FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

## Projekt iz predmeta Raspoznavanje uzoraka

## Ak. god. 2020/21

**Detekcija i lokalizacija nenormalnih događaja uporabom metode dubokog učenja**

Autori:

Marin Rađa

Ivona Farkaš

Fran Pugelnik

Branimir Ričko

Antonio Kuzminski

## siječanj, 2021.

Sadržaj

[Projekt iz predmeta Raspoznavanje uzoraka 1](#_Toc60861122)

[Ak. god. 2020/21 1](#_Toc60861123)

[siječanj, 2021. 1](#_Toc60861124)

[1. Projektni zadatak 3](#_Toc60861125)

[1.1 Opis projektnog zadatka 3](#_Toc60861126)

[1.2 Pregled i opis srodnih rješenja 3](#_Toc60861127)

[1.2.1 Skupovi podataka 4](#_Toc60861128)

[1.3 Konceptualno rješenje zadatka 5](#_Toc60861129)

[1.3.1 Arhitektura prostorno vremenske konvolucijske neuronske mreže 6](#_Toc60861130)

[2. Postupak rješavanja zadatka 7](#_Toc60861131)

[2.1 Priprema skupova podataka 7](#_Toc60861132)

[3. Ispitivanje rješenja 8](#_Toc60861133)

[4. Opis programske implementacije rješenja 9](#_Toc60861134)

[5. Zaključak 10](#_Toc60861135)

[6. Literatura 11](#_Toc60861136)

# Projektni zadatak

## Opis projektnog zadatka

Otkrivanje nenormalnih događaja u scenama gdje postoje gužve, izuzetno je izazovno područje računalnog vida zbog ozbiljnih zakrivanja objekata scene, različitih intenziteta gustoće gužve i kompleksnih mehanika ljudskih gužvi, odnosno nepravilnim kretanjima ili aktivnostima gužve u okolini.

Široko korištenje nadzornih sustava u kolodvorima, zračnim lukama, na cestama ili u kupovnim centrima rezultira sve većoj količini video zapisa. Osim što postoji velika potreba za prepoznavanjem objekata i njihovog ponašanja isto tako potreba za detekcijom nenormalnih događaja postaje sve veća. Automatska detekcija nenormalnih događaja ili aktivnosti tijekom dugih tokova video zapisa je jedna od najvažnijih komponenata inteligentnog video nadzora.

Dodatno, postoje potencijalno opasne aktivnosti u okolini gdje postoji mnoštvo. Prilikom kakvih društvenih događaja koji se odvijaju na javim mjestima poput stadiona, trgova i sl., postoji mogućnost izbijanja stampeda ili bilo kakvog iznenadnog događaja kojim bi se potencijalno narušila sigurnost mnoštva. Za izbjegavanje takvih situacija potrebno ih je detektirati na vrijeme i reagirati adekvatno stoga ima smisla uložiti napor u razvitak sustava za detekciju nenormalnih događaja.

## Pregled i opis srodnih rješenja

Za detekciju nenormalnih događaja iz perspektive statične kamere postoji dosta literature. Primjerice u [2] je dan opsežan pregled postojećih metoda poput histograma, Bayesovih mreža, neuronskih mreža, stroja potpornih vektora, metoda grupiranja i dr., ali gotovo sve te metode su ograničene činjenicom da rade zadovoljavajuće raštrkanim scenama, odnosno scenama gdje su ljudi jasno vidljivi i među pojedinih ljudi nema preklapanja.

Predloženo je nekoliko modela inspiriranih fizikom gdje se primjerice u [3] za detekciju kretanja mnoštva koristi tehnika optičkog toka kako bi se lokaliziralo kretanje mnoštva uz model socijalne sile (engl. *social force model* - SFM) kako bi se pokušalo izdvojiti nenormalne događaje. [32] predlaže metodu korištenja optimizacije rojem čestica gdje se pokušava skrenuti čestice u regiju slike s najvećim kretanjem mnoštva.

Metode bazirane na kretanju obično izvlače informaciju poput optičkog toka i trajektorija kretanja iz prostorno vremenskog volumena (nekoliko sličica unatrag u vremenu) i tako modeliraju kretanje mnoštva.

Jedan od glavnih izazova kod detekcije nenormalnih događaja je detekcija u vremenskoj i prostornoj domeni [4]. To znači da je potrebno naći sličice u kojima se događaju anomalije i lokalizirati regije unutar sličice u kojima je nenormalan događaj.

Osim što je problem detektirati nenormalne događaje, drugi problem je što nema formalne definicije nenormalnog događaja. Ta definicije ovisi uvelike o problemu koji se rješava.

