena olakšava prikupljanje raznovrsnih primjera anomalnih radnji, poput nasilja i opasnih kretnji, koje bi inače bile zahtjevne za prikupiti iz pravnih i etičkih razloga. Potencijalni nedostatak korištenja virtualnih videozapisa je pitanje hoće li se takav model moći uspješno primijeniti i na videozapise iz stvarnog svijeta. Da bi donekle odgovorili na to pitanje, autori su u [2] naučili CycleGAN [81] model za prebacivanje videozapisa iz virtualne domene skupa UBnormal u domenu stvarnih videozapisa skupova Avenue [46] i ShanghaiTech [48]. Tako dobivene podatke su ukomponirali u postupak učenja dotadašnjeg najboljeg modela za pronalaženje anomalija u skupovima Avenue i ShanghaiTech i empirijski utvrdili poboljšanje performansi modela.

Kvantitativna raspodjela podataka na skupove za učenje, validaciju i testiranje, kao i ostale statistike vezane uz skup podataka UBnormal vidljivi su u tablicama 3.1 i 3.2. Primjeri normalnih i anomalnih radnji prikazani su na ilustraciji 3.1.



Slika 3.1: Primjeri normalnih i abnormalnih radnji u različitim scenama skupa UBnormal. Crvenim obrubom lokalizirani su primjeri raznih anomalnih radnji, poput puzanja, tuče, spavanja, trčanja. Vrste anomalnih radnji u skupu za testiranje razlikuju se od onih u skupovima za učenje i validaciju. Ilustracija je uz manje preinake preuzeta iz [2].

Tablica 3.1: Usporedba statistika iz skupova podataka UBnormal i ShanghaiTech s ostalim popularnim skupovima za pronalaženje anomalija u videozapisima. ShanghaiTech se ističe kao skup s najvećim brojem okvira (engl. *frame*). Jedino ih skup UCF-Crime ima više, ali on sadrži zatvoreni skup anomalnih radnji namijenjen nadziranom učenju, stoga ga izuzimamo iz usporedbe.

	Broj okvira					
Skup podataka	**ukupno**	**učenje**	**validacija**	**testiranje**	**normalni**	**abnormalni**
CUHK Avenue [46]	30,652	15,328	-	15,324	26,832	3,820
Street Scene [59]	203,257	56,847	-	146,410	159,341	43,916
Subway Entrance [3]	144,250	76,453	-	67,797	132,138	12,112
Subway Exit [3]	64,901	22,500	-	42,401	60,410	4,491
UCF-Crime [69]	13,741,393	12,631,211	-	1,110,182	-	-
UCSD Ped1 [50]	14,000	6,800	-	7,200	9,995	4,005
UCSD Ped2 [50]	4,560	2,550	-	2,010	2,924	1,636
ShanghaiTech [48]	317,398	274,515	-	42,883	300,308	17,090
UBnormal [2]	236,902	116,087	28,175	92,640	147,887	89,015

Tablica 3.2: Usporedba statistika iz skupova podataka UBnormal i ShanghaiTech s ostalim popularnim skupovima za pronalaženje anomalija u videozapisima. Skup UBnormal se ističe velikim brojem raznovrsnih anomalija.

Skup podataka	**Anomalije**	**Scene**	**Vrste anomalija**	**Otvoreni skup**
CUHK Avenue [46]	77	1	5	✓
Street Scene [59]	205	1	17	✓
Subway Entrance [3]	51	1	5	✓
Subway Exit [3]	14	1	3	✓
UCF-Crime [69]	-	-	13	✗
UCSD Ped1 [50]	61	1	5	✓
UCSD Ped2 [50]	21	1	5	✓
ShanghaiTech [48]	158	13	11	✓
UBnormal [2]	660	29	22	✓

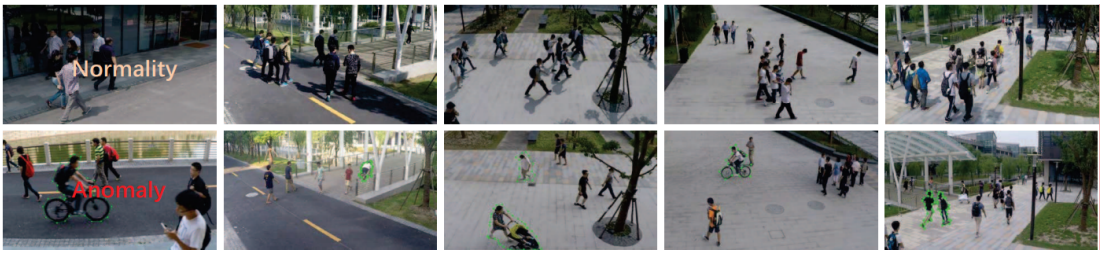
3.2. Skup podataka ShanghaiTech

ShanghaiTech [48], punog imena ShanghaiTech Campus, je najveći skup s više scena za pronalaženja otvorenog skupa anomalija u videozapisima. Sastoji se od 13 scena s

ukupno 317398 okvira.

Nedostatak starijih skupova podataka za pronalaženje anomalija u videozapisima je nedostatak varijacije u scenama i kutu kamere koja snima scenu. Većina starijih skupova su snimljeni jednim fiksnim kutem kamere. Stoga, ShanghaiTech uvodi veći broj scena snimljenih pod različitim kutevima i zahtjevnim svjetlosnim uvjetima. Osim toga, po prvi put uvode anomalije uzrokovane brzom kretnjom objekata kroz scenu. Ti dodaci čine ShanghaiTech pogodnijim za primjene u stvarnim sustavima.

Statistike za usporedbu s UBnormal i ostalim popularnim skupovima podataka za pronalaženje anomalija u videozapisima prikazane su u tablicama 3.1 i 3.2. Primjeri normalnih i anomalnih okvira iz skupa ShanghaiTech vidljivi su na slici 3.2.



Slika 3.2: Primjeri normalnih (prvi red) i abnormalnih (drugi red) radnji iz skupa ShanghaiTech. Zelenim obrubom lokalizirani su primjeri raznih anomalnih radnji. Veliki udio anomalija u ShanghaiTech čini brzo kretanje osobe kroz scenu na nekoj vrsti vozila, najčešće biciklu, kao što se vidi u drugom redu. Vidljiva je i velika raznolikost u kutevima kamere i osvjetljenju scene, što je jedno od dobrih svojstava ovog skupa podataka. Ilustracija je preuzeta iz [48].

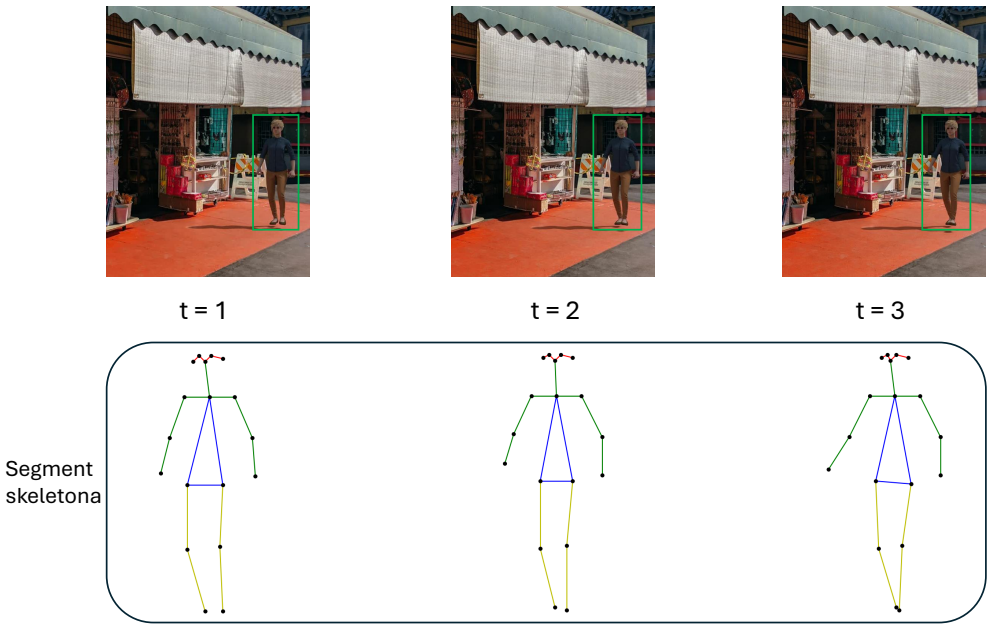
4. Detekcija anomalija u vremenskim sljedovima skeletona modeliranjem difuzijskog vremena

U ovom radu bavimo se pronalaženjem anomalnih ljudskih radnji u videozapisima. Cilj je razviti model koji će za svaki okvir u danom videozapisu odrediti događa li se u sceni anomalija ili normalni događaj. Općenitije, želimo izračunati mjeru anomalnosti za svaki okvir u videozapisu.

Model ne učimo izravno na pikselima već na skeletonima. Skeleton je skup koordinata ključnih točaka na slici ljudskog tijela. Primjeri ključnih točaka su zglobovi, oči, uši. Ako modelu na ulaz dovodimo samo jedan po jedan okvir odnosno skeleton iz videa, ograničavamo ga na korištenje isključivo prostornih informacija, poput relativnog pomaka koljena i glave u nekom trenutku. Ono što želimo je da model osim prostornih koristi i pomake ključnih točaka kroz vrijeme kako bi prepoznao pokret koji predstavlja anomalnu radnju. Stoga, u fazama učenja i zaključivanja radimo s *segmentom skeletona*, odnosno nizom skeletona iz nekoliko uzastopnih okvira u videu. Primjer jednog segmenta dan je na slici 4.1.

Anomalije pronalazimo modelom procjene difuzijskog vremena (DTE) kojeg smo detaljno opisali u potpoglavlju 2.2.2. DTE se pokazao kao zanimljiv i efikasan pristup pronalaženju anomalija u tabličnim skupovima podataka. Njegova primjena za pronalaženje anomalija u videozapisima je, koliko znamo, do sada neistražena. Glavna ideja rada je, stoga, naučiti DTE model za pronalaženje anomalija na segmentima skeletona i evaluirati njegovu uspješnost. U tu svrhu eksperimentiramo s kategoričkim i inverznim gama modelom za procjenu difuzijskog modela te različitim arhitekturama.

Korišteni skupovi podataka su UBnormal (opisan u poglavlju 3.1) i ShanghaiTech (opisan u poglavlju 3.2). Oba skupa koristimo isključivo za polunadzirano učenje, odnosno pristupamo pronalaženju anomalija kao klasifikaciji s jednim poznatim razredom (engl. *One-Class Classification, OCC*). To znači da se skup za učenje sastoji samo



Slika 4.1: Segment skeletona duljine 3. U prvom redu prikazana su 3 uzastopna okvira iz videozapisa koji pripada skupu podataka UBnormal [2]. Zelenim pravokutnikom označena je osoba u sceni. Okviri se zasebno dovode na ulaz modela za izvlačenje ključnih točaka na osobama. Izlaz modela je procjena poze osobe za svaki okvir u obliku skeletona, koje zatim konkateniramo po novoj dimenziji u jedan zajednički segment skeletona (drugi red ilustracije).

od segmenata skeletona na kojima se odvija normalna radnja. Skupovi za validaciju i testiranje uz normalne sadrže i abnormalne segmente skeletona.

4.1. Prethodne metode

U posljednje vrijeme vidljiv je značajan odmak od tradicionalnih plitkih modela prema dubokim modelima za pronalaženje anomalija u videozapisima. Ovisno o korištenoj mjeri anomalnosti, većinu modela možemo razvrstati u dva glavna smjera: modeli koji koriste rekonstrukcijski gubitak i modeli zasnovani na izglednosti.

