

# Semantička segmentacija 3D snimki unutrašnjih prostora primjenom umjetne neuronske mreže

---

**Stipić, Nikola**

**Undergraduate thesis / Završni rad**

**2020**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:200:415614>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-08-04**

*Repository / Repozitorij:*

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU  
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH  
TEHNOLOGIJA**

**Sveučilišni studij**

**SEMANTIČKA SEGMENTACIJA 3D SNIMKI UNUTRAŠNJIH  
PROSTORA PRIMJENOM UMJETNE NEURONSKE MREŽE**

**Završni rad**

**Nikola Stipić**

**Osijek, 2020.**

## Sadržaj

1. UVOD.....	3
1.1. Zadatak završnog rada.....	4
2. PREGLED PODRUČJA TEME .....	5
3. ARHITEKTURA MREŽE .....	7
3.1. GSPN .....	7
3.2. R-PointNet.....	12
3.3. Učenje mreže .....	14
3.4. Inferencija .....	15
4. ISPITNI PODATKOVNI SKUPOVI.....	16
4.1. ScanNet.....	16
4.2. PartNet.....	17
4.3. NYUv2 .....	17
5. EKSPERIMENTALNA EVALUACIJA.....	19
5.1. Sklopovlje.....	19
5.2. Učenje neuronske mreže .....	20
5.3. Rezultati .....	24
6. ZAKLJUČAK .....	29

## 1. UVOD

Računalni vid je područje umjetne inteligencije koje se razvija sve većom brzinom i koristi u sve više industrijskih primjena. Većina zadaća računalnog vida je analiziranje slika u svrhu donošenja odluke o izvršavanju radnje, raspoznavanju objekata, brojanju instanci objekata i slično.

Prepoznavanje i razumijevanje sadržaja slike osnovni je problem od kojeg polaze svi ostali problemi računalnog vida. Semantička segmentacija je problem određivanja i pripadnosti razredu svakog piksela na slici nekoj od semantičkih kategorija. Primjeri kategorija su stol, hladnjak, vrata, slika, krevet i umivaonik.

Razumijevanje scene jedna je od značajnijih zadaća u računalnom vidu te semantička segmentacija ima važnu ulogu u istoj. Razlikujemo semantičku segmentaciju otvorenih i zatvorenih prostora. Problemi s kojima se susrećemo u semantičkoj segmentaciji su nedostatak svjetla, različita osvjetljenja, kut gledanja, varijacije u boji, razlikovanje instanci istog objekta i slično. Do većine tih smetnji dolazi u zatvorenim prostorima u kojima se nalazi mnoštvo malih objekata.

Iako semantička segmentacija trenutno nema široku industrijsku primjenu, koristi se u robotima i autonomnim vozilima za raspoznavanje okoline i njihovo snalaženje u istoj te time pridonosi jednoj od najunosnijih industrija, automobilskoj.

Za razliku od uobičajenog dvodimenzionalnog pristupa, u kojem se obrađuje ulaz u dvodimenzionalnom formatu kao što su slike, u ovom se radu razmatraju zahtjevniji, ali zahvalniji trodimenzionalni pristupi problemu semantičke segmentacije. U trodimenzionalnim pristupima ulaz su 3D snimke predstavljene u formatima kao što su oblak točaka ili „voxel grid“ te je njihova uloga opisana u drugom poglavlju. Konkretni problem koji se razmatra u ovom radu je semantička segmentacija 3D snimki unutrašnjih prostora zgrada koja pronalazi primjenu u navigaciji robota.

Microsoft Kinect kamera, koja može zabilježiti RGB boje te dubinu piksela, dovela je do mnogih znanstvenih radova vezanih za unutrašnje prostore. U scenama s nejednakim osvjetljenjem ili nedostatkom istoga, korištenje RGB-a uz pomoć 3D pristupa znatno smanjuje te probleme. RGB-D kamere susreću se s nekim poteškoćama kao što je šum, oštećenost i pomaknutost piksela.

Semantička segmentacija može se provoditi na RGB i RGB-D slikama, videima te 3D oblacima točaka snimljenih 3D LiDAR-om, odnosno dobivenih fuzijom više snimki.

