Detekcija anomalnih sljedova poza osoba difuzijskim otklanjanjem šuma

Kukolj, Marko

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:168:293902

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-03-23**

Repository / Repozitorij:

FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repozitory

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1653

DETEKCIJA ANOMALNIH SLJEDOVA POZA OSOBA DIFUZIJSKIM OTKLANJANJEM ŠUMA

Marko Kukolj

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1653

DETEKCIJA ANOMALNIH SLJEDOVA POZA OSOBA DIFUZIJSKIM OTKLANJANJEM ŠUMA

Marko Kukolj

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

ZAVRŠNI ZADATAK br. 1653

Pristupnik:	**Marko Kukolj (0036540794)**
Studij:	Elektrotehnika i informacijska tehnologija i Računarstvo
Modul:	Računarstvo
Mentor:	prof. dr. sc. Siniša Šegvić

Zadatak: **Detekcija anomalnih sljedova poza osoba difuzijskim otklanjanjem šuma**

Opis zadatka:

Raspoznavanje akcija osoba u videu zanimljiv je problem računalnog vida s mnogim važnim primjenama. Zbog velikog broja stupnjeva slobode ljudskog tijela, ovaj problem često nije praktično formulirati kao standardno nadzirano učenje. Umjesto toga, ponekad je praktičnije oblikovati sustav koji detektira anomalne obrasce ponašanja s obzirom na skup slijedova za učenje. Ovaj rad istražit će postupke detekcije anomalija utemeljene na difuzijskom otklanjanju šuma. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje matricama i slikama. Proučiti i ukratko opisati postojeće duboke arhitekture utemeljene na konvolucijama i pažnji. Pažljivo proučiti jednadžbe generativnog modeliranja difuzijskim modelima za otklanjanje šuma. Vrednovati generalizacijsku moć postupaka iz literature na javno dostupnim podatcima. Procijeniti složenost učenja modela. Prikazati i ocijeniti provedene eksperimente. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 14. lipnja 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1653

DETEKCIJA ANOMALNIH SLJEDOVA POZA OSOBA DIFUZIJSKIM OTKLANJANJEM ŠUMA

Marko Kukolj

Zagreb, lipanj, 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

ZAVRŠNI ZADATAK br. 1653

Pristupnik:	**Marko Kukolj (0036540794)**
Studij:	Elektrotehnika i informacijska tehnologija i Računarstvo
Modul:	Računarstvo
Mentor:	prof. dr. sc. Siniša Šegvić

Zadatak: **Detekcija anomalnih sljedova poza osoba difuzijskim otklanjanjem šuma**

Opis zadatka:

Raspoznavanje akcija osoba u videu zanimljiv je problem računalnog vida s mnogim važnim primjenama. Zbog velikog broja stupnjeva slobode ljudskog tijela, ovaj problem često nije praktično formulirati kao standardno nadzirano učenje. Umjesto toga, ponekad je praktičnije oblikovati sustav koji detektira anomalne obrasce ponašanja s obzirom na skup slijedova za učenje. Ovaj rad istražit će postupke detekcije anomalija utemeljene na difuzijskom otklanjanju šuma. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje matricama i slikama. Proučiti i ukratko opisati postojeće duboke arhitekture utemeljene na konvolucijama i pažnji. Pažljivo proučiti jednadžbe generativnog modeliranja difuzijskim modelima za otklanjanje šuma. Vrednovati generalizacijsku moć postupaka iz literature na javno dostupnim podatcima. Procijeniti složenost učenja modela. Prikazati i ocijeniti provedene eksperimente. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 14. lipnja 2024.

Sadržaj

1. Uvod **3**

2. Duboko učenje **4**

2.1. Umjetne neuronske mreže 4

2.2. Prijenosne funkcije 5

2.2.1. Funkcija identiteta 6

2.2.2. Sigmoidalna funkcija 6

2.2.3. Zglobnica 7

2.3. Difuzijski modeli 8

2.4. Učenje neuronske mreže 10

2.4.1. Funkcija gubitka 10

2.4.2. Propagacija pogreške unatrag 11

2.4.3. Optimizacija 11

3. Korišteni alati **12**

3.1. Numpy 12

3.2. Pytorch 12

3.3. Pytorch Lightning 13

3.4. WANDB 13

4. Model **14**

5. Skupovi podataka **15**

5.1. HR-ShanghaiTech 15

5.2. HR-Avenue 16

6. Eksperimenti **17**

6.1. HR-ShanghaiTech 17
6.1.1. Hiperparametri 17
6.1.2. Rezultati 18
6.2. HR-Avenue 20
6.2.1. Hiperparametri 20
6.2.2. Rezultati 20

7. Rasprava 23

8. Zaključak 25

Literatura 26

Sažetak 29

Abstract 30

1. Uvod

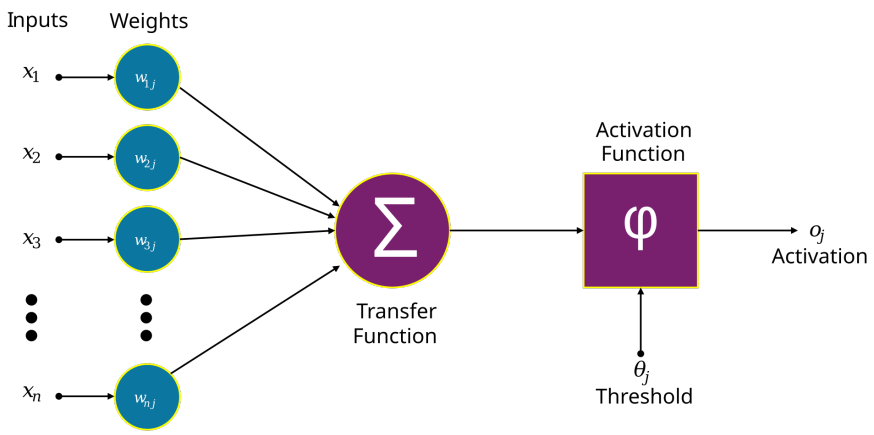
Jedan od bitnih problema u području računalnog vida je raspoznavanje kretnja i ponašanja osoba. Zbog velikog broja stupnjeva slobode ljudskog tijela, ovaj problem nije praktično modelirati pomoću nadziranog učenja.
Jedan alternativan pristup ovom problemu je modelirati sustav koji detektira anomalne kretnje s obzirom na skupove sljedova za učenje. Uočavanje anomalnih kretnji ima veliku primjenu u uočavanju nesreća ili opasnih ponašanja na snimkama.
Većina radova u ovom području se fokusira na OCC (One-Class Classsification) tehnike [1, 2, 3] koje nailaze na problem da normalnih kretnji, isto kao i anomalnih, ima beskonačno mnogo, dok u jednom skupu za učenje mogu obuhvatiti samo dio normalnih kretnji.
Iz tog razloga, ovaj se rad usredotočuje na difuzijsko uklanjanje šuma po uzoru na [4]. Ovim postupkom zapravo modeliramo model koji predviđa normalne obrasce kretnji ljudi te kao anomalne obrasce prepoznaje one koji se ne uklapaju s njegovim predviđanjima. Radovi koji koriste takve tehnike su [4, 5, 6]. Kako bi rješili problem beskonačno mnogo različitih normalnih kretnji generiramo više predviđanja te tako osiguravamo da će postojati predviđanje slično stvarnoj kretnji ako ona nije anomalna.