MoCoDAD (engl. *Multimodal Motion Conditioned Diffusion Model for Skeleton-based Video Anomaly Detection*) [18] je primjer rekonstrukcijskog pristupa pronalaženju anomalija u segmentima skeletona. Autori primjećuju da dosadašnje metode često ograničavaju latentne reprezentacije normalnih podataka u male prostore i sve izvan prostora normalnih podataka proglašavaju anomalijama. Na taj način uvažavaju raznolikost tipova anomalnih radnji koje se mogu pojaviti. Međutim, normalne rad-

nje također dijele ovo svojstvo raznolikosti. Zanemarivanje te činjenice često dovodi do pogrešne klasifikacije rijetkih normalnih radnji kao anomalnih. Stoga, MoCoDAD predlaže generativni model koji pretpostavlja da su i normalne i anomalne radnje multimodalno distribuirane. Odabiru varijacijske difuzijske modele zbog njihove odlične sposobnosti modeliranja distribucija s više modova [76]. Ulazni segment skeletona dijele na dva dijela, prošli i budući. Budući segment zašumljuju unaprijednim difuzijskim procesom. Zatim unatražnim procesom generiraju više mogućih rekonstrukcija, koje statistički agregiraju. Prošli segment enkoderom pretvaraju u latentnu reprezentaciju kojom dodatno uvjetuju difuzijski model prilikom računanja unatražnih prijelaza. Kao okosnicu (engl. *backbone*) za difuziju koriste U-Net [62] sa slojevima prostorno-vremenski razdvojivih konvolucija nad grafovima [67].

STG-NF (engl. *Normalizing Flows for Human Pose Anomaly Detection*) [29] uče model normalizirajućeg toka za pronalaženje anomalija u segmentima skeletona s kraja na kraj. Prilikom učenja, model uči bijektivno preslikavanje iz distribucije segmenata skeletona u latentnu Gaussovu distribuciju. Treniranje se provodi minimizacijom negativne log-izglednosti podataka. U afinim transformacijskim slojevima koriste prostorno-vremenske konvolucije nad grafovima [77] kako bi izvukli značajne reprezentacije iz segmenata skeletona. Zaključivanje se provodi transformacijom segmenta u latentni prostor, nakon čega se primjenjuje formula za zamjenu varijabli kako bi se izračunala izglednost podatka u prostoru segmenata. Što je izglednost manja, to se primjer više smatra anomalijom. STG-NF podržava i nadzirani način učenja u slučajevima kada su primjeri anomalnih radnji prisutni u skupu za učenje. Tada za apriornu distribuciju latentnih primjera koriste model Gaussove mješavine (engl. *Gaussian mixture*).

Naš model također pronalazi anomalije na segmentima skeletona. Također, po uzoru na ova dva modela i mi uključujemo slojeve konvolucija na grafovima u naš model. Poput MoCoDAD-a, mi također zašumljujemo segmente skeletona unaprijednim difuzijskim procesom. Razlika je u tome što mi ne učimo unatražni proces, već aproksimiramo rekonstrukcijski gubitak procjenom difuzijskom vremena.

Ostale relevantne metode koje detektiraju anomalne akcije na razini skeletona su [34, 51, 23, 73, 13, 21, 19, 7, 11, 66, 43, 77].

4.2. Učitavanje podataka

Na slici 4.1 prikazali smo kako izgledaju podaci kojima se bavimo u ovom radu. Po uzoru na [29], koristimo AlphaPose [16] model s YOLOX [20] detektorom za pronala-

ženje osoba u okvirima videozapisa i izvlačenje skeletona. Rezultati detekcije pohranjuju se u datoteke koje koristimo prilikom učitavanja podataka. Za svaki videozapis iz originalnog skupa podataka sprema se po jedna datoteka s podacima o skeletonima. Datoteke su strukturirane kao rječnik s dva ključa. Prvi ključ indeksira osobu, a drugi okvir u videozapisu. Vrijednost u rječniku je niz od 17 ključnih točaka, pri čemu je svaka ključna točka opisana 3 s koordinate. Uz dvije prostorne koordinate, treća koordinata predstavlja pouzdanost detektora za tu ključnu točku. Eksperimentalno smo utvrdili da modeli postižu bolje rezultate bez nje, stoga je izostavljamo prilikom učitavanja podataka i koristimo isključivo prostorne koordinate x i y. Skeletone pretvaramo u COCO [41] format, što znači da uz postojećih 17 točaka dodajemo i 18. koja označava vrat osobe, a računa se kao prosjek ključnih točaka lijevog i desnog ramena.

Dva su hiperparametra učitavanja podataka: duljina segmenta i korak (engl. *stride*). Duljina segmenta određuje koliko uzastopnih okvira spajamo u segment. Korak određuje pomake kojima se klizeće okno (engl. *sliding window*) "kreće" po videozapisu prilikom generiranja segmenata. Ovaj hiperparametar koristimo samo prilikom učitavanja skupa za treniranje. U skupovima za testiranje i validaciju korak uvijek postavljamo na 1. Jedna osoba ne mora biti prisutna u svakom okviru videa. Moguće je čak da u nekim okvirima osoba izađe iz scene pa se u kasnijim okvirima ponovo vrati. Stoga, prilikom stvaranja segmenata pazimo da su kontinuirani, odnosno da svaki okvir iz segmenta sadrži skeleton osobe. Proces stvaranja segmenata klizećim oknom prikazan je ilustracijom 4.2.

Nakon učitavanja, koordinate ključnih točaka najprije skaliramo na vrijednosti između 0 i 1. Potom normaliziramo segment na srednju vrijednost 0 i jediničnu varijancu. Kad učitavamo skup za učenje, na normalizirani segment primjenjujemo afinu transformacijom. Afinim transformacijama stvaramo nove podatke u kojima su ključne točke blago pomaknute, čime nastojimo povećati robusnost modela. Korištene transformacije su zrcaljenje, pomak i nakošenje (engl. *shear*).

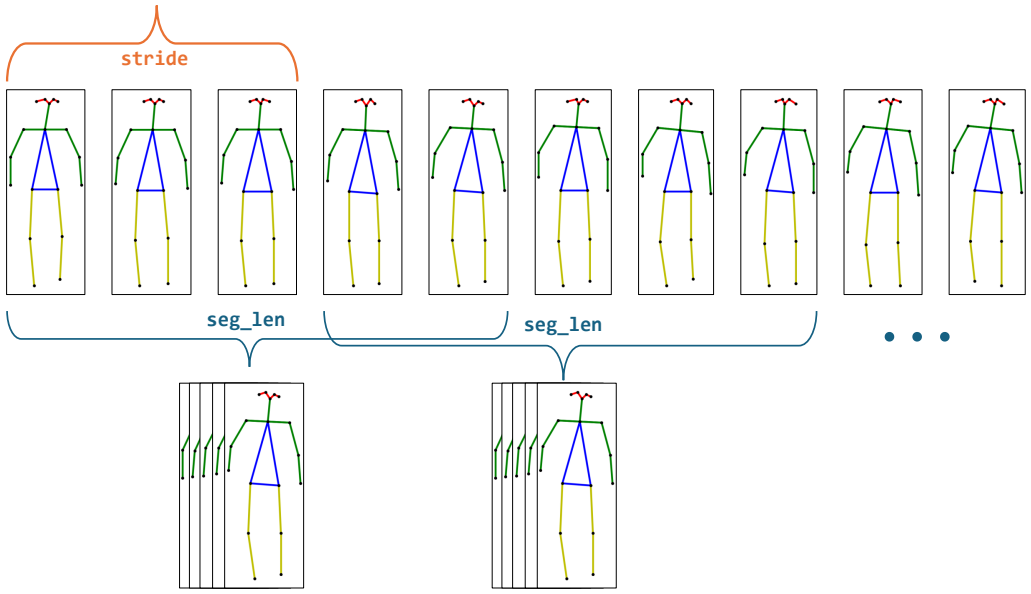
Nakon učitavanja, segmenti su pohranjeni u memoriji kao tenzori oblika (`batch_size`, 2, `seg_len`, 18), pri čemu je `batch_size` hiperparametar koji određuje veličinu minigrupe, a `seg_len` duljina segmenta skeletona. Cijeli postupak učitavanja podataka najbolje je prikazan sljedećim algoritmom:

Algoritam 4 Učitavanje segmenata

Ulaz: `seg_len`, `train_stride`, `split` ("train", "test" ili "val")

Izlaz: Segmenti skeletona oblika (N, 2, `seg_len`, 18)

Inicijaliziraj $\mathcal{D} = [\varnothing]$
Za svaki videozapis v **radi**
    Za svaku osobu o koja se pojavljuje u videozapisu v **radi**
        Učitaj iz datoteke sve skeletone iz videozapisa v koji pripadaju osobi o
        `stride` $=$ `train_stride` ako `split` $=$ "train" inače 1
        Za svaki skeleton koji pripada osobi o, s korakom `stride` **radi**
            Ako postoji idućih `seg_len` okvira **onda**
                Konkateniraj `seg_len` uzastopnih okvira počevši od trenutnog u segment **seg**
                Normaliziraj segment **seg** tako da ima srednju vrijednost 0 i jediničnu varijancu
                Ako `split` $=$ "train" **onda**
                    Primijeni afinu transformaciju na segment **seg**
                Dodaj segment **seg** u $\mathcal{D}$
Vrati $\mathcal{D}$



Slika 4.2: Stvaranje segmenata klizećim oknom. Parametar `seg_len` određuje veličinu okna, a parametar `stride` iznos za koji se okno pomiče u svakom koraku. U ovom primjeru `seg_len` iznosi 5, a `stride` 3.

4.3. Osnovni model

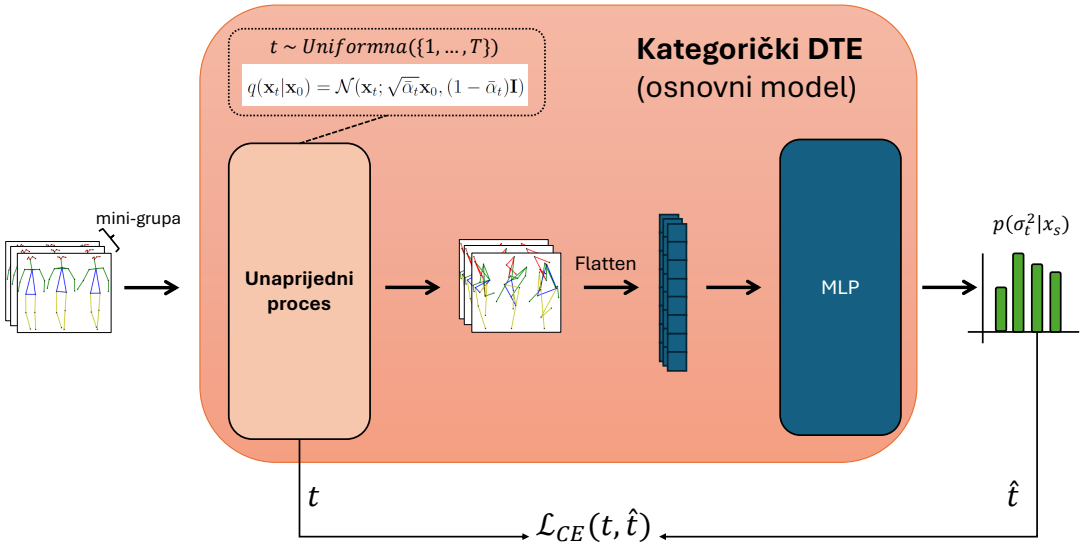
Kao osnovni model (engl. *baseline*) prilagođavamo DTE [44] iz poglavlja 2.2.2 na najjednostavniji mogući način za rad sa segmentima skeletona. "Peglamo" sve dimenzije segmenata osim prve, odnosno spremamo podatke u tenzor oblika (N, D), gdje je N broj primjeraka u skupu podataka, a $D = 2 \cdot 16 \cdot \texttt{seg\_len}$. Time efektivno tretiramo skup segmenata skeletona kao tablični skup podataka, za koje je DTE originalno i implementiran.