U ovom radu opisan je postupak semantičke segmentacije 3D snimki unutrašnjih prostora primjenom umjetne neuronske mreže. U drugom poglavlju bit će opisani tradicionalni pristupi, a zatim pristupi korištenjem umjetnih neuronskih mreža te prethodno korišteni pristupi i aktualni primjeri rješenja ovog problema. U trećem poglavlju opisani su neuronska mreža za semantičku segmentaciju i segmentaciju na instance objekata R-PointNet te sastavni dio te mreže Generative Shape Proposal Network(GSPN) i njegov položaj u R-PointNet-u, njihova arhitektura i način rada. Zatim su opisani ispitni podatkovni skupovi koji se koriste za ispitivanje semantičke segmentacije prostorija. U četvrtom poglavlju opisana je konkretna implementacija razmatrane mreže na osobnom računalu te su na kraju opisani rezultati evaluacije uspješnosti mreže nad ispitnim podatkovnim skupovima i dan zaključak o korisnosti razmatranog pristupa te neke ideje za unaprjeđenje istoga.

## **1.1. Zadatak završnog rada**

Zadatak ovog završnog rada je istražiti i opisati aktualne pristupe semantičkoj segmentaciji 3D snimki unutrašnjih prostora primjenom umjetne neuronske mreže. Uz to, potrebno je i implementirati jedan od postojećih pristupa na osobnom računalu.

## 2. PREGLED PODRUČJA TEME

U tradicionalnim pristupima problemu semantičke segmentacije, slike se većinom segmentiraju u povezana područja ili super-piksele, zatim se iz slike izlučuju značajke kao što su SIFT, HOG, SURF i slični.[\[3\]](#) Ručno definirane značajke se razlikuju od suvremenih pristupa u kojima konvolucijska neuronska mreža sama izlučuje značajke kao što su otkrivanje ruba. Takve, ručno definirane značajke, se zatim prosljeđuju u klasifikator koji označuje svako područje ili super-piksel.

Korištenje RGB uz dubinske slike znatno je unaprijedilo rezultate semantičke segmentacije. Budući da su za suvremene metode u računalnom vidu ključni veliki izvori podataka, Silberman i Fergus [\[8\]](#) su izradili NYU-V1 ispitni podatkovni skup koji sadrži snimke unutrašnjih prostorija snimljene RGB i dubinskim kamerama iz Microsoft Kinecta. Ovaj je podatkovni skup kao pionir u tom području utjecao na primjenu i razvoj 3D semantičke segmentacije. Silberman i Fergus, također autori NYU-V1 ispitnog podatkovnog skupa, koristili su CRF(conditional random fields) pristup problemu semantičke segmentacije. Ren, Bo i Fox [\[9\]](#) koristili su MRF(Markov random field) i hijerarhijsko segmentacijsko stablo kako bi iskoristili kontekstualno modeliranje.

Konvolucijske neuronske mreže pružaju najbolje rezultate u skoro svim primjenama računalnog vida. Začetnik ideje konvolucijske neuronske mreže bio je LeCun, koji ju je primijenio na rukom pisane brojke 1998. godine na MNIST ispitnom podatkovnom skupu.

Konvolucijske neuronske mreže nisu našle širu upotrebu do 2012. godine, a glavni razlog tomu su bila tehnološka ograničenja, tj. prevelika opterećenja na računalnu snagu. Razvojem računala veće snage i omogućavanjem pristupa velikim ispitnim podatkovnim skupovima postigli su se značajni uspjesi u računalnom vidu. Model konvolucijske neuronske mreže AlexNet [\[12\]](#) 2012. godine je pokrenuo novi val upotrebe neuronskih mreža u računalnom vidu kada je smanjio grešku klasifikacije s 26% na 15%. Do danas se taj postotak smanjio na nevjerovatnih 1.3%.

Razvojem konvolucijskih neuronskih mreža iste su postale najuspješniji pristup problemu semantičke segmentacije. Jedan od problema pri korištenju RGB-D slika je održati vezu između RGB i dubinskog dijela slika te su se pojavili razni pristupi tom problemu. Razlikujemo ranu, srednju i kasnu fuziju. Razlika između pojedinih pristupa je u kojem stadiju obrade slike se spajaju RGB i dubinski dio slike, tj. spajaju li se prije kodera, između kodera i dekodera ili nakon dekodera.