2. Duboko učenje

Duboko učenje je podskup strojnog učenja koji proučava modele izražene kompozicijom naučenih nelinearnih transformacija. Duboke neuronske mreže postižu jako dobre rezultate u području umjetne inteligencije te je njihova primjena vrlo raširena.

2.1. Umjetne neuronske mreže

Umjetne neuronske mreže su algoritam strojnog učenja korišten u strojnom učenju koji je inspiriran biološkim neuralnim mrežama.

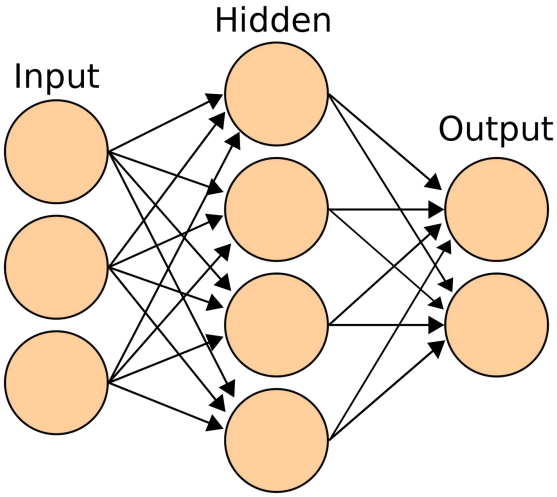


Slika 2.1. Struktura umjetnog neurona

Umjetni neuroni su funkcije koje ulaze množe odgovarajućim težinama, sumiraju ih, dodaju im pomak te primjenjuju prijenosnu fukciju na rezultat. Matematički to možemo zapisati kao (2.1), pri čemu je $\vec{y_i}$ izlaz i-tog sloja neuronske mreže, W_i matrica težina i-tog sloja, $\vec{x_{i-1}}$ ulaz u i-ti sloj, $\vec{b_i}$ pomak i-tog sloja i f prijenosna funkcija neuronske mreže.

$$\vec{y_i} = f(W_i \vec{x_{i-1}} + \vec{b_i}) \tag{2.1}$$

Umjetne neuronske mreže (Slika 2.2.) sastoje se od više slojeva, pri čemu je sloj agregacija neurona. Prvi sloj umjetne neuronske mreže je ulazini sloj, dok je zadnji izlazni sloj, a svi slojevi između nazivaju se skrivenim slojevima. Najjednostavniji tip umjetnih neuronskih mreža je potpuno povezana neuronska mreža u kojoj su svi neuroni nekog sloja potpuno povezani sa svim neuronima prethodnog sloja.



Slika 2.2. Primjer arhitekture neuronske mreže koja se satoji od 3 ulazna neurona, 4 skrivena neurona te 2 izlazna neurona

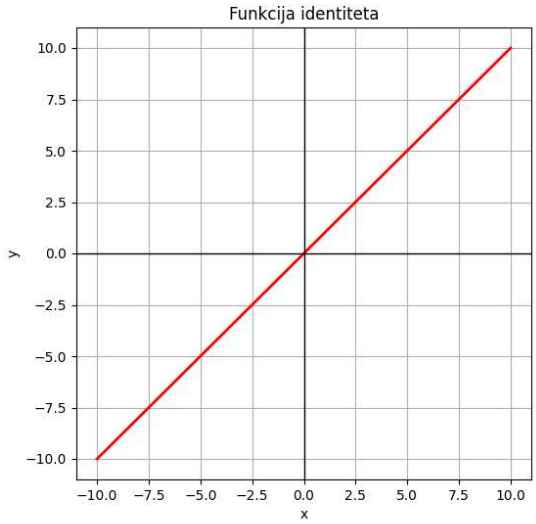
2.2. Prijenosne funkcije

Ako promotrimo svaki sloj neuronske mreže bez prijenosnih fukcija uočiti ćemo da je svaki sloj linearna funkcija. Prolaz ulaza kroz više slojeva linearnih funkcija je matematički zapravo kompozicija linearnih funkcija, a kompozicija linearnih funkcija je uvijek linearna funkcija. Kako bi rješili taj problem koristimo prijenosne funkcije koje su nelinearne funkcije, kako bi kompozicijama mogli dobiti bilo koju funkciju. Uobičajene prijenosne funkcije su funkcija identiteta, sigmoidalna funkcija, zglobnica te propusna zglobnica. Osim prijenosnih funkcija, danas se u slojevima umjetnih neuronskih mreža često koristi pažnja, koja osigurava nelinearnost sloja.