Eksperimentiramo s kategoričkim i inverznim gama modelom na skupovima UBnormal i ShanghaiTech. Tijekom učenja, zašumljujemo segmente unaprijednim difuzijskim procesom. Postupak zašumljivanja skeletona prikazan je na slikama 4.4 i 4.5. Kao i DTE, oslanjamo se na zanimljivu sposobnost difuzijskog procesa da "popunjava" prostor. Podsjetimo se, zašumljivanjem podatak udaljavamo od mnogostrukosti na kojoj leže normalni podaci. Klasični difuzijski modeli tada uče rekonstrukciju podatka natrag na mnogostrukost. DTE umjesto toga procjenjuje difuzijsko vrijeme koje korespondira udaljenosti zašumljenog primjera od mnogostrukosti. Pritom pretpostavljamo da se anomalije nalaze izvan mnogostrukosti normalnih primjera, kao na slici

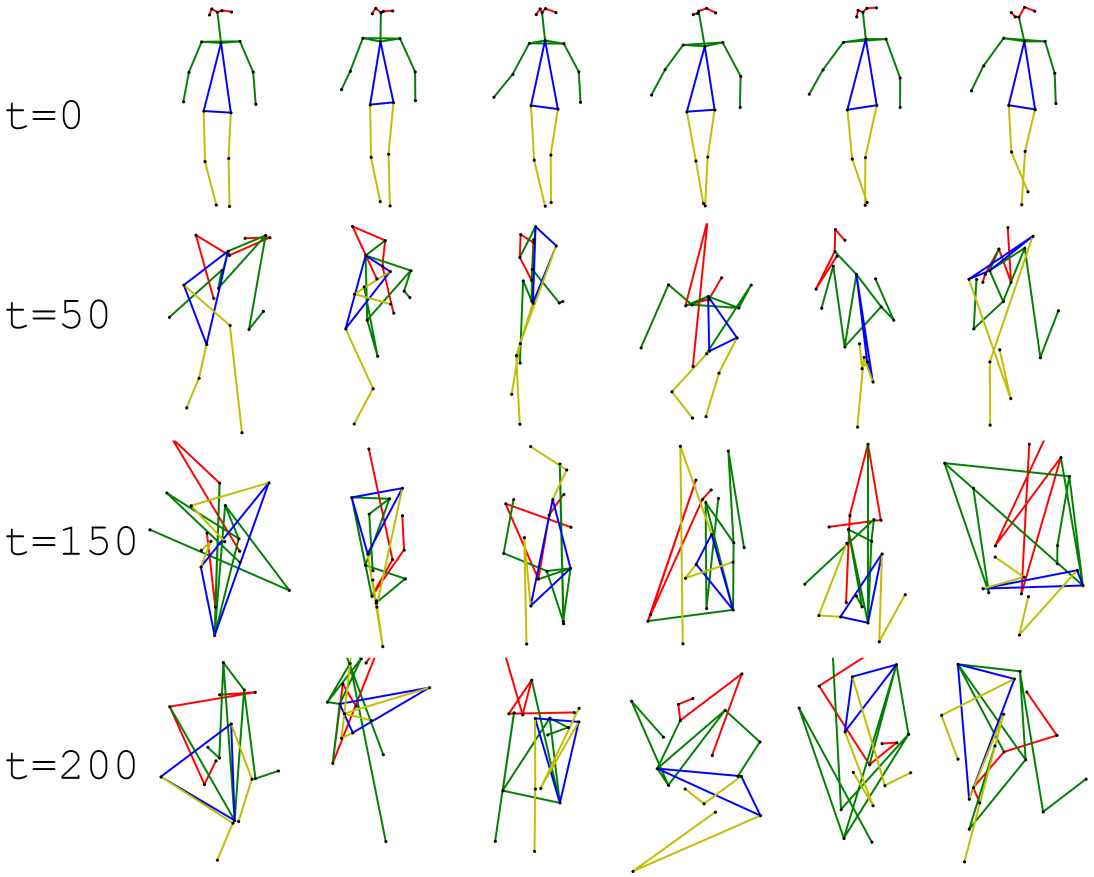
2.4. Stoga, nadamo se da ćemo zašumljivanjem normalnih segmenata pokriti dobar dio prostora u kojem se nalaze abnormalni segmenti, odnosno segmenti skeletona koji predstavljaju anomalnu radnju.

Ponovimo, kategorički model koristi duboki neuronsku mrežu $f_{\boldsymbol{\theta}}$ za procjenu parametara kategoričke distribucije difuzijskog vremena. Inverzni gama model mrežom $f_{\boldsymbol{\theta}}$ predviđa parametar b inverzne gama distribucije kojom modeliramo difuzijsko vrijeme. Po uzoru na [44], koristimo jednostavni višeslojni perceptron (engl. *Multilayer Perceptron, MLP*) za aproksimaciju funkcije $f_{\boldsymbol{\theta}}$.

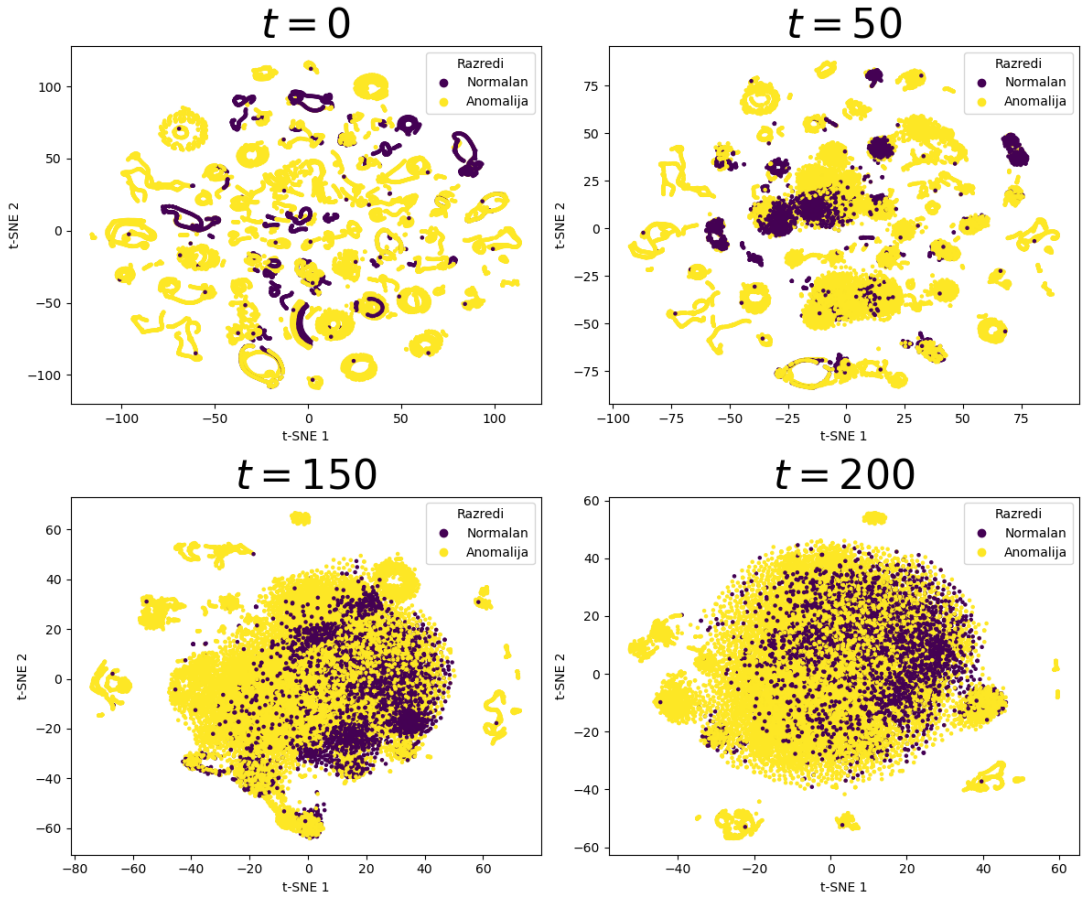
MLP se sastoji od niza slojeva potpuno povezanih slojeva. Nakon svakog potpuno povezanog sloja podaci prolaze kroz aktivacijsku funkciju zglobnicu (engl. *Rectified Linear Unit, ReLU*). U svim skrivenim slojevima koristi se isključivanje neurona (engl. *dropout*) s parametrom $p = 0.5$ za regularizaciju. Izlazna dimenzija zadnjeg sloja je jednaka 1 za inverzni gama model, a u slučaju kategoričkog modela odgovara broju klasifikacijskih pretinaca. Dodatno, u kategoričkom modelu izlaz posljednjeg potpuno povezanog sloja dovodimo na ulaz funkcije `softmax` da izlaz modela možemo interpretirati kao distribuciju.



Slika 4.3: Arhitektura za učenje osnovnog modela. Okosnica DTE modela je višeslojna potpuno povezana mreža (MLP). Zbog toga se zašumljeni segmenti izravnaju (engl. *flatten*) u vektore prije ulaza u okosnicu. Na slici je prikazan kategorički DTE model, što znači da na izlazu predviđamo kategoričku distribuciju difuzijskog vremena i koristimo gubitak unakrsne entropije. S druge strane, izlaz kod inverznog gama modela je parametar b inverzne gama distribucije kojom modeliramo difuzijsko vrijeme, a gubitak je negativna log-izglednost. U fazi zaključivanja arhitektura izgleda identično, osim što podatke ne zašumljujemo nego izravno dovodimo na ulaz okosnice.



Slika 4.4: Unaprijedni difuzijski proces duljine $T = 200$ na segmentu skeletona. Prvi red predstavlja originalni segment skeletona. Svaki idući red je originalni segment u koraku t unaprijednog difuzijskog procesa. Unaprijedni proces provodi se u jednom koraku, prema formuli 2.28.



Slika 4.5: t-SNE ugrađivanja segmenata skeletona provedenih kroz unaprijedni difuzijski proces duljine $T = 200$. Prikazan je samo podskup validacijskog skupa iz UBnormal. Visoko dimenzionalni segmenti skeletona projicirani su u 2 dimenzije metodom t-SNE [72]. Segment ovdje proglašavamo anomalijom ako sadrži barem jedan anomalan okvir. Vidimo da je distribucija posljednjeg koraka unaprijednog procesa jedinična Gaussova, odnosno potpuni šum.

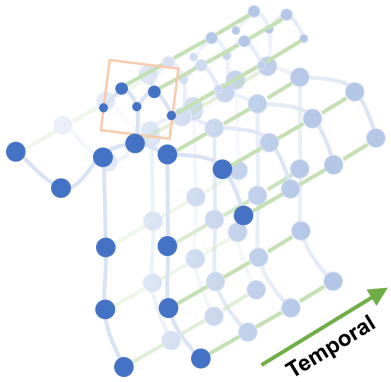
4.4. Proširenje osnovnog modela prostorno-vremenski razdvojivim konvolucijama nad grafovima

U osnovnom modelu funkciju f_θ zbog jednostavnosti aproksimiramo višeslojnim perceptronom. To je ujedno i model koji se pokazao uspješnim u originalnom DTE radu [44]. Međutim, oni rade isključivo s tabličnim podacima i vektorskim reprezentacijama slika, što opravdava takav izbor arhitekture. S druge strane, mi radimo sa segmentima skeletona. Ključne točke osoba su semantički bogati podaci. Promatranjem njihovih odnosa u prostoru i vremenu možemo otkriti korisne informacije o kretanju

osobe kroz scenu. Prepoznavanje uzoraka u segmentima skeletona vrlo je složen zadatak za običnu potpuno povezanu neuronsku mrežu. Idealno, želimo u okosnicu modela za procjenu difuzijskog vremena uvesti induktivnu pristranost prema skeletonskim podacima. To nas prirodno dovodi do neuronskih modela nad grafovima.

4.4.1. Konvolucije nad grafovima

Konvolucijski modeli nad grafovima (engl. *Graph Convolutional Networks, GCN*) mogu se promatrati kao generalizacija klasičnih konvolucijskih modela na susjedstva proizvoljne, ne nužno rešetkaste, strukture. Koristimo ih za izvlačenje korisnih reprezentacija iz grafova. Da bismo ih mogli primijeniti na naše podatke, segmente skeletona interpretiramo kao prostorno-vremenske grafove. Vrhovi grafa su ključne točke kroz sve okvire segmenta. Prostorni bridovi povezuju ključne točke koje pripadaju istom okviru, a vremenski bridovi povezuju ključne točke s istim ključnim točkama u ostalim vremenskim trenucima (slika 4.6).



Slika 4.6: Segment skeletona kao prostorno-vremenski graf. Preuzeto iz [77].