Postojeći pristupi detekciji nenormalnih događaja mogu se podijeliti u dvije kategorije: 1) metode gdje je objekt u centru razmatranja i 2) holističke metode gdje je u centru zbivanja pojedinac. U klasičnom modelu gdje je u centru zbivanja objekt, mnoštvo se tretira kao skup pojedinaca. Za razumijevanje zbivanja mnoštva, potrebno je segmentirati mnoštvo interesa u objekte. Metode gdje je objekt u centru razmatranja suočavaju se s problemima kod detekcije objekata, praćenja trajektorija i prepoznavanja ponašanja u krcatim scenama. Jednom kada kompleksnost scene naraste u smislu količine objekata, može se primijetiti značajna degradacija kvalitete detekcije objekata. Drugi vid modela, holistički model, cilj nije detekcija i praćenje svakog pojedinca, nego, mnoštvo se gleda kao cijeli entitet. Tipične značajke poput prostorno vremenskih gradijenata ili optičkog toka su korištene kod ove metode za lokalizaciju regija sa zanimljivim kretanjem.

Detekcija nenormalnih događaja može biti definirana kao problem binarne klasifikacije u razrede „normalan događaj“ i „nenormalan događaj“.

U skoro vrijeme puno radova demonstriralo je korisnost konvolucijskih neuronskih mreža (engl. CNN) [6] u različitim zadatcima poput klasifikacije i detekcije, prepoznavanje teksta, detekcije rubova, detekcija lica i dr. Korištenje konvolucijskih neuronskih mreža kod video zapisa pokazano je u [7] gdje se koristi 3D CNN za prepoznavanje ljudskih akcija gdje 3D označava niz slika u vremenu.

### Skupovi podataka

Korišteni su skupovi podataka UCSD i UMN. UCSD dolazi s dva skupa podataka: „ped1“ i „ped2“ dok UMN dolazi s 3 podskupa: „indoor“, „lawn“ i „plaza“. Neki od primjera uzoraka slika iz svakog skupa podataka dani su na slikama koje slijede.

#### UCSD skup podataka



Slika 1 Prikaz nekih slika iz ped1

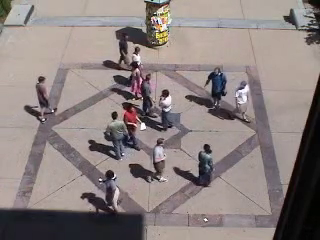
Slika 2 Prikaz nekih slika iz ped2

Moguće je vidjeti kako se na slikama iz USCD skupova nalazi mnoštvo promjenjive gustoće. Kamera se nalazi pod nekim kutom, a ujedno se događa puno prekrivanja pojedinih ljudi na slikama. Ovi skupovi dolaze s direktorijima „Train“ i „Test“, međutim direktorij „Train“ ne sadrži niti slike s nenormalnim događajima kako je i napisano od strane autora skupa podataka te se za učenje mreže koristi direktorij „Test“ koji se odgovarajuće podijeli na manje dijelove.

#### UMN skup podataka



  
Slika 3 Prikaz nekih slika iz skupa "indoor"

  
Slika 4 Prikaz nekih slika iz skupa "lawn"

Slika 5 Prikaz nekih slika iz skupa "plaza"

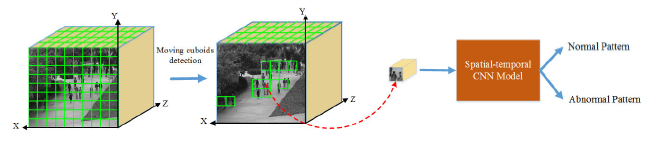
Prikazom primjera iz skupa podataka UMN moguće također se vidi kako se pojavljuje mnoštvo različite gustoće, objekti se međusobno prekrivaju, a onda u jednom periodu nastupa nenormalan događaj koji je u ovom skupu označen u gornjem lijevom uglu crvenim slovima. Taj natpis ne utječe na kvalitetu klasifikacije mreže u smislu da bi ga se trebalo izrezati jer je natpis statičan, a statične stvari se ignoriraju prilikom izvlačenja SVOI-ja.

## Konceptualno rješenje zadatka

U ovom radu se fokusira na izazov detekcije nenormalnih događaja u prostornoj i vremenskoj domeni video zapisa (pojedinih sličica) gdje se nalazi mnoštvo promjenjive gustoće.

Za detekciju lokalnih anomalija koje se nalaze u lokalnim regijama tipična metoda bi bila da se po cijela slika podijeli u više manjih nepreklapajućih dijelova pa se onda svaki od tih manjih dijelova obrađuje bez obzira nosi li taj dio kakvu informaciju ili ne. Takva metoda je uvelike računalno zahtjevna i lako daje pogrešne detekcije anomalija. Bolja metoda koja bi smanjila računalnu složenost i poboljšali detekciju anomalija ima smisla obrađivati samo dijelove slike koji pružaju neku korisnu informaciju poput kretanja ljudi. S druge strane svi dijelovi slike koji ne pružaju nikakvu informaciju se ne obrađuju.