2.2.1. Funkcija identiteta

Funkcija identiteta (2.2) je osnovna prijenosna funkcija. Pošto je i ona sama linearna funkcija, koristi se isključivo u jednostavnijim problemima u kojima je linearna funkcija dovoljna.

$$f(x) = x \tag{2.2}$$

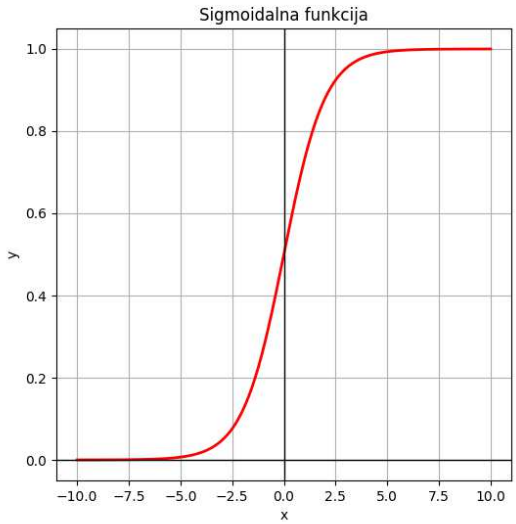


Slika 2.3. Funkcija identiteta

2.2.2. Sigmoidalna funkcija

Sigmoidalna funkcija (2.3) je funkcija koja je preslikava sve ulaze u interval $[0, 1]$ te ima jednu točku infleksije za $x = 0$. Za razliku od funkcije identiteta, ona je nelinearna funkcija, a osnovna primjena sigmoidalne funkcije je u binarnoj klasifikaciji.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.3}$$

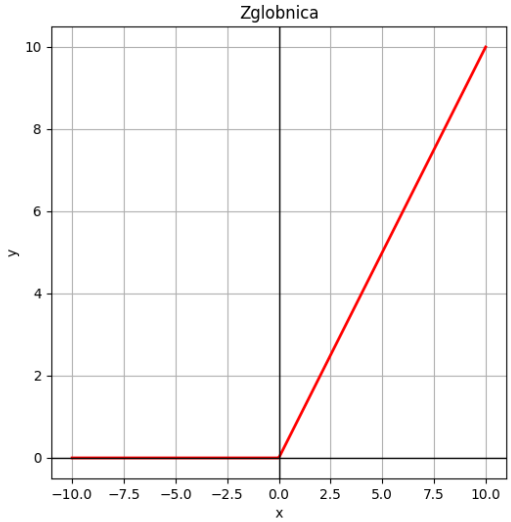


Slika 2.4. Sigmoidalna funkcija

2.2.3. Zglobnica

Zglobnica (2.4) je popularna aktivacijska funkcija. Njezine prednosti nad sigmoidalnom funkcijom su smanjenje problema nestajućeg gradijenta i računalna učinkovitost.

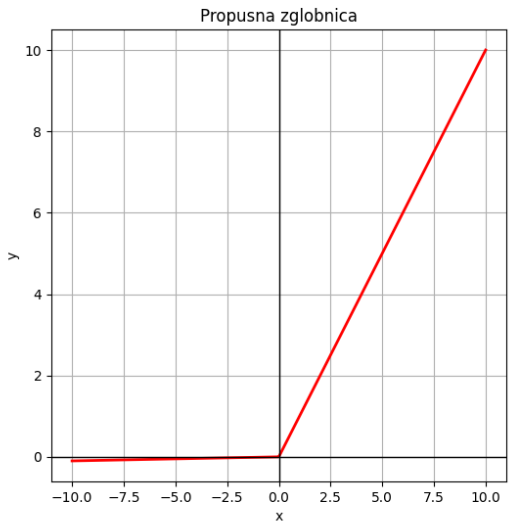
$$f(x) = max(0, x) \tag{2.4}$$



Slika 2.5. Zglobnica

Ipak, zglobnica ima problem umiruće zglobnice, koji je vrsta problema nestajućih gradijenata. Zbog toga često se koristi propusna zglobnica (2.5).

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases} \tag{2.5}$$



Slika 2.6. Propusna zglobnica

2.3. Difuzijski modeli

Difuzijski modeli javljaju se 2015. godine [7] i temelje se na tehnikama neravnotežne termodinamike, pogotovo difuzije. Oni su generativni algoritam strojnog učenja. Uče gustoću p(x) be oznaka te ju mogu uzrokovati. Difuzijski modeli su uče unatražni difuzijski proces, iz nasumično generiranog gaussovog šuma u svakom koraku izoštruju sliku i povećavaju broj detalja. Uglavnom se koriste u području računalnog vida. Glavne primjene su u područjima generiranja slika [8, 9], uklanjanju šuma [10, 11] te restoraciji slika [12, 13]. Difuzijski modeli stvaraju markovljev lanac difuzijskih koraka dodavanjem šuma na podatke te uče rekonstruirati podatke iz šuma. Cilj modela je naučiti funkciju gustoće razdiobe $\hat{p}(x)$ koja aproksimira stvarnu funkciju gustoće razdiobe $p(x)$(2.6), pri čemu su x varijable dostupne u skupu podataka, koje nazivamo osmotrivim varijablama.

$$\hat{p}(x) \approx p(x) \tag{2.6}$$

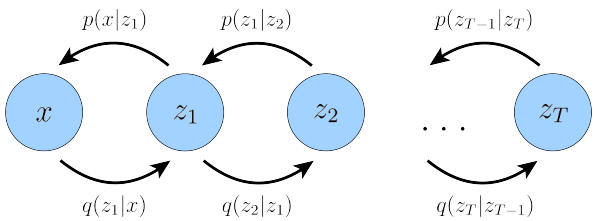
Osim ulaznih varijabli, u difuzijskim modelima koristimo i skrivene varijable, koje označavamo sa z, koje nisu dio skupa podataka, no one generiraju promatrane podatke. Združenu gustoću vjerojatnosi označavamo sa $\hat{p}(x, z)$. Do gustoće vjerojatnosti preko isključivo osmotrivih vjerojatnosti možemo doći izrazom (2.7).

$$\hat{p}(x) = \frac{\hat{p}(x, z)}{\hat{p}(z|x)} \tag{2.7}$$

Zbog nepoznavanja razdiobe $\hat{p}(z|x)$ aproksimiramo ju razdiobom $\hat{q}(z|x)$ (2.8).

$$\hat{q}(z|x) \approx \hat{p}(z|x) \tag{2.8}$$

Često su skrivene varijable višeg reda. Tada se nadalje ograničavamo da je svaka skrivena varijabla z_t uvjetovana isključivo neposrednom varijablom z_{t+1}. Tako stvaramo markovljev lanac skrivenih i osmotrivih varijabli (Slika 2.7.).