Malo formalnije [67], definiramo graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ s TV vrhova, od kojih svaki predstavlja jednu ključnu točku u jednom okviru segmenta. Bridovi $(i, j) \in \mathcal{E}$ su reprezentirani prostorno-vremenskom matricom susjedstva $A^{st} \in \mathbb{R}^{VT \times VT}$. Ulaz u l-ti sloj konvolucije nad grafom je tenzor $\mathcal{H}^{(l)} \in \mathbb{R}^{C^{(l)} \times V \times T}$. Taj tenzor sadrži skrivenu reprezentaciju dimenzije $C^{(l)}$ za svaku od V ključnih točaka kroz T okvira. Ulaz u prvi sloj je originalni graf opisan x i y koordinatama ključnih točaka, pa vrijedi $C^{(1)} = 2$. Izlaz l-tog sloja konvolucije nad grafom računa se prema idućem izrazu

$$\mathcal{H}^{(l+1)} = \sigma(A^{st-(l)}\mathcal{H}^{(l)}W^{(l)}) \tag{4.1}$$

pri čemu je $A^{st-(l)} \in \mathbb{R}^{VT \times VT}$ prostorno-vremenska matrica susjedstva sloja l, $W^{(l)} \in \mathbb{R}^{C^{(l)} \times C^{(l+1)}}$ parametri koje učimo za projekciju skrivene reprezentacije vrha iz $C^{(l)}$ u

$C^{(l+1)}$ dimenzija, a σ neka nelinearna aktivacijska funkcija.

ST-GCN [77] (engl. *Spatial Temporal GCN*) primjer je metode koja uspješno koristi konvolucije nad grafovima za prepoznavanje ljudske akcije na temelju njegova skeletona. Ova metoda ograničava reprezentacije vrhova na promatranje isključivo prostornih odnosa ključnih točaka, koristeći prostornu matricu susjedstva A^s. Vremenske odnose između korespondentnih ključnih točaka obrađuje klasičnim konvolucijskim slojem s jezgrom dimenzija $T \times T \times 1 \times 1$.

STS-GCN [67] (engl. *Space-Time-Separable GCN*), s druge strane, modelira sve tri moguće vrste interakcija između ključnih točaka: prostor-prostor, vrijeme-vrijeme i prostor-vrijeme. Međutim, ograničava prijenos informacija u prostorno-vremenskim interakcijama faktorizacijom matrice susjedstva na umnožak prostorne i vremenske matrice susjedstva kao u jednadžbi 4.2.

$$\mathcal{H}^{(l+1)} = \sigma(A^{s-(l)} A^{t-(l)} \mathcal{H}^{(l)} W^{(l)}) \tag{4.2}$$

Ovakva formulacija smanjuje broj parametara modela te postiže impresivne rezultate na zadatku predikcije budućih poza. Za potrebe dekodiranja reprezentacija, koriste temporalne konvolucijske mreže (engl. *Temporal Convolutional Networks, TCN*).

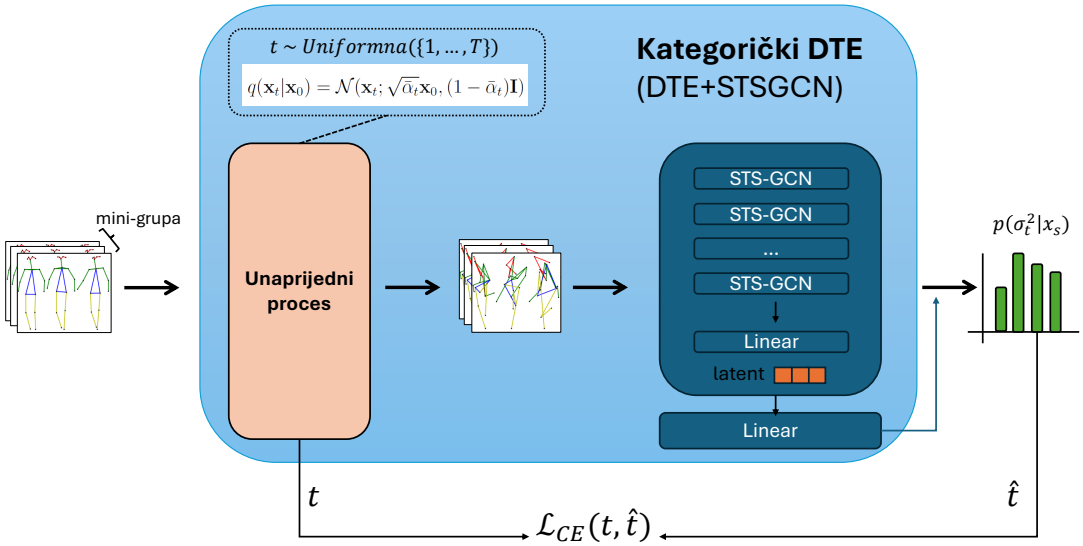
4.4.2. Arhitektura

Okosnicu našeg DTE modela čine dva modula: 1. prostorno-vremensko odvojivi enkoder (engl. *Space-Time-Separable Encoder, STSE*) koji je zadužen za izvlačenje bogate latentne reprezentacije iz segmenta skeletona te 2. potpuno povezani sloj. Potpuno povezani sloj služi kao klasifikacijska glava nakon koje slijedi funkcija `softmax` u slučaju kategoričkog DTE modela. U inverznom gama modelu služi za regresiju parametra b. Enkodirana reprezentacija provodi se kroz zglobnicu (ReLU).

STSE se sastoji od niza STS-GCN slojeva koji izvlači skrivenu reprezentaciju za svaku od TV ključnih točaka. Nakon toga se skrivena reprezentacija $\mathcal{H}$ cijelog segmenta projicira linearnim slojem u latentnu reprezentaciju.

STS-GCN sloj također se sastoji od dva modula: GCN i TCN opisanih u prethodnom potpoglavlju, uz preskočnu vezu i parametarsku zglobnicu (PReLU) kao aktivacijsku funkciju.

Predložena arhitektura DTE modela s okosnicom koja provodi konvoluciju nad grafovima prikazana je na slici 4.7.



Slika 4.7: Predložena arhitektura za procjenu difuzijskog vremena na segmentima skeletona. Za razliku od osnovnog modela sa slike 4.3, ovdje na ulaz okosnice dovodimo segmente skeletona, bez da ih izravnamo u vektore. Segmente skeletona interpretiramo kao prostorno-vremenski graf. Okosnica sadrži slojeve prostorno-vremenski razdvojive konvolucije nad grafovima, popraćene potpuno povezanim slojevima.

4.5. Evaluacija

4.5.1. Evaluacijske metrike

Za potrebe računanja evaluacijskih metrika, zadatak pronalaženja anomalija promatramo kao običnu binarnu klasifikaciju [82]. Uz dani klasifikacijski prag i stvarnu oznaku primjera, ovisno o mjeri anomalnosti, primjer svrstavamo u jednu od četiri kategorije: stvarno pozitivni (engl. *true positive, TP*), lažno pozitivni (engl. *false positive, FP*), lažno negativni (engl. *false negative, FN*) i stvarno negativni (engl. *true negative, TN*). Popularne metrike koje možemo izračunati na temelju tih podataka su preciznost i odziv. Preciznost se definira kao udio stvarno pozitivnih primjera (TP) u skupu svih primjera koje je klasifikator označio kao pozitivne (TP + FP).

$$Preciznost = \frac{TP}{TP + FP} \tag{4.3}$$

Odziv je udio stvarno pozitivnih primjera (TP) u skupu svih pozitivnih primjera (TP + FN). Naziva se još i stopa stvarnih pozitiva.

$$Odziv = TPR = \frac{TP}{TP + FN} \tag{4.4}$$

Odabir klasifikacijskog praga uvelike ovisi o samoj primjeni modela. Poželjno bi bilo imati mjeru koja će evaluirati model za detekciju anomaliju u raznim primjenama, što znači pri raznim vrijednostima klasifikacijskog praga [63]. Jedna takva metrika je AUROC, površina ispod ROC (engl. *Receiver Operating Characteristic*) krivulje. ROC krivulja na x-osi iscrtava stopu lažnog alarma (FPR), a na y-osi stopu stvarnih primjera (TPR) za različite vrijednosti klasifikacijskog praga. Stopa lažnog alarma računa se kao:

$$FPR = \frac{FP}{FP + TN} \tag{4.5}$$

AUROC se računa kao integral ROC krivulje i tako objedinjuje performanse modela preko svih klasifikacijskih pragova. Idealni klasifikator imat će AUROC $= 1$. Zgodno svojstvo AUROC-a je to što će njegova vrijednost uvijek biti 0.5 za slučajni klasifikator (koji radi nasumične predikcije sa šansom od 50% za pozitivnu, odnosno negativu predikciju). To vrijedi čak i kad je jedan razred značajno više zastupljen u skupu odnosu na drugi, što ga čini popularnim izborom za evaluacijsku mjeru u zadatku detekcije anomalija.

Druga popularna mjera je prosječna preciznost (engl. *Average Precision, AP*), koja se računa kao površina ispod PR krivulje (engl. *Precision-Recall Curve*). Na x-osi PR krivulje nalazi se odziv, a na y-osi preciznost izračunati na cijelom skupu za različite vrijednosti klasifikacijskog praga. Slučajni klasifikator u ovom slučaju uvijek postiže AP koji odgovara udjelu anomalija u skupu podataka, što donekle otežava interpretaciju i usporedbu ove metrike u različitim primjenama.

Obe metrike se raširenо koriste u literaturi. U ovom radu rezultate prijavljujemo na metrici AUROC.

4.5.2. Evaluacija korištenih modela

U ovom radu evaluaciju modela provodimo na razini okvira videozapisa. Raspolažemo oznakama za svaki okvir, pri čemu 1 predstavlja anomaliju, a 0 normalnu radnju. Mjeru anomalnosti računamo za svaki segment skeletona. Proces stvaranja segmenata opisan je u poglavlju 4.2.

Logiku za računanje mjera anomalnosti za okvire na temelju postojećih mjera anomalnosti za segmente preuzimamo iz [29]. Iteriramo po svim videozapisima i učitavamo njihove oznake. Zatim inicijaliziramo mjeru anomalnosti za svaki okvir na najmanju moguću vrijednost, $-\infty$. Time se osiguravamo da će se okviri u kojima nije prisutna niti jedna osoba tretirati kao normalne radnje. Nakon toga prolazimo kroz

sve segmente u podacima i okviru koji se nalazi na sredini segmenta pridajemo mjeru anomalnosti cijelog segmenta. U slučajevima kada se više osoba nalazi u istom okviru, konačnu mjeru anomalnosti za taj okvir računamo kao maksimalnu vrijednost s obzirom na sve prisutne osobe. Mjere anomalnosti za sve okvire svih videozapisa spajamo u jedan dugi vektor. To isto radimo i za oznake. Mjere anomalnosti, dodatno, zaglađujemo Gaussovim filtrima s različitim varijancama. Motivacija za to je ublažavanje utjecaja jako velikih ili jako malih iznosa mjere anomalnosti koji odskaču od susjednih iznosa. Radi se često o neobičnim ili lažno detektiranim skeletonima koji su nastali zbog greške detektora ili nekim drugim artefaktima. Vektore mjera anomalnosti i oznaka prosljeđujemo kao parametre funkciji za izračun metrike AUROC.