Koder i dekođer su funkcionalni dijelovi u mreži koji uglavnom dolaze u paru. Koder iz ulaza, kao što je slika, kroz niz slojeva izlučuje značajke u vektor značajki koji se zatim prosljeđuje u dekođer koji iz tog vektora značajki daje najbližu pretpostavku prvotnog ulaza. Primjena koderi i dekođera je široka u dubokom učenju te su neki od primjera prevođenje s jednog jezika na drugi te u raznim generativnim modelima (slika, tekst, oblak točaka i drugi).

Iako se 2D modeli također mogu koristiti za semantičku segmentaciju unutrašnjih prostora, u nastavku će biti opisani neki od 3D modela za semantičku segmentaciju.

PointNet [\[10\]](#) iz 2017. godine jedna je od najpoznatijih neuronskih mreža području 3D semantičke segmentacije. PointNet kao ulaz prima oblak točaka koji za razliku od konvencionalnog načina zapisivanja 3D snimaka u mreže (engl. mesh) znatno rasterećuje opterećenje na računalnu snagu. Mreže se za razliku od oblaka točaka sastoje od mnogokuta i zbog toga zauzimaju više prostora te je potrebno više računalne snage za njihovu obradu. Budući da je 3D znatno zahtjevniji od 2D za računalne operacije oblak točaka se uz „voxel grid“ pokazao kao najbolje rješenje za probleme u 3D-u. „Voxel grid“ se za razliku od neorganiziranog oblaka točaka sastoji od jednakostraničnih kvadara složenih u trodimenzionalnu matricu koji imaju preddefinirane dimenzije. PointNet je korišten kao osnova za rad GSPN-a koji je objašnjen u sljedećem poglavlju.

SGPN (Similarity Group Proposal Network) [\[11\]](#) iz 2017. korišten je za usporedbu rezultata GSPN-a na ScanNet ispitnom podatkovnom skupu te on isto kao ulaz prima oblik točaka. Glavna ideja iza ovog rada je grupiranje točaka koje se izvodi pomoću matrice sličnosti za svaku točku. Ovakav pristup donio je značajne rezultate pri razlikovanju svake instance objekata koje je teško razlikovati 2D pristupom.

GSPN [\[1\]](#) iz 2018. donosi novi doprinos u području 3D računalnog vida time što na osnovu vidljivog dijela objekta vrši rekonstrukciju ostatka objekta te time dodatno pospješuje iskoristivost 3D pristupa u semantičkoj segmentaciji. Taj model je detaljno objašnjen u sljedećem poglavlju.

### 3. ARHITEKTURA MREŽE

U ovom poglavlju opisana je neuronska mreža R-PointNet, čiji je glavni dio generativna neuronska mreža GSPN te ću ih detaljnije opisati u nastavku.