Slika 2.7. Markovljev lanac skrivenih i osmotrivih varijabli, preuzeto iz [14]

Detaljniji opis difuzijskih modela može se naći u [15].

Modeli koji provode otklanjanje šuma su obično transformeri ili U-net. Difuzijski modeli mogu modelirati vrlo kompleksne vjerojatnosne razdiobe, jer dodavanjem šuma iz nepoznate početne razdiobe dobivamo uravnoteženu razdiobu, koja je Gaussova razdioba sa funkcijom gustoće $\alpha e^{-\frac{||x||^2}{2}}$. 2020. godine javljaju se difuzijski probabilistički modeli uklanjanja šuma (DDPM) [16] koji poboljšavaju difuzijske modele.

2.4. Učenje neuronske mreže

Učenje umjetne neuronske mreže sastoji se od modela, gubitka i optimizacijskog postupka. Dijeli se na nadzirano učenje, kod kojeg su podatci u obliku (ulaz, izlaz) te tražimo preslikavanje iz ulaza u izlaz, nenadzirano učenje, kod kojeg nema ciljnih vrijednosti i traže se pravilnost u podatcima te podržano učenje, koje se temelji na pokušavanju s odgođenom nagradom. Algoritam učenje pri učenju ažurira težine neuronske mreže kako bi minimizirao grešku, koja je dana funkcijom gubitka. Gradijenti gubitka propagiraju se unatrag kroz mrežu te model optimizira težine koristeći optimizacijske metode poput stohastičkog gradijentnog spusta ili Adam-a.

2.4.1. Funkcija gubitka

Uobičajena funkcija gubitka kod regresije je funkcija srednje kvadratne greške (2.9), gdje je n broj podataka, Y_i stvaran podatk te $\hat{Y}_i$ predikcija modela.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.9}$$

Jedna od funkcija gubitka je L1 gubitak (2.10), sa istim oznakama kao i kod funkcije srednje kvadratne pogreške.

$$L1 = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.10}$$

Kao podvrsta funkcije L1 gubitka postoji zaglađeni L1 gubitak koji je manje osjetljiv na stršeće vrijednosti.

$$L1 = \frac{1}{n}\sum_{i_1}^{n} l_i \tag{2.11}$$

$$l_n = \begin{cases} \frac{1}{2}\frac{(y_n - \hat{y_n})^2}{\beta}, & |y_n - \hat{y_n}| < \beta \\ |y_n - \hat{y_n}| - \frac{1}{2}\beta, & |y_n - \hat{y_n}| \geq \beta \end{cases} \tag{2.12}$$

2.4.2. Propagacija pogreške unatrag

Propagacija pogreške unatrag (eng. Error Backpropagation) je postupak računanja svih parcijalnih derivacija, računamo parcijalne derivacije gubitka svakog podatka po težinama i po pomaku $\vec{b}$, koje koristimo kako bi optimizirali iznose težina neuronske mreže koristeći izračunate gradijente funkcije gubitka. Gradijente računamo korištenjem pravila ulančavanja kod funkcija derivacije (2.13), gdje je z funkcija ovisna o y, a y funkcija ovisna o x.

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} \tag{2.13}$$

2.4.3. Optimizacija

Kao metodu optimizacije koristimo gradijentni spust. Gradijent je smjer najbržeg rasta funkcije te koristeći njegovu negativnu vrijednost dobivamo smjer najbržeg pada funkcije, što koristimo kako bi minimizirali funkciju gubitka. Negativni gradijent množimo s parametrom ϵ, koji je broj između 0 i 1, kojeg nazivamo stopa učenja. Zbog brže konvergencije i sprečavanja zapinjanja u lokalnim minimumima koristi se promjenjiva stopa učenja. Primjer takvog optimizatora je stohastički gradijentni spust (2.14), pri čemu je w_{i+1} težina u sljedećoj iteraciji, w_i trenutna težina, η stopa učenja, ∇ gradijent težine i $L(w_i, x)$ gubitak trenutne težine, te Adam [17], koji donosi poboljšanja u odnosu na osnovni stohastički gradijentni spust.

$$w_{i+1} = w_i - \eta \cdot \nabla \cdot L(w_i, x) \tag{2.14}$$

3. Korišteni alati

U radu je korišten programski jezik Python i njegove biblioteke Numpy[1], Pytorch[2] i Pytorch Lightning[3]. Za vizualizaciju testiranja korišten je WANDB[4]. GPU na kojem je treniranje provedeno je NVIDIA GeForce RTX 3050.

3.1. Numpy

Numpy je osnovna biblioteka za rad s matricama i tenzorima. Omogućuje korištenje širokog spektra matematičkih funkcija, od kojih su za područje računalnog vida izuzetno bitne funkcije linearne algebre, pogotovo one za rad s matricama i vektorima. Numpy također nudi izuzetno dobre performanse koristeći optimizirani C kod. Osim toga, Numpy ima sintaksu visoke razine što ga čini vrlo jednostavnim za korištenje.

3.2. Pytorch

Pytorch je popularan okvir koja se koristi u području dubokog učenja. Pruža potporu za automatsku diferencijaciju i rad s više GPU-a, module i funkcije za olakšanu konstrukciju neuronskih mreža, aktivacijske funkcije, funkcije gubitka, optimizatore i slično. Osnovan tip objekta u Pytorchu je Pytorch tenzor koji nalikuje na Numpyeva višedimenzionalna polja.