Postupak evaluacije modela možemo sažeto prikazati sljedećim algoritmom:

Algoritam 5 Evaluacija modela

Ulaz: mjera anomalnosti $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$ za svaki segment skeletona
oznake razreda $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$ za sve okvire u svim videozapisima
metapodaci $\mathcal{M}$ o segmentima (duljina segmenta, indeks početka segmenta u videozapisu)

Izlaz: AUROC

Inicijaliziraj $\mathcal{P} = [\varnothing]$
Za svaki videozapis v **radi**
    $\mathcal{P}_v = \{-\infty, -\infty, \ldots, -\infty\}$ ▷ *Inicijaliziraj mjeru anomalnosti za svaki okvir na* $-\infty$
    Za svaku osobu o koja se pojavljuje u videozapisu v **radi**
        Za svaki segment seg osobe o u videozapisu v **radi**
            Dohvati mjeru anomalnosti $\mathbf{s}$ iz $\mathcal{S}$ za segment seg
            Dohvati indeks i početka segmenta u videozapisu iz $\mathcal{M}$
            Dohvati duljinu segmenta len iz $\mathcal{M}$
            Ako $\mathbf{s} > \mathcal{P}_v\left[i + \frac{len}{2}\right]$ **onda**
                $\mathcal{P}_v\left[i + \frac{len}{2}\right] = \mathbf{s}$ ▷ *Postavi mjeru anomalnosti srednjeg okvira segmenta na mjeru anomalnosti cijelog segmenta*
    $\mathcal{P} = \mathcal{P} \cup \mathcal{P}_v$ ▷ *Proširi ukupnu listu s mjerama anomalnosti trenutnog videozapisa*
$\mathcal{P} = \texttt{gauss\_filter\_1d}(\mathcal{P})$
AUROC $= \texttt{calc\_auroc}(\mathcal{Y}, \mathcal{P})$
Vrati AUROC

5. Eksperimenti

U ovom poglavlju prikazujemo provedene eksperimente i njihove rezultate. Programska implementacija ostvarena je radnim okvirom za razvoj dubokih modela *PyTorch*. Za treniranje i evaluaciju korištena je grafička kartica NVIDIA Titan Xp s 12 GiB radne memorije.

U svim eksperimentima pratimo gubitak i točnost predikcije difuzijskog vremena na skupu za učenje. Na skupu UBnormal pratimo AUROC na skupu za validaciju. UBnormal sadrži i anomalne primjere u skupu za učenje koje ne koristimo za učenje, ali ih koristimo za praćenje metrike AP na proširenom skupu za učenje `train_with_ab`. Konačnu mjeru uspješnosti modela računamo metrikom AUROC na skupu za testiranje.

5.1. Rezultati osnovnog i proširenog modela

Sada uspoređujemo rezultate najboljeg osnovnog DTE modela (poglavlje 4.3) s najboljim DTE modelom koji je proširen slojevima konvolucija nad grafovima (poglavlje 4.4). U oba slučaja najbolji model dala je kategorička varijanta procjene difuzijskog vremena.

U tablicama 5.1 i 5.2 prikazani su hiperparametri najboljih modela, a rezultati u 5.3. Očekivano, uvođenje konvolucija nad grafovima umjesto višeslojnog perceptrona pomaže modelu pri procjeni difuzijskog vremena.

U tablicama 5.4 i 5.5 uspoređujemo našu metodu s ostalim iz literature. DTE proširenim slojevima konvolucije nad grafovima postiže novo stanje tehnike na skupu UBnormal.

Hiperparametar	UBnormal	ShanghaiTech
DTE varijanta	kategorička	kategorička
T (broj vremenskih koraka difuzije)	200	200
Raspored šuma	linearni	linearni
Broj klasifikacijskih spremnika	7	7
Duljina segmenta	16	16
Stopa učenja	1e-3	3e-4
Optimizator	Adam	Adam
Veličina mini-grupe	512	512
Broj epoha	200	1
Skriveni slojevi	[512, 128, 512]	[256, 64, 256]
Vjerojatnost isključivanja neurona p	0.5	0.5
λ (L2 regularizacija)	5e-4	5e-4

Tablica 5.1: Hiperparametri najboljih osnovnih DTE modela za skupove UBnormal i ShanghaiTech. Zanimljivo je da je za učenje najboljeg modela za ShanghaiTech potrebna samo 1 epoha, za razliku od UBnormal gdje smo model učili znatno dulje. To pripisujemo jednostavnosti i maloj raznolikosti tipova anomalija u ShanghaiTech, koje većinom odgovaraju brzoj kretnji osobe kroz scenu na nekom prijevoznom sredstvu.

Hiperparametar	UBnormal	ShanghaiTech
DTE varijanta	kategorička	kategorička
T (broj vremenskih koraka difuzije)	200	200
Raspored šuma	linearni	linearni
Broj klasifikacijskih spremnika	7	7
Duljina segmenta	16	16
Stopa učenja	1e-3	2e-4
Optimizator	Adam	SGD
Veličina mini-grupe	512	512
Broj epoha	50	1
Kanali u slojevima konvolucija nad grafovima	[512, 256, 512]	[512, 256, 128, 256, 512]
Veličina latentne reprezentacije	128	128
Veličina skrivene reprezentacije	128	128
λ (L2 regularizacija)	5e-4	5e-4

Tablica 5.2: Hiperparametri najboljih DTE modela proširenih konvolucijama nad grafovima za skupove UBnormal i ShanghaiTech. Kao i kod osnovnog modela (tablica 5.1), model za ShanghaiTech učimo kraće nego model za UBnormal.)

Metoda	UBnormal	ShanghaiTech
Osnovni model	71.0	79.6
DTE+STSGCN	**75.8**	**82.6**

Tablica 5.3: Usporedba metrike AUROC na UBnormal i ShanghaiTech između osnovnog i DTE modela proširenog konvolucijama nad grafovima.

Metoda	AUROC-HR	AUROC-Full
BiPOCO [34] †	52.3	50.7
GEPC [51] †	55.2	53.4
BAF [23]	-	59.3
Jigsaw [73]	-	56.4
MPED-RNN [13] †	61.2	60.6
SSMTL++ [21]	-	62.1
COSKAD [19] †	65.5	65.0
MoCoDAD [18] †	68.4	68.3
STG-NF [29] †	-	71.8
DTE+STSGCN (naša) †	-	**75.8**
BAF [23]	-	61.3
STA [7]	-	68.5
SGCN [11] †	-	64.6
TSAGCN [66] †	-	74.1
UGC [43] †	-	77.8
ST-GCN [77] †	-	78.1
STG-NF [29] †	-	79.2

Tablica 5.4: Rezultati na skupu UBnormal u nenadziranom/polunadziranom (gore) i nadziranom (dolje) načinu rada. Znakom † označene su metode koje rade isključivo sa skeletonima. Naša metoda nadmašuje trenutno stanje tehnike (STG-NF) u polunadziranom načinu rada za 4 postotna boda. Stupac AUROC-HR (engl. *Human Related*) predstavlja AUROC rezultate na podskupu UBnormala koji se sastoji isključivo od anomalija uzrokovanih ljudskim akcijama (bez videozapisa u kojima je anomalija npr. požar u sceni). Taj stupac ovdje dodajemo zbog cjelovitosti.

Metoda	AUROC-HR	AUROC-Full
sRNN [48]	-	68.0
Conv-AE [28]	69.8	70.4
LSA [1]	-	72.5
FFP [42]	72.7	72.8
BiPOCO [34] †	74.9	-
MPED-RNN [13] †	75.4	73.4
MTP [61] †	77.0	76.0
GEPC [51] †	74.8	76.1
PoseCVAE [33] †	75.5	-
Normal Graph [49] †	76.5	-
COSKAD [19] †	77.1	-
STGCAE-LSTM [40] †	77.2	-
MoCoDAD [18] †	77.6	-
GCL [78]	-	79.6
unmasking [32]	-	80.6
BAF [23]	-	83.6
SSMTL++ [8]	-	83.8
Jigsaw [73]	84.7	84.2
STG-NF [29] †	**87.4**	**85.9**
DTE+STSGCN (naša) †	84.1	82.6

Tablica 5.5: Rezultati na skupu ShanghaiTech u polunadziranom načinu rada. Znakom † označene su metode koje rade isključivo sa skeletonima. Objašnjenje stupca AUROC-HR je dano u opisu tablice 5.4.

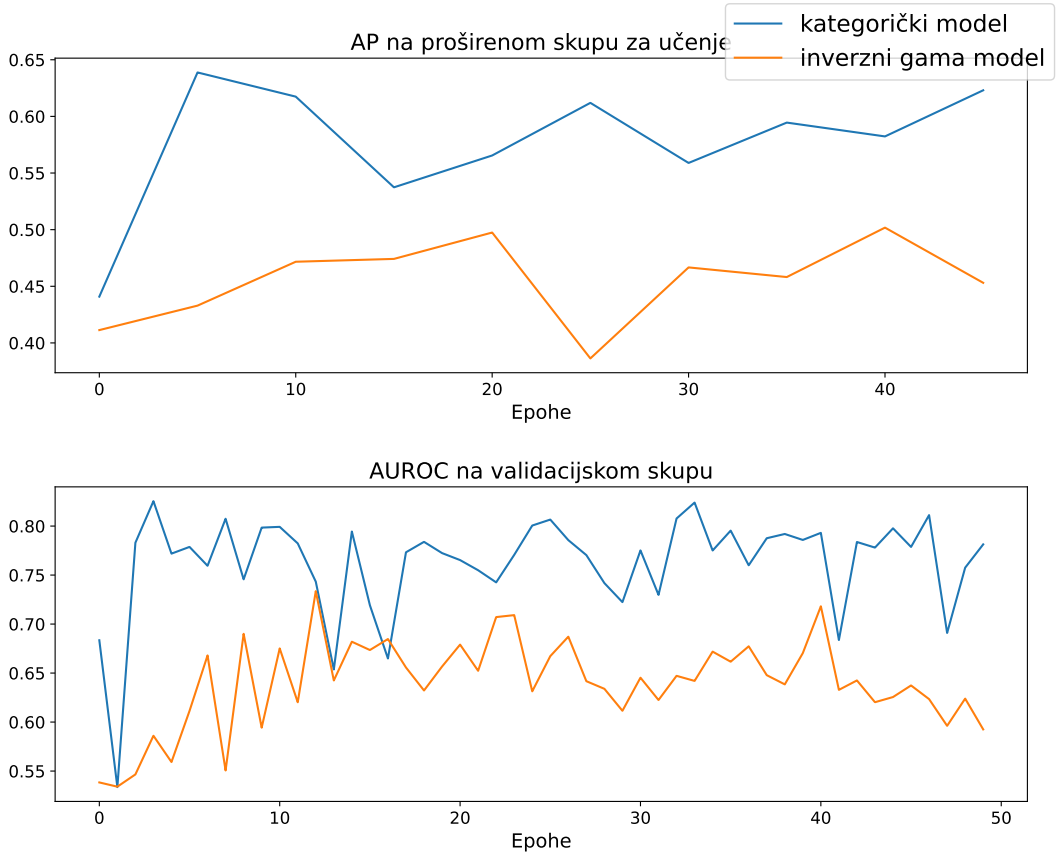
5.2. Validacija hiperparametara

U svrhu pronalaženja najbolje konfiguracije hiperparametara, validiramo razne kombinacije njihovih vrijednosti. Hiperparametri koje razmatramo su: varijanta DTE modela (kategorička i inverzna gama), broj vremenskih koraka difuzije T, raspored varijance unaprijednog procesa, optimizator, stopa učenja, broj kanala (veličine skrivenih slojeva) okosnice, broj klasifikacijskih pretinaca, duljina segmenta skeletona, jačina L2 regularizacije. U nastavku prikazujemo validaciju različitih vrijednosti za samo neke od njih.

5.2.1. Varijante DTE modela

Kategorička varijanta DTE modela pokazala je bolje performanse u [44]. To se pokazalo i u našem slučaju. Na ilustraciji 5.1 uspoređujemo kategorički i inverzni gama model naučene na skupu UBnormal. Pratimo prosječnu preciznost (AP) na skupu za učenje (proširenog anomalnim primjerima) i AUROC na validacijskom skupu.

Mogući razlog ovome je ograničena ekspresivnost inverznog gama modela koji predviđa samo parametar beta predodređene inverzne gama distribucije. Kategorički model pruža veću fleksibilnost pri procjeni difuzijskog vremena [44].