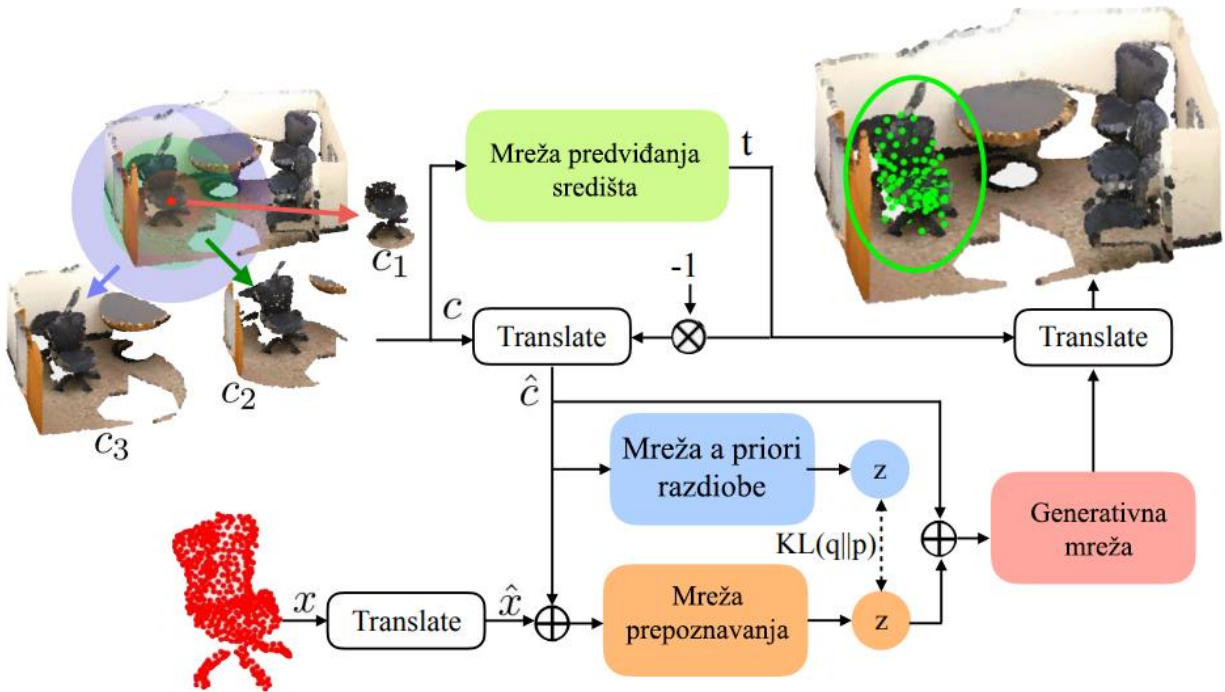
Ulaz u neuronsku mrežu R-PointNet je oblak točaka scene dobiven pomoću 3D senzora kao što su 3D LiDAR ili RGB-D kamera. U slučaju ScanNet ispitnog podatkovnog skupa potrebno je procesirati snimljene scene da bi se dobio oblak točak, taj postupak opisan je u poglavlju 6.2.

#### 3.1. GSPN

GSPN ili „Generative Shape Proposal Network“ metoda iz ulaznog oblika točaka generira pretpostavke o objektima prisutnim na sceni. Spomenute pretpostavke nazivaju se prijedlozi objekata (engl. object proposal). Prijedlozi objekata generiraju se u obliku oblaka točaka; one točke scene koje se nalaze u neposrednoj blizini generiranog prijedloga objekta pridružuje tom objektu. Generiranje prijedloga objekata tako da se ulazni oblak točaka uzorkuje metodom najdalje točke (engl. Farthest point sampling)[4]. Ono funkcionira tako da opetovano postavlja točku uzorkovanja u sredinu najmanje poznatog područja unutar domene uzorkovanja. Točke dobivene ovim postupkom predstavljaju ključne točke za generiranje prijedloga objekta na sceni. Analizira se okolina (susjedstvo) svake izabrane ključne točke u svrhu generiranja hipoteze o postojanju određenog objekta u njenoj okolini. Opisani postupak ključan je za segmentiranje instanci objekata.

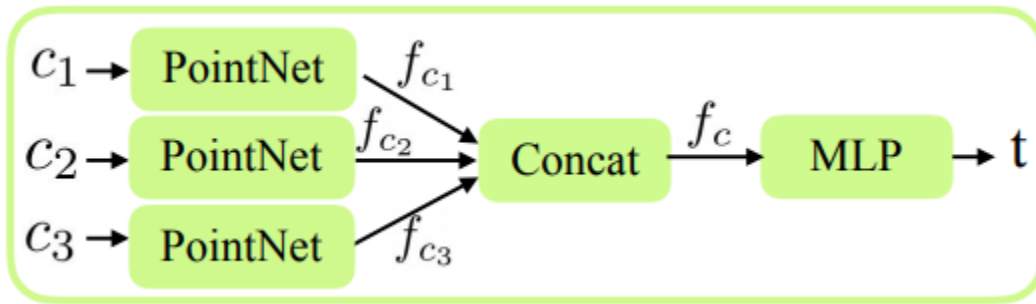
GSPN je zapravo uvjetni varijabilni autoenkoder (CVAE) koji generiranjem „nevidljivog“ dijela objekta pospješuje segmentaciju na fizičke objekte. Arhitektura GSPN-a se sastoji od 2 funkcijska dijela. Prvi funkcijski dio predstavlja „Mreža Predviđanja Središta“, koji određuje središte objekta, dok se drugi dio mreže sastoji od „Mreže a priori razdiobe“, „Mreže Prepoznavanja“ te „Generativne Mreže“ koje određuju oblik objekta. U nastavku je dan opis svake od navedenih pod-mreža. Na slici 3.1. nalazi se shematski prikaz GSPN mreže.