Postupak kojim se izvlače korisni dijelovi slike naziva se optički tok. Konkretnije, optički tok se izračunava Farnebackovom metodom [8] između pojedinih sličica video zapisa. Koliko sličica sudjeluje u izračunavanju optičkog toka može se odrediti proizvoljno, a ovdje se koristi 7 sličica u vremenskoj domeni. Jednom kada se izračuna optički tok, koriste se dijelovi slike na kojima je optički tok najizraženiji, a ta se izraženost može iskazati kao postotak piksela koji se kreću unutar pojedinih kvadrata (ovdje 65%). Ta područja su kvadrati dimenzija koje se također mogu odrediti proizvoljno, ovisno o skupu podataka koji se koristi. Ti kvadrati se protežu u vremenskoj domeni kroz broj sličica koji je određen prije i na taj način tvore volumen u prostorno vremenskoj domeni (engl. *spatial-temporal volume of interest* – SVOI). Svaki SVOI daje se na ulaz CNN mreže na analizu je li u tom SVOI-ju nenormalan događaj ili nije. Opisani postupak moguće je vidjeti na slici 1 koja je preuzeta iz [1].



Slika 6 Prikaz izvlačenja SVOI-ja iz niza sličica iz video zapisa

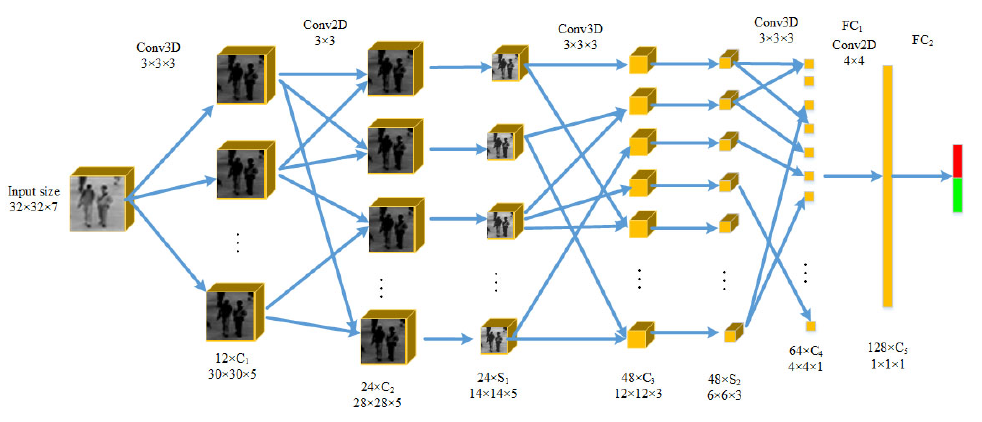
CNN je lokalno povezan model neuronske mreže koji izvlači lokalne značajke ograničavajući ulaze koje dobivaju skriveni slojevi. Tipična CNN se sastoji od konvolucijskih slojeva nakon kojih slijede potpuno povezani slojevi. Konvolucijski sloj koristi skup filtara kako bi obradio male lokalne regije ulazne slike. Korak poduzorkovanja (engl. *subsampling*) generira filter manje rezolucije na način da uzima najveću vrijednost unutar nekog definiranog prozora koji se kreće po ulaznom filteru.

U CNN-u konvolucija se računa u konvolucijskim slojevima. Formalno, *j-*ta aktivacijska mapa *aij* u *i­*-tom sloju je definirana kao gdje označava nelinearnu funkciju definiranu kao . predstavlja jezgru filtera, indeks skupa aktivacijskih mapa povezanih s trenutnom aktivacijskom mapom u -tom sloju, \* označava operaciju konvolucije, je dodatan skalar trenutnog sloja. Postupak učenja mijenja vrijednosti i za bolje izvlačenje značajki iz slike.

Za bolje izvlačenje informacija primjenjuju se prostorno vremenske konvolucije, a to se postiže korištenjem 3D jezgre koja zatim prolazi prostorno vremenskim volumenom, odnosno SVOI-jem. Takvim konvolucijama su aktivacijske mape povezane s više sličica u prethodnim slojevima čime se hvata koncept kretanja. Formalno, prostorno vremenska konvolucija između 3D jezgre i prostorno vremenskog volumena može se definirati kao:

Gdje predstavljaju visinu, širinu i vremensku domenu 3D jezgre, odgovara veličini prostorno vremenskog volumena .