[1] https://numpy.org
[2] https://pytorch.org
[3] https://lightning.ai/docs/pytorch/stable/
[4] https://wandb.ai/site

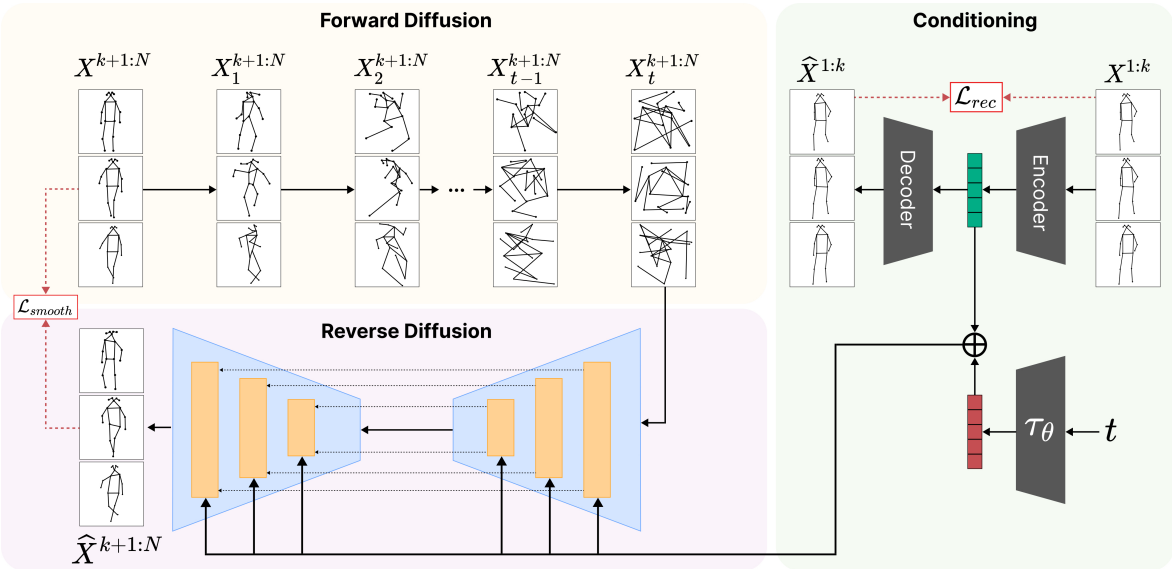
3.3. Pytorch Lightning

Pytorch Lightning je okvir otvorenog koda koja pruža sučelje visoke razine za Pytorch. Omogućuje provođenje treniranja i testiranja modela na visokoj razini. Zbog izrazito visoke razine abstrakcije izuzetno je jednostavan za korištenje, mogućuje lako upravljanje parametrima i lako praćenje parametara, kao i jako pregledan kod.

3.4. WANDB

Weights and Biases je platforma otvorenog koda koja omogućuje jednostvano praćenje i vizualizaciju parametara i rezultata. Integrairana je s bibliotekama Pytorch i Pytorch lightning. Zbog svoje integracije izuzetno je jednostavna za korištenje pri korištenju biblioteka Pytorch i Pytorch Lightning.

4. Model



Slika 4.1. Pregled modela [4]

Funkcionalnost modela je prikazana na slici 4.1. Ulaz modela je vektorska reprezentacija čovjeka. Pri treniranju koristi buduće slike te im dodaje šum, pomičući ključne točke čovjeka nasumičnim vektorom pomaka. Model potom uči kako ukloniti šum kako bi dobio stvarne buduće poze, to jest stvarne koordinate ključnih točaka čovjeka. Pri predviđanju model koristi poze osobe do tog trenutka, enkodira ih te ih koristi sa sljedećom pozom, kojoj prvo dodaje šum, a zatim s prošlim enkodiranim pozama ju pokušava rekonstruirati. Model radi više rekostrukcija buduće poze, te za svaku računa gubitak. Funkcija gubitka koju model koristi je zaglađeni L1 gubitak (2.11). Zbog činjenice da se generira više predviđanja, pri treniranju gubitak uzet je najmanji gubitak. Mjera anomalnosti modela, koja predstavlja razliku između predviđanja modela i stvarne poze, je zaglađeni L1 gubitak između rekonstrukcije i stvarnog podatka. Kao konačnu mjeru točnosti model koristi ROC-AUC(Receiver Operating Characteristic Area Under the Curve), koji je uobičajena generalizacijska mjera za balansirane probleme binarne klasifikacije.

5. Skupovi podataka

5.1. HR-ShanghaiTech

Slika 5.1. Skup podataka ShanghaiTech Campus [3]

Skup podataka HR-ShanghaiTech predstavljen je u [2] kao verzija skupa podataka ShanghaiTech Campus. ShanghaiTech Campus skup podataka je dobiven iz snimaka 13 kamera na kampusu sveučilišta ShanghaiTech, što ga čini vrlo opsežnim. Različite kamere također daju različite kuteve pogleda na scenu, dok snimanje kampusa osigurava realistične primjere i veliki broj različitih anomalnih događaja. Skup sadrži 107 testna videa sa raznim anomalnim događajima, 274515 slika za treniranje, 42883 slika za testiranje i 130 anomalna događaja. Od 107 videa, 6 imaju anomalne događaje koji nisu povezani s ljudima te je iz stoga napravljen podskup HR-ShanghaiTech koji izbacuje ta 6 videa, i u skupu podataka koristi preostalih 101 video. Rezolucija videa u ovom skupu podataka je 856 x 480.

5.2. HR-Avenue

Slika 5.2. Skup podataka CUHK Avenue [18]

CUHK Avenue je skup podataka dobiven iz avenije CUHK kampusa. Sadrži 16 videa za treniranje i 21 video za testiranje, snimljenih jednom kamerom, koji uključuju 15328 slika za treniranje i 15324 slika za testiranje. Skup podataka je zanimljiv jer sadrži neke slike koje uključuju kameru koja se trese i neke normalne sljedove poza koji se rijetko pojavljuju u skupu za treniranje. HR-Avenue je podskup skupa podataka CUHK Avenue predstavljen u [2]. Podskup je dobiven izbacivanjem slika sa anomalnim događajima koji ne uključuju ljude i slika u kojima osoba uključena u anomalan događaj nije vidljiva. Rezolucija videa u ovom skupu podataka je 640 x 360.

6. Eksperimenti

Funkcija gubitka korištena pri treniranju je zaglađeni L1 gubitak (2.11), optimizator korišten je Adam. Konačna preciznost izračunata je pomoću AUC-a.