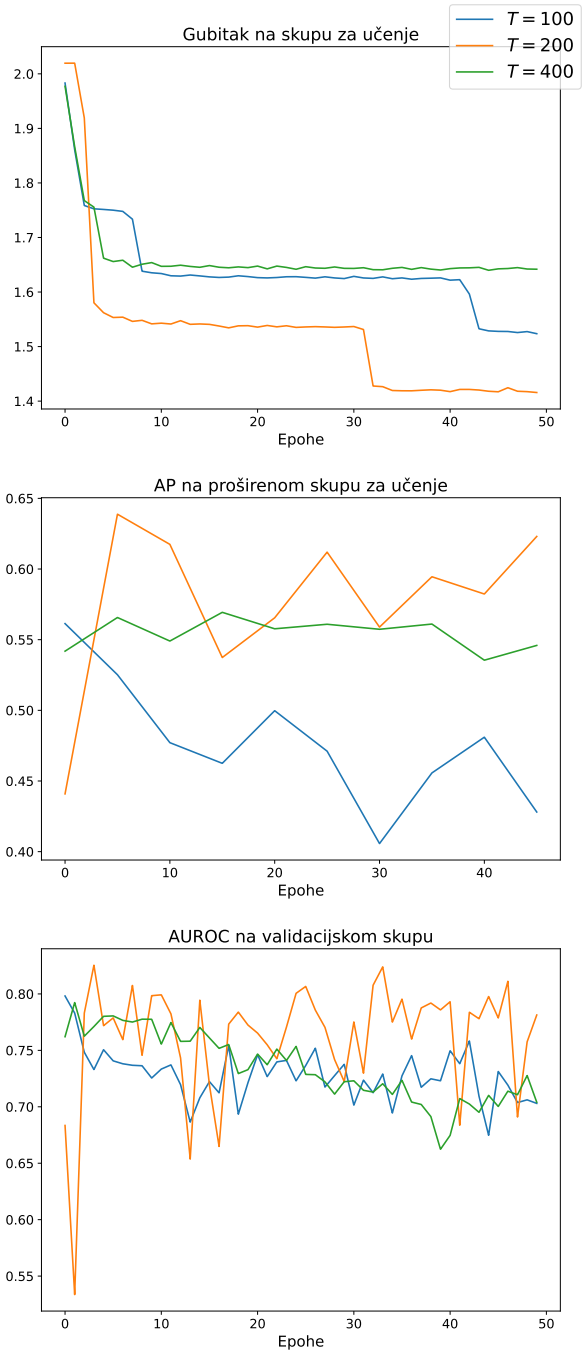
Slika 5.1: Usporedba kategoričke i inverzne gama varijanta modela za procjenu difuzijskog vremena. Kategorički model iskazuje bolje performanse od inverznog gama modela.

5.2.2. Broj vremenskih koraka difuzije T

Hiperparametar T određuje duljinu unaprijednog difuzijskog procesa. Ako odaberemo premalu vrijednost za T, postoji mogućnost da posljednja latentna varijabla $\mathbf{x}_T$ neće izgledati kao da pripada jediničnoj Gaussovoj distribuciji i da nećemo pokriti sve potencijalne anomalije [44]. S druge strane, preveliki izbor za T nam može značajno

produljiti i otežati učenje, jer gubitak računamo kao očekivanje preko svih vremenskih koraka (jednadžbe 2.45 i 2.46).

Eksperimentiramo s tri vrijednosti za T: 100, 200 i 400. Najbolja generalizacija postiže se uz duljinu unaprijednog procesa $T = 200$.



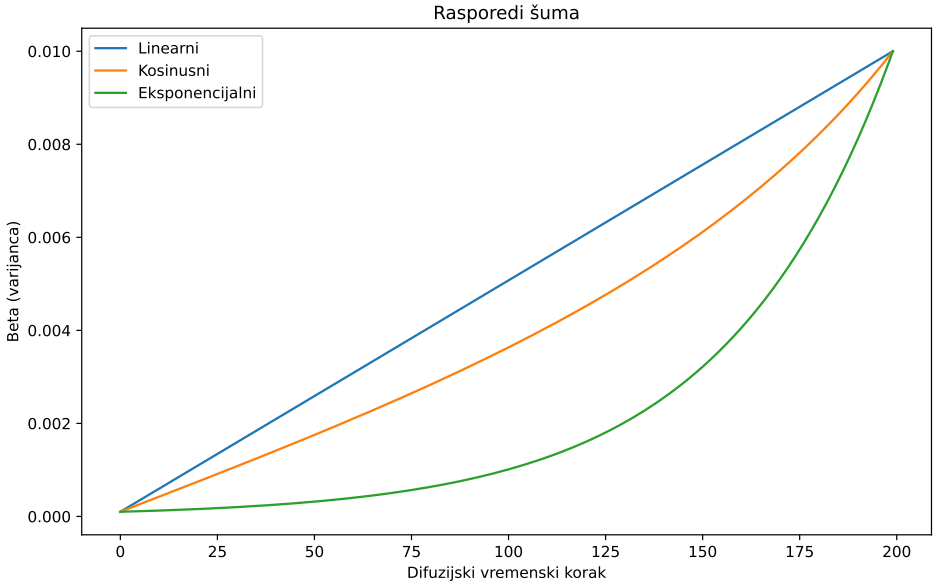
Slika 5.2: Validacija različitih duljina T unaprijednog difuzijskog procesa. Najbolja generalizacijska točnost postiže se uz duljinu $T = 200$.

5.2.3. Raspored šuma

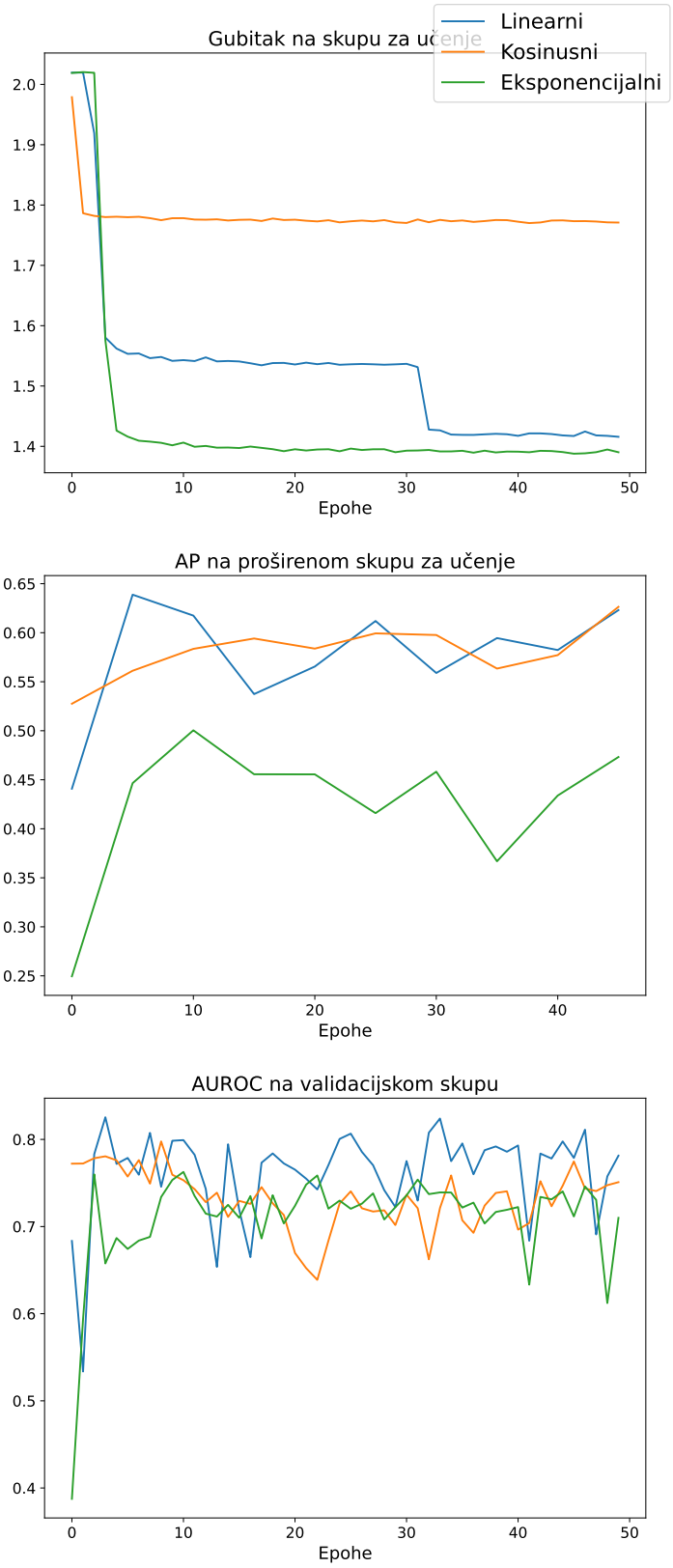
Raspored šuma definira parametre $\beta_t = 1 - \alpha_t$, odnosno varijancu unaprijednog difuzijskog procesa. Uz fiksni raspored β_t, unaprijedni proces je potpuno određen za neki podatak $\mathbf{x}_0$, prema jednadžbi 2.28.

Uloga ovog hiperparametra je bitna jer definira kojom brzinom "uništavamo" informaciju iz ulaznog podatka i izgled latentnih varijabli. U originalom DDPM modelu [30] koristi se jednostavni linearni raspored šuma. U [54] predlažu kosinusni raspored šuma, s ciljem boljeg očuvanja informacija u srednjem dijelu unaprijednog procesa. Motivirani time, eksperimentiramo s tri različita rasporeda šuma: linearni, kosinusni i eksponencijalni. Rasporedi su prikazani na slici 5.3.

Rezultati su vidljivi na slici 5.4. Eksponencijalni raspored šuma daje najbolju empirijsku točnost, ali linearni raspored najbolje generalizira. Stoga, odabiremo ga kao glavnog u ostalim eksperimentima.



Slika 5.3: Tri različita rasporeda varijance (šuma) u unaprijedom difuzijskom procesu. Plavom bojom prikazan je linearni, narančastom kosinusni, a zelenom eksponencijalni raspored šuma.



Slika 5.4: Usporedba performansi različitih rasporeda varijance (šuma). Plavom bojom prikazan je linearni, narančastom kosinusni, a zelenom eksponencijalni raspored šuma. Eksponencijalni raspored ostvaruje najmanji gubitak na skupu za učenje, ali linearni raspored najbolje generalizira.

5.3. Usporedba s kNN-om

Premda vrlo jednostavna, metoda k najbližih susjeda impresivno detektira anomalije u raznim tipovima podataka. U originalnom DTE radu [44], ne uspijevaju nadmašiti kNN na tabličnim podacima, što dokazuje njenu robusnost.

Mi također evaluiramo metodu kNN, na skupovima UBnormal i ShanghaiTech. Dobivene rezultate prikazujemo u tablici 5.6. Vidimo da kNN postiže bolji AUROC na skupu UBnormal, a naša metoda na skupu ShanghaiTech.

Glavni nedostatak metode k najbližih susjeda je njena vremenska i prostorna složenost. kNN mora pohraniti sve podatke iz skupa za učenje. U fazi zaključivanja, mora pronaći k najbližih susjeda iz skupa za učenje za svaki podatak u evaluacijskom skupu. Naša metoda je prostorno jeftinija jer ne mora pamtiti cijeli skup podataka, nego samo parametre modela. Osim toga, vremenski je učinkovitija jer joj je za evaluaciju jednog podatka potreban samo jedan unaprijedni prolaz kroz mrežu. Da to dokažemo, mjerimo trajanje evaluacije oba modela na nasumično generiranom skupu podataka u tablici 5.7. Očekujemo da bi razlike s porastom broja primjera u skupu za učenje bile još veće.

Metoda	UBnormal	ShanghaiTech
kNN ($k = 5$)	**77.7**	79.7
DTE+STSGCN	75.8	**82.6**

Tablica 5.6: Usporedba metrike AUROC na UBnormal i ShanghaiTech između metode k najbližih susjeda i našeg najboljeg modela. Korištena je veličina susjedstva $k = 5$.