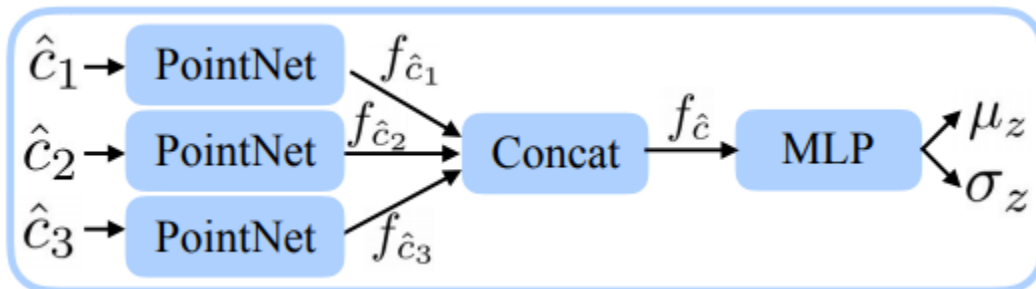
Slika 3.1. Grafički prikaz arhitekture GSPN mreže

„Mreža predviđanja središta“ kao ulaz prima tri oblaka točkaka, od kojih svaki predstavlja susjedstvo jedne ključne točke ulaznog oblaka točkaka unutar zadanog radijusa. Tri različita radijusa definiraju tri susjedstva ( $c_1, c_2, c_3$ ), tj. tri oblaka točkaka koje autori nazivaju kontekst. Razmatrana mreža kao rezultat vraća pretpostavljenu točku središta  $t$  traženog objekta  $x$ . Za svaki polumjer  $K$ , tj. za svaki kontekst  $c_k$ , postoji zaseban PointNet koji obrađuje sukladni kontekst te vraća zasebne vektore značajki (engl. feature vector)  $f_{ck}$  koje zatim spaja u jedinstveni vektor značajki  $f_c$  koji se prosljeđuje višeslojnom perceptronu (engl. multilayer perceptron - MLP) te kao rezultat daje mjesto središta  $t$ . Kontekstu  $c$  se središte postavlja u dobivenoj točki  $t$ , tj. u pretpostavljeno središte objekta, te se predaje mreži za određivanje oblika objekta kao  $c'$ . Na slici 3.2. nalazi se shematski prikaz opisane mreže.



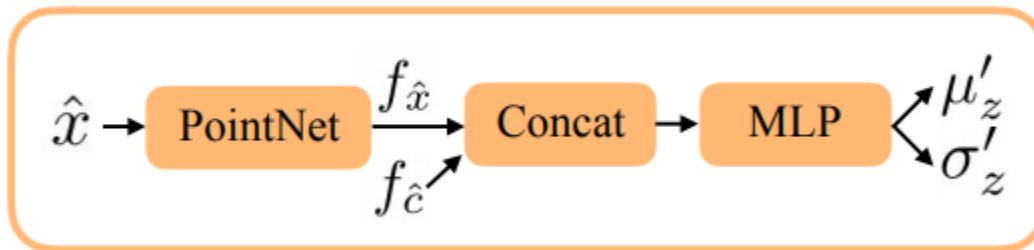
Slika 3.2. Grafički prikaz arhitekture Mreže predviđanja središta

„Mreža a priori razdiobe“ kao ulaz prima centrirani kontekst  $c'$  te se sastoji od iste arhitekture kao i „Mreža predviđanja središta“, ali za razliku od „Mreže predviđanja središta“ koristi višeslojni perceptron nad spojenim vektorom značajki  $f_c'$  za izračun Gaussove a priori razdiobe (engl. Gaussian prior distribution). Na slici 3.3. nalazi se shematski prikaz opisane mreže.