### Arhitektura prostorno vremenske konvolucijske neuronske mreže



Slika 7 Arhitektura prostorno vremenske CNN

Arhitektura prostorno vremenske konvolucijske neuronske mreže prikazana je na slici 2 koja je preuzeta iz [1]. Mreža se sastoji od 8 slojeva, a na ulaz dolazi SVOI koji je dimenzija , gdje 32 označava veličinu kvadrata, a 7 vremensku dubinu SVOI-ja. Iako je veličina kvadrata SVOI-ja različite veličine za različite skupove podataka, svaka se veličina skalira na 32 kako je i prikazano na slici. Nakon ulaza u mrežu, primjenjuje se prva 3D konvolucija s veličinom jezgre pri čemu se koristi 12 filtera kojima se želi učiti različite značajke i zatim se generira 12 različitih aktivacijskih mapa kako je prikazano u sloju . Veličina svake mape je jer dolazi do gubitka rubne informacije sa svake strane volumena pošto nije korišteno nikakvo ispunjavanje slike kako bi se očuvala veličina nakon konvolucije. Nakon sloja primjenjuje se 2D konvolucija s veličinom jezgre koja prolazi svakim kvadratom SVOI-ja čime se dimenzija smanjuje na , ali ne prolazi vremenskom domenom SVOI-ja i time ta dimenzija ostaje očuvana. Nakon 2D konvolucije u sloju dolazi na red prvo poduzorkovanje veličine čime se želi dalje prenijeti važna informacija, ali u manjem formatu i nakon ove operacije svaka aktivacijska mapa je veličine (sloj ). Sljedeća je 3D konvolucija s jezgrom veličine čime se dimenzija svake mape smanjuje na . Nakon toga slijedi još jedno poduzorkovanje i 3D konvolucije gdje se u sloju dobivaju 64 aktivacijske mape veličine . Zadnja 3D konvolucija generira 128 aktivacijskih mapa koje su veličine u sloju . Svim konvolucijama i poduzorkovanjima se na kraju dobio jedan vektor koji ima 128 vrijednosti i koji se zatim linearnim slojem preslikava u dvije vrijednosti koje su na izlazu iz mreže. Dva broja su iz razloga jer želimo razlikovati normalne i abnormalne SVOI-je. Nakon dobivenih brojeva, primjenjuje se logistička regresija kako bi se normalizirale vjerojatnosti pojedinih razreda. Vjerojatnosti pojedinih razreda su dane s:

gdje označavaju prvi odnosno drugi broj koji su dobiveni na izlazu mreže.

# Postupak rješavanja zadatka

## Priprema skupova podataka

Skupovi podataka UCSD i UMN dolaze s označenim nenormalnim sličicama. UCSD dolazi s jednom matlab datotekom za svaki podskup (ped1 i ped2) gdje je zapisano koje su sličice abnormalni događaji. Radi pogodnosti se takav zapis prebacio u jednu tekstualnu datoteku kako bi se lakše učitao i koristio u programu. UMN skup podataka je jedan video zapis koji traje malo više od 4 minute gdje su pojedinih podskupovi spojeni zajedno i jedan počinje kada drugi završi. Svaki od tih podskupova u jednom trenutku počinje imate nenormalne događaje, a oni su označeni crvenim tekstom u gornjem lijevom uglu. Kako nije pogodno koristiti videozapis jer nema oznake po svakoj sličici, prvo je taj videozapis podijeljen u tri dijela. Svaki dio je jedan podskup (indoor, lawn i plaza).

Sljedeći korak je podijeliti svaki taj videozapis u pojedine sličice. Jednom kada je to napravljeno, u tekstualnu datoteku za svaki podskup je zapisano koje sve sličice sadrže abnormalne događaje.

## Izvlačenje SVOI-ja iz skupa podataka

Kako bi dobili SVOI potrebno je izračunati optički tok Farnebackovom metodom na nekom (unaprijed određenom) broju sličica.

# Ispitivanje rješenja

# Opis programske implementacije rješenja

# Zaključak

# Literatura

[1] Zhou S., Shen W., Zeng D., et al.: Spatial-temporal Convolutional Neural Networks for Anomaly Detection and Localization in Crowded Scenes. Signal Processing: Image Communication 47, pp. 358-368 (2016)

[2] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM computing surveys (CSUR) 41(3), 15 (2009)

[3] Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 935–942. IEEE (2009)

[4] Nam, Y., Hong, S. Real-time abnormal situation detection based on particle advection in crowded scenes. J Real-Time Image Proc 10, 771–784 (2015)

[5] Cong, Y., Yuan, J., Liu, J.: Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition 46(7), 1851–1864 (2013)

[6] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)

[7] Ji, S., Xu, W., Yang, M., Yu, K.: 3d convolutional neural networks for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(1), 221–231 (2013)

[8] OpenCV, Optical flow, <https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html>, posjećeno 6.12.2020.