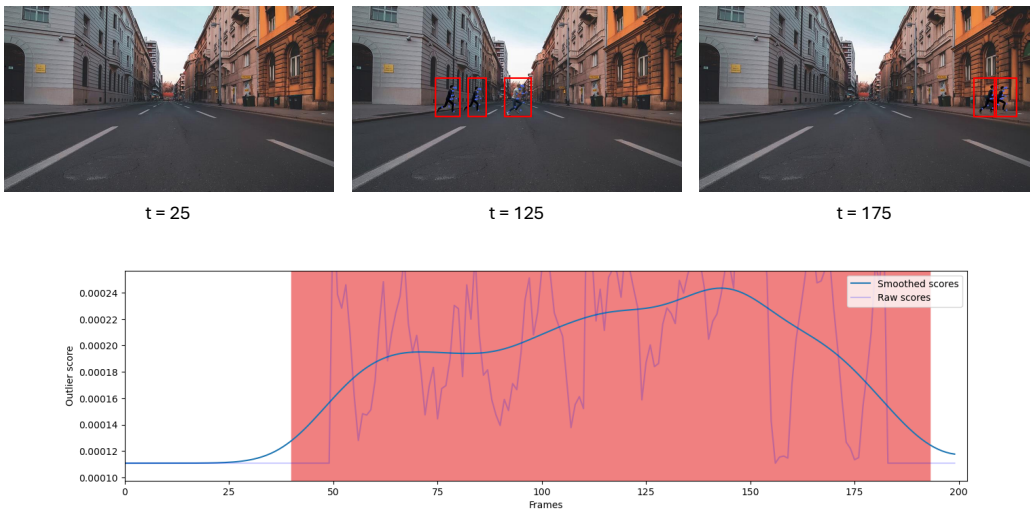
Metoda	Trajanje
kNN ($k = 5$)	31.77 sekundi
DTE+STSGCN	**28.12** sekundi

Tablica 5.7: Usporedba efikasnosti metode k najbližih susjeda i našeg najboljeg modela. Mjerimo trajanje evaluacije modela mjereno na nasumično generiranom skupu segmenata skeletona sa 100000 primjera u skupu za učenje i 300000 primjera u skupu za testiranje, što otprilike odgovara stanju stvari u stvarnim skupovima koje koristimo u radu.

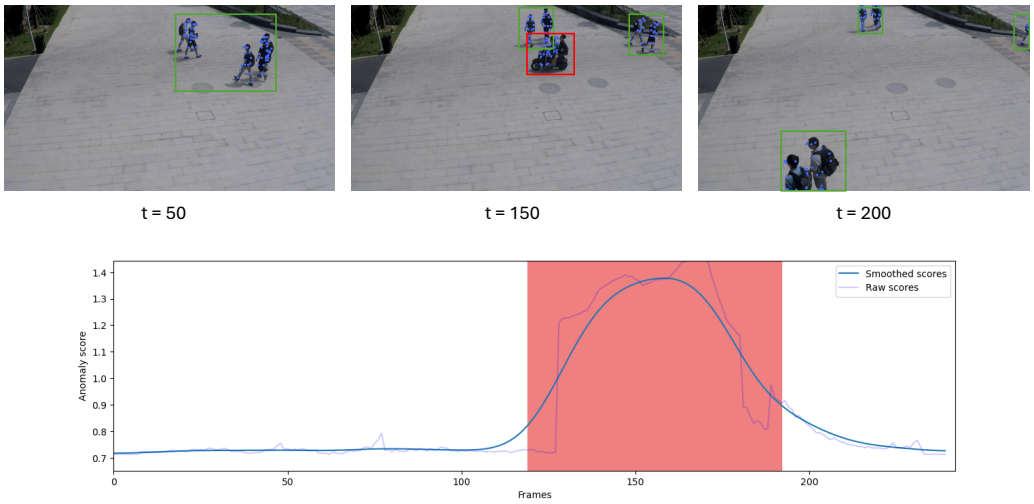
5.4. Kvalitativna analiza

5.4.1. Dobri primjeri

Na slici 5.5 prikazan je rad našeg DTE+STSGCN modela na podacima iz UBnormal i ShanghaiTech. Vidimo da se model dobro snalazi u detekciji različitih tipova anomalija u scenama s više osoba. Zaglađivanje mjere anomalnosti Gaussovim filtrom povećava robusnost modela umanjujući utjecaj individualnih segmenata skeletona.



(a) UBnormal



(b) ShanghaiTech

Slika 5.5: Primjeri dobrih detekcija anomalnih radnji našeg modela na skupovima UBnormal i ShanghaiTech. U obje ilustracije prikazujemo okvire iz videozapisa (gore) i mjeru anomalnosti (dolje). Crvenom pozadinom označen je dio videozapisa koji sadrži anomalnu radnju. U okvirima, crvenim pravokutnikom označavamo aktere koji sudjeluju u anomalnoj, a zelenim aktere koji sudjeluju u normalnoj radnji.

5.4.2. Teški primjeri

Na slici 5.6 prikazan je primjer lažno pozitivne detekcije na okviru iz UBnormal. Primijetili smo da se lažno pozitivna detekcija često odvija na okvirima u kojima osoba

upravo ulazi ili izlazi iz scene. To su trenuci u kojima model za detekciju osoba uspješno pronađe osobu, ali ekstraktor ključnih točaka ne uspije kvalitetno smjestiti sve ključne točke na pravo mjesto. Za DTE model takav segment skeletona izgleda vrlo šumovito zbog čega mu pridjeljuje veliku procjenu difuzijskog vremena, odnosno mjeru anomalnosti. Sličan problem dogodi se kada dvije osobe prolaze jako blizu jedna kraj druge u sceni. Ovaj problem bi se potencijalno mogao riješiti filtriranjem nepouzdanih predikcija ekstraktora ključnih točaka, što ostavljamo za budući rad.

Slika 5.6: Primjer lažno pozitivne detekcije u trenutku u kojem osoba izlazi iz scene.

6. Zaključak

U ovom radu bavimo se pronalaženjem anomalnih ljudskih radnji u videozapisima. Iz svakog okvira videozapisa izvlačimo koordinate ključnih točaka osoba i slažemo ih u segmente skeletona. Razvijamo polunadziranu metodu koja će na temelju isključivo poze osobe, odnosno prostorno-vremenskih relacija njenih ključnih točaka, prepoznati radi li se o neobičnom ponašanju koje odskače od naučenog koncepta normalnosti.

U poglavlju o teorijskoj podlozi predstavljamo varijacijske difuzijske modele, posebnu vrstu hijerarhijskih varijacijskih autoenkodera, koji su se pokazali kao moćan generativni model za pronalaženje anomalija, ali su računalno skupi. Procjena difuzijskog vremena (DTE) je moderna tehnika inspirirana difuzijskim modelima koja ih nadmašuje u zadatku detekcije anomalija uz značajno manji broj računalnih operacija. Vođeni time, učimo kategorički i inverzni gama DTE model na segmentima skeletona. Koristeći jednostavnu višeslojnu potpuno povezanu mrežu za procjenu difuzijskog vremena, postižemo rezultate koji se približavaju trenutnom stanju tehnike. Proširivanjem arhitekture konvolucijama nad grafovima značajno nadmašujemo stanje tehnike na skupu podataka UBnormal što ističemo kao glavni doprinos ovog rada.

Predložena metoda nije bez nedostataka. Klasična metoda kNN, premda manje računalno efikasna, nadmašuje DTE za 2 postotna boda na skupu za testiranje od UBnormal. U budućem istraživanju korisno bi bilo ispitati razlike između naše metode i kNN-a analizirajući primjere na kojima naš model radi lošije. Osim toga, naša metoda je dosta ovisna o ekstraktoru ključnih točaka na osobama. Krive predikcije ekstraktora povećavaju lažno pozitivne predikcije našeg modela. Stoga, buduća istraživanja bi mogla ići u smjeru eksperimentiranja s novijim ekstraktorima i uklanjanja lažnih skeletona iz podataka. Naša metoda temelji se na udaljenostima anomalnih i normalnih primjera, što može biti problematično u visoko dimenzionalnim prostorima. Zanimljiva ideja je nekom samonadziranom metodom naučiti ugrađivanja skeletona u niždedimenzionalni latentni prostor te u njemu provoditi difuziju. Dobra polazišna točka je naučiti varijacijski autoenkoder sa slojevima konvolucije nad grafovima, zatim latentne vektore koristiti kao ulaz u DTE.

LITERATURA

[1] Davide Abati, Angelo Porrello, Simone Calderara, i Rita Cucchiara. Latent space autoregression for novelty detection. U *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, stranice 481–490. Computer Vision Foundation / IEEE, 2019.

[2] Andra Acsintoae, Andrei Florescu, Mariana-Iuliana Georgescu, Tudor Mare, Paul Sumedrea, Radu Tudor Ionescu, Fahad Shahbaz Khan, i Mubarak Shah. Ubnormal: New benchmark for supervised open-set video anomaly detection. U *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, stranice 20143–20153, 2022.

[3] Amit Adam, Ehud Rivlin, Ilan Shimshoni, i Daviv Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE transactions on pattern analysis and machine intelligence*, 30(3):555–560, 2008.

[4] Mohiuddin Ahmed, Abdun Naser Mahmood, i Md Rafiqul Islam. A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, 55:278–288, 2016.

[5] Stephen D Bay i Mark Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. U *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, stranice 29–38, 2003.

[6] Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[7] Gedas Bertasius, Heng Wang, i Lorenzo Torresani. Is space-time attention all you need for video understanding? U Marina Meila i Tong Zhang, urednici, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, svezak 139 od *Proceedings of Machine Learning Research*, stranice 813–824. PMLR, 2021.

[8] Antonio Bărbălău, Radu Tudor Ionescu, Mariana-Iuliana Georgescu, Jacob Velling Dueholm, Bharathkumar Ramachandra, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas Baltzer Moeslund, i Mubarak Shah. Ssmtl++: Revisiting self-supervised multi-task learning for video anomaly detection. *Comput. Vis. Image Underst.*, 229:103656, 2022.

[9] Varun Chandola, Arindam Banerjee, i Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009.

[10] Xiaoran Chen i Ender Konukoglu. Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. U *Medical Imaging with Deep Learning*, 2018.

[11] Ke Cheng, Yifan Zhang, Xiangyu He, Weihan Chen, Jian Cheng, i Hanqing Lu. Skeleton-based action recognition with shift graph convolutional network. U *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, stranice 180–189. Computer Vision Foundation / IEEE, 2020.

[12] Thomas M. Cover i Peter E. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory*, 13(1):21–27, 1967. URL `http://dblp.uni-trier.de/db/journals/tit/tit13.html#CoverH67`.

[13] Romero F. A. B. de Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Reda Mansour, i Svetha Venkatesh. Learning regularity in skeleton trajectories for anomaly detection in videos. U *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, stranice 11996–12004. Computer Vision Foundation / IEEE, 2019.

[14] Prafulla Dhariwal i Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[15] Sahibsingh A. Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4):325–327, 1976. doi: 10.1109/TSMC.1976.5408784.

[16] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, i Cewu Lu. Alphapose: Whole-body regional multi-person pose

estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7157–7173, 2022.

[17] William Feller. On the theory of stochastic processes, with particular reference to applications. 1949.

[18] Alessandro Flaborea, Luca Collorone, Guido Maria D'Amely Di Melendugno, Stefano D'Arrigo, Bardh Prenkaj, i Fabio Galasso. Multimodal motion conditioned diffusion model for skeleton-based video anomaly detection. U *Proceedings of the IEEE/CVF International Conference on Computer Vision*, stranice 10318–10329, 2023.

[19] Alessandro Flaborea, Guido D'Amely, Stefano D'Arrigo, Marco Aurelio Sterpa, Alessio Sampieri, i Fabio Galasso. Contracting skeletal kinematics for human-related video anomaly detection. *Arxiv*, 2023.

[20] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, i Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[21] Mariana-Iuliana Georgescu, Antonio Bărbălău, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Claudiu Popescu, i Mubarak Shah. Anomaly detection in video via self-supervised and multi-task learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, stranice 12737–12747, 2020. URL `https://api.semanticscholar.org/CorpusID:226964553`.

[22] Mariana Iuliana Georgescu, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, i Mubarak Shah. A background-agnostic framework with adversarial training for abnormal event detection in video. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4505–4523, 2021.

[23] Mariana-Iuliana Georgescu, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, i Mubarak Shah. A background-agnostic framework with adversarial training for abnormal event detection in video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(9):4505–4523, 2022.

[24] Koosha Golmohammadi i Osmar R Zaiane. Time series contextual anomaly detection for detecting market manipulation in stock market. U *IEEE International Conference on Data Science and Advanced Analytics*, stranice 1–10, 2015.

[25] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, i Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[26] Piotr S Gromski, Alon B Henson, Jarosław M Granda, i Leroy Cronin. How to explore chemical space using algorithms and automation. *Nature Reviews Chemistry*, 3(2):119–128, 2019.