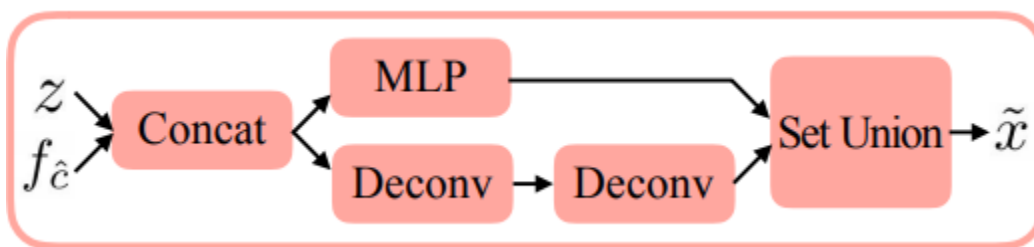
Slika 3.3. Grafički prikaz arhitekture Mreže a priori razdiobe

„Mreža Prepoznavanja“ dijeli kontekstni koder (engl. encoder) s „Mrežom a priori razdiobe“. Kao ulaz prima centrirani objekt  $x'$  te generira vektor značajki objekta  $f_{x'}$  pomoću PointNet-a. Vektor značajki objekta se zatim spaja s vektorom značajki konteksta  $f_c$ , i provlači kroz višeslojni perceptron za izračun Gaussove razdiobe prijedloga (engl. Gaussian proposal distribution). Na slici 3.4. nalazi se shematski prikaz opisane mreže.



Slika 3.4. Grafički prikaz arhitekture Mreže prepoznavanja

Iz rezultata „Mreže a priori razdiobe“ i „Mreže predviđanja središta“ uzorkuje se latentni vektor  $z$ . Latentni vektor je zapis 3D oblika fizičkog objekta na temelju kojeg se može rekonstruirati cjeloviti oblik 3D objekta. Latentni vektor  $z$  se prosljeđuje u „Generativnu mrežu“. „Generativna mreža“ također dijeli koder konteksta (engl. context encoder) s „Mrežom a priori razdiobe“. Nakon spajanja vektora značajki konteksta  $f_c$  i latentne pretpostavke fizičkog objekta  $z$ , dekodira oblak točaka  $x'$  te mjeru uvjerenosti (engl. confidence score) za svaku točku oblaka točaka koja predstavlja vjerojatnost pojave svake generirane točke. Na slici 3.5. nalazi se shematski prikaz opisane mreže.



Slika 3.5. Grafički prikaz arhitekture Generativne mreže

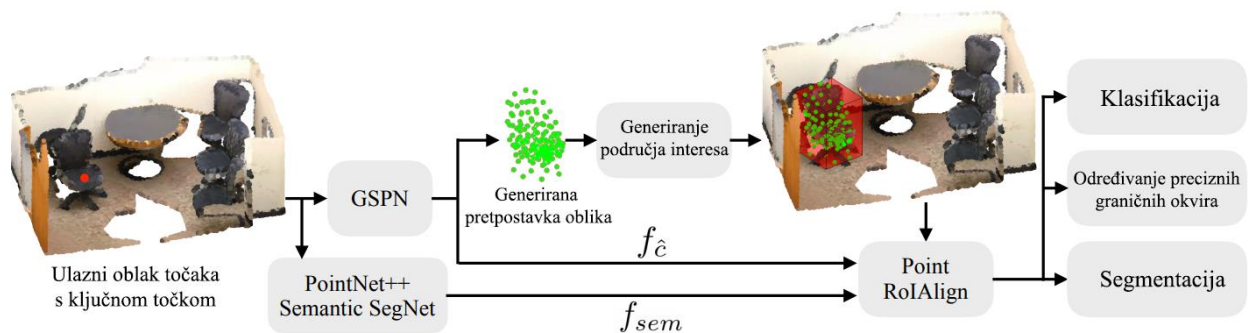
GSPN sadrži dodatni višeslojni perceptron koji uzimanjem vektora značajki konteksta  $f_c$ , određuje mjeru vjerojatnosti da razmatrani kontekst sadrži objekt od interesa, koji se u literaturi naziva *engl. objectness score*. Mreža uz *objectness score* daje i pretpostavku graničnog okvira u obliku kvadra

čiji su bridovi paralelni s osima koordinatnog sustava ulaznog oblaka točaka. Iste uči generirati na osnovu referentnih podataka koji sadrže granične okvire stvarnih predmeta prisutnih na sceni. Točnom pretpostavkom se smatra pretpostavka čiji se pripadni granični okvir podudara s referentnim graničnim okvirom u omjeru volumena presjeka i unije većim od 50%, a netočne pretpostavke su s omjerom manjim od 50%. Koeficijent podudaranja pripadnih i referentnih graničnih okvira oblaka točaka se označava kao *IoU* (engl. Intersection over Union).