6.1. HR-ShanghaiTech

6.1.1. Hiperparametri

Za treniranje na skupu podataka HR-ShanghaiTech korišteni su sljedeći hiperparametri:

- Broj epoha: 20
- Broj difuzijskih koraka: 10
- Broj predviđenih poza: 5
- Početna stopa učenja: 0.0001
- Veličina grupa: 4096

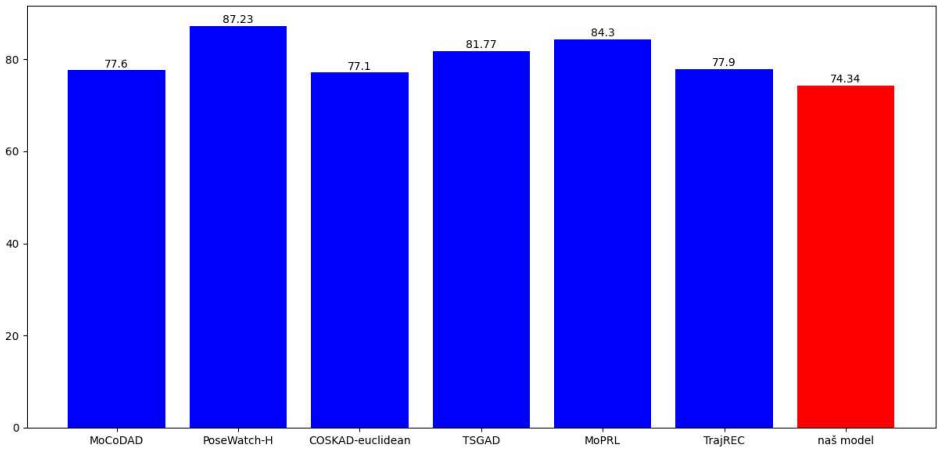
6.1.2. Rezultati

Slika 6.1. Funkcija gubitka kroz epohe treniranja

Zbog veličine skupa podataka, treniranje je provedeno na samo 20 epoha. Unatoč tome, možemo vidjeti kako je funkcija gubitka već počela konvergirati, što upućuje da povećanje epoha ne bi znatno poboljšalo performanse modela. Funkcija gubitka minimum postiže u petnaestoj epohi, gdje joj je vrijednost 0.037. Kao mjera točnosti modela korišten je AUC.

Model	AUC
MoCoDAD [4]	77.6
PoseWatch-H [19]	87.2
COSKAD-euclidean [1]	77.1
TSGAD [20]	81.8
MoPRL [21]	84.3
TrajREC [22]	77.9
MoCoDAD (naš)	74.3

Tablica 6.1. Rezultati na skupu podataka HR-ShanghaiTech



Slika 6.2. Vizualizacija AUC-a modela u usporedbi s drugim modelima

U usporedbi s današnjim state-of-the-art modelima[1], možemo primjetiti da naš model nešto slabije generalizira. Uočljivo je da i MoCoDAD, koji je uz naš model jedini difuzijski model, koji, za razliku od našeg modela koji koristi samo enkoder za prethodne poze, koristi enkoder i dekoder za prethodne poze kojima dodatno informira model u generiranju predviđanja, postiže slabije rezultate od najboljih modela. HR-ShanghaiTech je veliki skup podataka, koji stoga obuhvaća veliki broj poza, kako normalnih, tako i anomalnih, kao i različite perspektive zbog različitih korištenih kamera.

[1]Stranica s najboljim modelima na skupu podataka: https://paperswithcode.com/sota/video-anomaly-detection-on-hr-shanghaitech

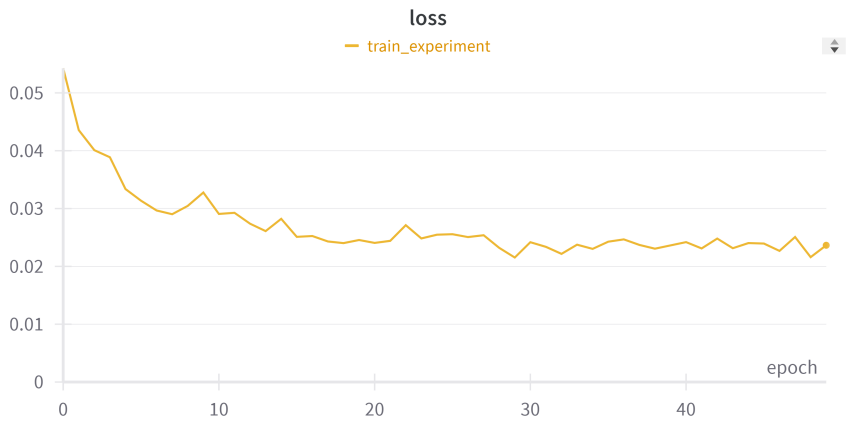
6.2. HR-Avenue

6.2.1. Hiperparametri

Za treniranje na skupu podataka HR-Avenue korišteni su sljedeći hiperparametri:

- Broj epoha: 50
- Broj difuzijskih koraka: 10
- Broj predviđenih poza: 5
- Početna stopa učenja: 0.0001
- Veličina grupa: 1024

6.2.2. Rezultati

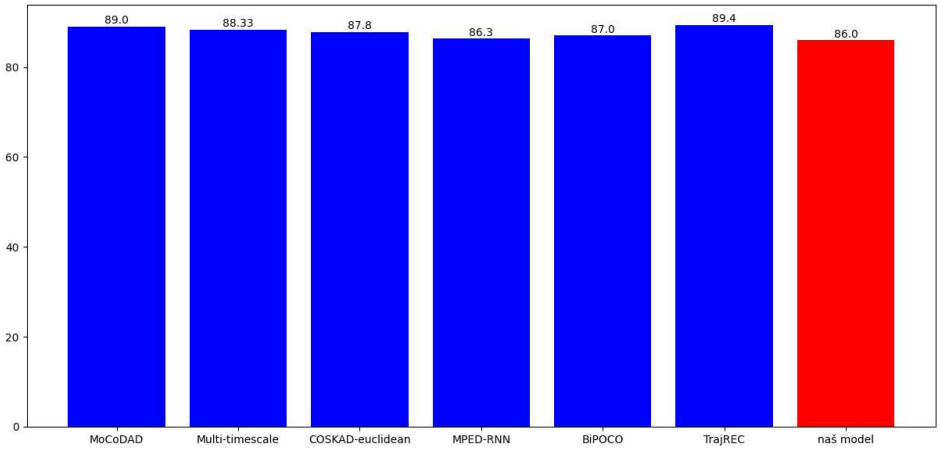


Slika 6.3. Funkcija gubitka kroz epohe treniranja

Treniranje je provedeno u 50 epoha i vidimo da je funkcija gubitka počela konvergirati. Minimum funkcije gubitka postignut je u dvadesetdevetoj epohi i iznosi 0.022. Ponovno koristimo AUC kao mjeru preciznosti.