[27] Songqiao Han, Xiyang Hu, Hailiang Huang, Mingqi Jiang, i Yue Zhao. Adbench: Anomaly detection benchmark. U *Neural Information Processing Systems (NeurIPS)*, 2022.

[28] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K. Roy-Chowdhury, i Larry S. Davis. Learning temporal regularity in video sequences. U *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, stranice 733–742. IEEE Computer Society, 2016.

[29] Or Hirschorn i Shai Avidan. Normalizing flows for human pose anomaly detection. U *Proceedings of the IEEE/CVF International Conference on Computer Vision*, stranice 13545–13554, 2023.

[30] Jonathan Ho, Ajay Jain, i Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[31] Heiko Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3): 863–874, 2007.

[32] Radu Tudor Ionescu, Sorina Smeureanu, Bogdan Alexe, i Marius Popescu. Unmasking the abnormal events in video. U *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, stranice 2914–2922. IEEE Computer Society, 2017.

[33] Yashswi Jain, Ashvini Kumar Sharma, Rajbabu Velmurugan, i Biplab Banerjee. Posecvae: Anomalous human activity detection. U *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*, stranice 2927–2934. IEEE, 2020.

[34] Asiegbu Miracle Kanu-Asiegbu, Ram Vasudevan, i Xiaoxiao Du. Bipoco: Bidirectional trajectory prediction with pose constraints for pedestrian anomaly detection. *CoRR*, abs/2207.02281, 2022.

[35] Diederik P. Kingma i Max Welling. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019. ISSN 1935-8245. doi: 10.1561/2200000056. URL http://dx.doi.org/10.1561/2200000056.

[36] Diederik P Kingma i Max Welling. Auto-encoding variational bayes, 2022.

[37] Edwin M Knorr, Raymond T Ng, i Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3):237–253, 2000.

[38] Flip Korn i Suresh Muthukrishnan. Influence sets based on reverse nearest neighbor queries. *ACM Sigmod Record*, 29(2):201–212, 2000.

[39] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[40] Nanjun Li, Faliang Chang, i Chunsheng Liu. Human-related anomalous event detection via spatial-temporal graph convolutional autoencoder with embedded long short-term memory network. *Neurocomputing*, 490:482–494, 2021. URL https://api.semanticscholar.org/CorpusID:245456764.

[41] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, i C Lawrence Zitnick. Microsoft coco: Common objects in context. U *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, stranice 740–755. Springer, 2014.

[42] Wen Liu, Weixin Luo, Dongze Lian, i Shenghua Gao. Future frame prediction for anomaly detection - A new baseline. U *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, stranice 6536–6545. Computer Vision Foundation / IEEE Computer Society, 2018.

[43] Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, i Wanli Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. U *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, stranice 140–149. Computer Vision Foundation / IEEE, 2020.

[44] Victor Livernoche, Vineet Jain, Yashar Hezaveh, i Siamak Ravanbakhsh. On diffusion modeling for anomaly detection. *arXiv preprint arXiv:2305.18593*, 2023.

[45] Felipe Lopez, Miguel Saez, Yuru Shao, Efe C Balta, James Moyne, Z Morley Mao, Kira Barton, i Dawn Tilbury. Categorization of anomalies in smart manufacturing systems to support the selection of detection mechanisms. *Robotics and Automation Letters*, 2(4):1885–1892, 2017.

[46] Cewu Lu, Jianping Shi, i Jiaya Jia. Abnormal event detection at 150 fps in matlab. U *Proceedings of the IEEE international conference on computer vision*, stranice 2720–2727, 2013.

[47] Calvin Luo. Understanding diffusion models: A unified perspective, 2022.

[48] Weixin Luo, Wen Liu, i Shenghua Gao. A revisit of sparse coding based anomaly detection in stacked rnn framework. U *Proceedings of the IEEE international conference on computer vision*, stranice 341–349, 2017.

[49] Weixin Luo, Wen Liu, i Shenghua Gao. Normal graph: Spatial temporal graph convolutional networks based prediction network for skeleton based video anomaly detection. *Neurocomputing*, 444:332–337, 2020. URL `https://api.semanticscholar.org/CorpusID:229470119`.

[50] Vijay Mahadevan, Weixin Li, Viral Bhalodia, i Nuno Vasconcelos. Anomaly detection in crowded scenes. U *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, stranice 1975–1981, 2010. doi: 10.1109/CVPR.2010.5539872.

[51] Amir Markovitz, Gilad Sharir, Itamar Friedman, Lihi Zelnik-Manor, i Shai Avidan. Graph embedded pose clustering for anomaly detection. U *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, stranice 10536–10544. Computer Vision Foundation / IEEE, 2020.

[52] Shakir Mohamed i Balaji Lakshminarayanan. Learning in implicit generative models, 2017.

[53] Benjamin Nachman i David Shih. Anomaly detection with density estimation. *Physical Review D*, 101:075042, Apr 2020.

[54] Alexander Quinn Nichol i Prafulla Dhariwal. Improved denoising diffusion probabilistic models. U *International conference on machine learning*, stranice 8162–8171. PMLR, 2021.

[55] Tudor I Oprea. Chemical space navigation in lead discovery. *Current Opinion in Chemical Biology*, 6(3):384–389, 2002.

[56] Poojan Oza i Vishal M Patel. One-class convolutional neural network. *IEEE Signal Processing Letters*, 26(2):277–281, 2019.

[57] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[58] Miloš Radovanović, Alexandros Nanopoulos, i Mirjana Ivanović. Reverse nearest neighbors in unsupervised distance-based outlier detection. *IEEE transactions on knowledge and data engineering*, 27(5):1369–1382, 2014.

[59] Bharathkumar Ramachandra i Michael Jones. Street scene: A new dataset and evaluation protocol for video anomaly detection. U *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, stranice 2569–2578, 2020.

[60] Sridhar Ramaswamy, Rajeev Rastogi, i Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. U *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, stranice 427–438, 2000.

[61] Royston Rodrigues, Neha Bhargava, Rajbabu Velmurugan, i Subhasis Chaudhuri. Multi-timescale trajectory prediction for abnormal human activity detection. U *IEEE Winter Conference on Applications of Computer Vision, WACV 2020, Snowmass Village, CO, USA, March 1-5, 2020*, stranice 2615–2623. IEEE, 2020.

[62] Olaf Ronneberger, Philipp Fischer, i Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. U *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, stranice 234–241. Springer, 2015.

[63] Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G Dietterich, i Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5):756–795, 2021.

[64] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, i Ursula Schmidt-Erfurth. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019.

[65] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, i Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[66] Lei Shi, Yifan Zhang, Jian Cheng, i Hanqing Lu. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, stranice 12018–12027, 2018.

[67] Theodoros Sofianos, Alessio Sampieri, Luca Franco, i Fabio Galasso. Space-time-separable graph convolutional network for pose forecasting. U *Proceedings of the IEEE/CVF International Conference on Computer Vision*, stranice 11209–11218, 2021.

[68] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, i Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. U *International conference on machine learning*, stranice 2256–2265. PMLR, 2015.

[69] Waqas Sultani, Chen Chen, i Mubarak Shah. Real-world anomaly detection in surveillance videos. U *Proceedings of the IEEE conference on computer vision and pattern recognition*, stranice 6479–6488, 2018.

[70] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, i Jinwoo Shin. CSI: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in Neural Information Processing Systems*, 2020.

[71] Lionel Tarassenko, Paul Hayton, Nicholas Cerneaz, i Michael Brady. Novelty detection for the identification of masses in mammograms. U *International Conference on Artificial Neural Networks*, stranice 442–447, 1995.

[72] Laurens van der Maaten i Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL `http://jmlr.org/papers/v9/vandermaaten08a.html`.

[73] Guodong Wang, Yunhong Wang, Jie Qin, Dongming Zhang, Xiuguo Bao, i Di Huang. Video anomaly detection by solving decoupled spatio-temporal jigsaw

puzzles. U Shai Avidan, Gabriel J. Brostow, Moustapha Cissé, Giovanni Maria Farinella, i Tal Hassner, urednici, *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part X*, svezak 13670 od *Lecture Notes in Computer Science*, stranice 494–511. Springer, 2022.

[74] Julia Wolleb, Florentin Bieder, Robin Sandkühler, i Philippe C Cattin. Diffusion models for medical anomaly detection. U *International Conference on Medical image computing and computer-assisted intervention*, stranice 35–45. Springer, 2022.

[75] Julian Wyatt, Adam Leach, Sebastian M Schmon, i Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. U *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, stranice 650–656, 2022.

[76] Zhisheng Xiao, Karsten Kreis, i Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804*, 2021.

[77] Sijie Yan, Yuanjun Xiong, i Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. U *Proceedings of the AAAI conference on artificial intelligence*, svezak 32, 2018.

[78] Muhammad Zaigham Zaheer, Arif Mahmood, Muhammad Haris Khan, Mattia Segù, Fisher Yu, i Seung-Ik Lee. Generative cooperative learning for unsupervised video anomaly detection. U *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, stranice 14724–14734. IEEE, 2022.

[79] Hui Zhang, Zheng Wang, Zuxuan Wu, i Yu-Gang Jiang. Diffusionad: Denoising diffusion for anomaly detection. *arXiv preprint arXiv:2303.08730*, 2023.

[80] Lisheng Zhang, Zehua He, i Dajiang Lei. Shared nearest neighbors based outlier detection for biological sequences. *International Journal of Digital Content Technology and its Applications*, 6(12):1–10, 2012.

[81] Jun-Yan Zhu, Taesung Park, Phillip Isola, i Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. U *2017 IEEE International Conference on Computer Vision (ICCV)*, stranice 2242–2251, 2017. doi: 10.1109/ICCV.2017.244.

[82] Jan Šnajder. *Vrednovanje modela, predavanja iz kolegija Strojno učenje*. 2022.

Pronalaženje anomalija difuzijskim uklanjanjem šuma

Sažetak

Varijacijski difuzijski modeli iskazuju impresivne performanse na zadatku detekcije anomalija. Njihova glavna mana je računalna složenost, zbog prolaska kroz dugi unatražni Markovljev proces. Procjena difuzijskog vremena (DTE) varijanta je difuzijskih modela koja ih nadmašuje u zadatku detekcije anomalija uz značajno manji broj potrebnih računalnih operacija. U ovom radu učimo model difuzijske procjene vremena na segmentima skeletona koji opisuju poze osoba u videozapisima, s ciljem pronalaženja anomalnih radnji. U osnovnom eksperimentu pokazujemo da difuzijska procjena vremena primijenjena izravno na skeletone daje kompetitivne rezultate. Zatim, segment skeletona promatramo kao prostorno-vremenski graf te u model za procjenu difuzijskog vremena ubacujemo slojeve konvolucije nad grafovima. Evaluiramo predloženu arhitekturu na skupovima UBnormal i ShanghaiTech u polunadziranom načinu rada. Naši rezultati postižu kompetitivne rezultate na skupu ShanghaiTech te nadmašuju stanje tehnike na skupu UBnormal za 4 postotna boda. Provodimo diskusiju rezultata i predlažemo pravce za budući rad.

Ključne riječi: detekcija anomalija, ljudska poza, varijacijski difuzijski modeli, procjena difuzijskog vremena, konvolucije nad grafovima

Anomaly detection through denoising diffusion

Abstract

Variational diffusion models show impressive performance on the anomaly detection task. Their main drawback is computational complexity, due to the long traversal of the backward Markov process. Diffusion Time Estimation (DTE) is a variant of diffusion models that outperforms them in the task of anomaly detection with a significantly lower number of parameters and computational power needed. In this paper, we train a diffusion time estimation model on skeleton segments describing the poses of people in videos, with the goal of detecting anomalous actions. For our baseline experiment, we show that diffusion time estimation applied directly to skeletons gives competitive results. Next, we interpret skeleton segments as spatio-temporal graphs and introduce graph convolutional layers into the diffusion time estimation model. We evaluate the proposed architecture on the UBnormal and ShanghaiTech sets in the semi-supervised setting. Our results achieve competitive results on ShanghaiTech and significantly surpass the state of the art on UBnormal. We perform a discussion of the results and suggest directions for future work.

Keywords: anomaly detection, human pose, variational diffusion models, diffusion time estimation, graph convolutional networks