Model	AUC
MoCoDAD [4]	89.0
Multi-timescale Prediction [23]	88.3
COSKAD-euclidean [1]	87.8
MPED-RNN [2]	86.3
BiPOCO [24]	87.0
TrajREC [22]	89.4
MoCoDAD (naš)	86.0

Tablica 6.2. Rezultati na skupu podataka HR-Avenue



Slika 6.4. Vizualizacija AUC-a modela u usporedbi s drugim modelima

Na ovom jednostavnijem skupu podataka model postiže bolje rezultate, iako treba napomenuti da se ovaj skup podataka rjeđe koristi za testiranje od prethodnog. Ipak, rezultati modela su približni najboljim modelima[2] na ovom skupu podataka, što upućuje da je model vrlo dobar u prepoznavanju anomalnih poza, kada su normalne poze međusobno konzistentne.

[2]Stranica s najboljim modelima: https://paperswithcode.com/sota/video-anomaly-detection-on-hr-avenue

Nakon ovog testiranja, skup podataka za treniranje smo podjelili na dva skupa, koje ćemo nazvati train1 i train2, te smo ih testirali. Rezultati su prikazani u sljedećoj tablici.

skup za učenje	AUC
train1	86.17
train2	84.44

Tablica 6.3. AUC skupova za treniranje train1 i train2

Vidimo da model treniran manjem skupu podataka train1 postiže bolje rezultate od modela treniranog cijelog skupa za treniranje, no razlika je mala, dok drugi skup postiže nešto lošije rezultate. Jedna od interpretacija ovakvih rezultata mogla bi biti da je bitnije da skup podataka obuhvaća što više različitih normalnih kretnji, no da povećanje vrlo sličnih normalnih kretnji u skupu podataka nema veliki utjecaj na rezultate.

7. Rasprava

Model najbolje rezultate postiže sa 10 difuzijskih koraka 7.1., to jest kada pozi dodajemo šum 10 puta. Ovo možemo objasniti činjenicom da ako dodajemo manje šuma, predviđanje će više odgovarati trenutnoj pozi, te će biti puno osjetljivije na bilo kakve promjene. S druge strane, ako koristimo više koraka, rekonstrukcija buduće poze će biti puno nasumičnija, te će se sve poze činiti normalnim. Unatoč tome bitno je napomenuti kako je AUC u svim slučajevima jako slučan te razliku također možemo pripisati nasumičnosti te bi drukčiji seed rezultatirao drukčijim rezultatima.

Broj difuzijskih koraka	AUC
4	85.33
7	85.50
10	86.00
12	85.99
15	85.08

Tablica 7.1. AUC u ovisnosti o broju difuzijskih koraka na HR-Avenue skupu podataka

Model je kod predviđanja buduće poze generirao 5 predviđanja. Bitno je napomenuti da to ustvari znači da je generirano 5· broj difuzijskih koraka poza kao predviđanje, zato što svaki difuzijski korak generira 5 predviđanja te se sva uzimaju u obzir. Ovaj dio nije bio mijenjan te je nepoznato kako bi promjena broja generiranih predviđanja utjecala na preciznost modela. U [4] za predviđanje generirali su 50 poza. U svojem članku izvještavaju da preciznost raste povećanjem broja generairanih poza, no da nakon 50 poza preciznost prestaje rasti.

Model je pokazao dobre rezultate na jednostavnijem HR-Avenue skupu podataka, no imao je lošije rezultate na opsežnijem i zahtjevnijem HR-ShanghaiTech skupu podataka. Ovakvi rezultati dovode u pitanje skalabilnost ovakvog modela. Ovakav difuzijski model stvara predviđanja temeljena na zapaženim normalnim sljedovima poza, no rezultati

dovode u pitanje koliko je to dobro na velikim skupovima podataka, gdje je jako puno različitih normalnih sljedova poza. Rješenje ovog problema moglo bi biti u podešavanju broja difuzijskih koraka i broja generiranih predviđanja. Još jedan pristup rješavanju ovog problema bio bi mjenjanje distribucije iz koje model stvara nasumični šum. Minimiziranje distribucije iz koje se šum izvlači te povećavanje broja difuzijskih koraka bi moglo biti jedno rješenje.

Daljnja istraživanje bi se osim na prethodno navedene probleme mogla i usredotočiti na problem ljudi koji su samo djelomično u kadru. Naš model pretpostavlja da su sve ključne točke osobe vidiljive te je napravljen da predviđa na takvim osobama. Proširenje bi osim toga moglo biti i u samoj vektorskoj reprezentaciji osobe. Difuzijski bi modeli možda postizali bolje rezultate za veći broj ključnih točaka osobe. Zanimljivo područje daljnjeg istraživanja moglo bi i biti testiranje modela na skupovima podataka drukčijim od onog na kojem je treniran kako bi se bolje mogla ocjeniti sposobnost generalizacije modela. Možda se poboljšanje nalaz u treniranju modela na jednostavnim skupovima podataka te dodatnom podešavanju hiperparametra kako bi model mogao dobro generalizirati.

8. Zaključak

Difuzijski modeli sve prisutniji su u području računalnog vida. U ovom radu govorili smo o primjeni difuzijskih modela za prepoznavanje anomalnih poza kod ljudi. Definirali smo što su umjetne neuronske mreže i difuzijski modeli. Dali smo pregled o aktivacijskim funkcijama te postupku učenja umjetnih neuronskih mreža i njihovoj optimizaciji, gdje smo spomenuli gradijenti spust.

Nakon definiranja pojmova spomenuli smo korištene tehnologije te smo nakon toga objasnili kako model funkcionira te skupove podataka na kojima je model treniran i testiran. Potom smo proveli eksperimente. Eksperimenti su nam pokazali da model postiže dobre rezultate za jednostavnije skupove podataka, no rezultati su bili gori za veće skupove. Razlog tome prepoznali smo kao veliki broj različitih sljedova normalnih poza, kao i veliki broj različitih sljedova anomalnih poza.

Daljnja istraživanja mogla bi se usredotočiti na poboljšavanje hiperparametara. Također se mogu baviti prepoznavanjem anomalnih poza kod ljudi koji nisu potpuno u kadru. Model bi se isto tako moglo testirati na još skupova podataka kako bismo bolje ustanoviti sposobnosti modela.

Literatura

[1] A. Flaborea, G. D'Amely, S. D'Arrigo, M. A. Sterpa, A. Sampieri, i F. Galasso, "Contracting skeletal kinematics for human-related video anomaly detection", 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2301.09489

[2] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, i S. Venkatesh, "Learning regularity in skeleton trajectories for anomaly detection in videos", 2019. [Mrežno]. Adresa: https://arxiv.org/abs/1903.03295

[3] W. Liu, W. Luo, D. Lian, i S. Gao, "Future frame prediction for anomaly detection – a new baseline", 2018. [Mrežno]. Adresa: https://arxiv.org/abs/1712.09867

[4] A. Flaborea, L. Collorone, G. D'Amely, S. D'Arrigo, B. Prenkaj, i F. Galasso, "Multimodal motion conditioned diffusion model for skeleton-based video anomaly detection", 2023. [Mrežno]. Adresa: https://arxiv.org/abs/2307.07205

[5] O. Hirschorn i S. Avidan, "Normalizing flows for human pose anomaly detection", 2023. [Mrežno]. Adresa: https://arxiv.org/abs/2211.10946

[6] C. Yan, S. Zhang, Y. Liu, G. Pang, i W. Wang, "Feature prediction diffusion model for video anomaly detection", u *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023., str. 5504–5514. https://doi.org/10.1109/ICCV51070.2023.00509

[7] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, i S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics", 2015. [Mrežno]. Adresa: https://arxiv.org/abs/1503.03585

[8] P. Xing, H. Wang, Y. Sun, Q. Wang, X. Bai, H. Ai, R. Huang, i Z. Li, “Csgo: Content-style composition in text-to-image generation”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.16766

[9] A. Eldesokey i P. Wonka, “Build-a-scene: Interactive 3d layout control for diffusion-based image generation”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.14819

[10] C. Niu, C. Wiedeman, M. Li, J. S. Maltz, i G. Wang, “3d photon counting ct image super-resolution using conditional diffusion model”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.15283

[11] H. Asgariandehkordi, S. Goudarzi, M. Sharifzadeh, A. Basarab, i H. Rivaz, “Denoising plane wave ultrasound images using diffusion probabilistic models”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.10987

[12] C. Yue, Z. Peng, J. Ma, i D. Zhang, “Enhanced control for diffusion bridge in image restoration”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.16303

[13] S. Tu, W. Yang, i B. Fei, “Taming generative diffusion for universal blind image restoration”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.11287

[14] C. Luo, “Understanding diffusion models: A unified perspective”, 2022. [Mrežno]. Adresa: https://arxiv.org/abs/2208.11970

[15] J. Srzić, “Pronalaženje anomalija difuzijskim uklanjanjem šuma”, 2024.

[16] J. Ho, A. Jain, i P. Abbeel, “Denoising diffusion probabilistic models”, 2020. [Mrežno]. Adresa: https://arxiv.org/abs/2006.11239

[17] D. P. Kingma i J. Ba, “Adam: A method for stochastic optimization”, 2017. [Mrežno]. Adresa: https://arxiv.org/abs/1412.6980

[18] C. Lu, J. Shi, i J. Jia, “Abnormal event detection at 150 fps in matlab”, 2013.

[19] G. A. Noghre, A. D. Pazho, i H. Tabkhi, “Posewatch: A transformer-based architecture for human-centric video anomaly detection using spatio-temporal pose tokenization”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2408.15185

[20] ——, “An exploratory study on human-centric video anomaly detection through variational autoencoders and trajectory prediction”, 2024. [Mrežno]. Adresa: https://arxiv.org/abs/2406.15395

[21] S. Yu, Z. Zhao, H. Fang, A. Deng, H. Su, D. Wang, W. Gan, C. Lu, i W. Wu, “Regularity learning via explicit distribution modeling for skeletal video anomaly detection”, 2021. [Mrežno]. Adresa: https://arxiv.org/abs/2112.03649

[22] A. Stergiou, B. D. Weerdt, i N. Deligiannis, “Holistic representation learning for multitask trajectory anomaly detection”, 2023. [Mrežno]. Adresa: https://arxiv.org/abs/2311.01851

[23] R. Rodrigues, N. Bhargava, R. Velmurugan, i S. Chaudhuri, “Multi-timescale trajectory prediction for abnormal human activity detection”, 2019. [Mrežno]. Adresa: https://arxiv.org/abs/1908.04321

[24] A. M. Kanu-Asiegbu, R. Vasudevan, i X. Du, “Bipoco: Bi-directional trajectory prediction with pose constraints for pedestrian anomaly detection”, 2022. [Mrežno]. Adresa: https://arxiv.org/abs/2207.02281

Sažetak

Detekcija anomalnih sljedova poza osoba difuzijskim otklanjanjem šuma

Marko Kukolj

Raspoznavanje kretnji osoba bitan je problem u području računalnog vida, koji nije praktično modelirati nadziranim učenjem. Ovaj rad promatra alternativni pristup utemeljen na detekciji anomalnih poza koristeći skupove sljedova za učenje. Model anomalne poze prepoznaje modeliranjem predviđanja stvarnih kretnji te odstupanjem predviđene kretnje od stvarne kretnje.

Ključne riječi: Difuzija; anomalne poze; otklanjanje šuma

Abstract

Detection of anomalous sequences of human poses by denoising diffusion

Marko Kukolj

Recognizing movement of people is an important problem in the field of computer vision that is not practical to model with supervised learning. This work looks at an alternative approach based on the detection of anomalous poses using sets of sequences for learning. The model recognizes anomalous poses by modeling the prediction of real movement and using the deviation between the prediction and the actual movement.

Keywords: Diffusion; anomalous poses; denoising