## 3.2. R-PointNet

R-PointNet ili Region-based PointNet je neuronska mreža namijenjena semantičkoj segmentaciji, detekciji fizičkih objekata te klasifikaciji objekata (engl. object classification) u oblaku točaka snimljenom 3D senzorom.

U drugom dijelu R-PointNet-a se rezultati GSPN-a, tj. prijedlozi objekata, dalje obrađuju. Određuju se klase objekata, predloži se preciznije određuju te se odvaja pozadina od objekata. Iz GSPN-a se dobivaju područja interesa (engl. Region of Interest - RoI), tj. za točke generiranog prijedloga objekta koje imaju, već spomenutu, mjeru uvjerenosti veću od 0.5 određuje se granični okvir. Iz svakog se područja interesa pomoću „Point RoIAlign“ sloja, koji će biti objašnjen u sljedećem odlomku, izvode značajke područja, koje se prosljeđuju neuronskim pod-mrežama za klasifikaciju, određivanje preciznih graničnih okvira i segmentaciju. Precizni granični okviri se određuju predviđanjem promjena relativnog središta i dimenzija graničnog okvira u odnosu na prvotno procijenjene vrijednosti. Segmentacija se postiže predviđanjem binarne maske za svaku točku i to za svaku kategoriju. Na slici 3.6. nalazi se shematski prikaz arhitekture R-PointNet mreže.



Slika 3.6. Grafički prikaz arhitekture R-PointNet mreže

U ovom odlomku opisan je proces pridruživanja značajki svakoj točki ulaznog oblaka točaka. R-PointNet koristi PointNet++ neuronsku mrežu koja je prethodno naučena da klasificira svaku točku u klasu objekta, a sastoji se od 4 „sample-and-group“ sloja te 4 sloja interpolacije značajki (engl. feature interpolation). Vektori značajki se prikupljaju nakon svakog „sample-and-group“ sloja, za svaku točku, te se spajaju u jedan vektor značajki. Taj proces svakoj točki pridružuje jedan vektor značajki  $f_{sem}$  koji opisuje okolinu točke na različitim skalama. Budući da se nakon svakog „sample-and-group“ sloja smanjuje broj točaka kojima su pridruženi vektori značajki, značajke se

nakon svakog „sample-and-group“ sloja distribuiraju susjednim interpolacijskim slojevima.  $f_{sem}$  vektor značajki spaja se sa značajkom konteksta  $f_{c'}$  u „Point RoIAlign“ sloju opisanom u sljedećem odlomku.

„Point RoIAlign“ sloj koristi se za izračunavanje skupa značajki fiksne veličine za svako područje interesa. Za taj proces sloj uzorkuje fiksni broj točaka  $N_{ROI}$  koje sadrže vektor značajki iz ulaznog oblaka točaka. „Point RoIAlign“ prima vektore značajki  $f_{sem}$  i  $f_{c'}$ , te budući da je izračun značajki konteksta  $f_{c'}$  iznimno računski zahtjevan, „Point RoIAlign“ izračunava značajke na podskupovima oblaka točaka, a zatim se te značajke interpolacijom pridružuju uzorkovanim točkama.

R-PointNet se uči tako da minimizira složenu funkciju gubitka (engl. loss function) definiranu za svako područje interesa. Funkcija gubitka se računa zbrajanjem tri člana; gubitak klasifikacije  $L_{cls}$ , gubitak graničnog okvira  $L_{box}$  i gubitak maske  $L_{mask}$ .

Kao što je objašnjeno u poglavlju 3.1, GSPN generira granične okvire, tj. područja interesa za svaki prijedlog objekta, i tokom učenja svaki se od tih graničnih okvira uspoređuje s nekim od referentnih graničnih okvira sadržanih u podatkovnom skupu za učenje. Ona područja interesa čiji se granični okviri podudaraju s nekim od referentnih graničnih okvira tako da im je  $IoU > 0.5$ , proglašavaju se točnim prijedlogom objekta. Dakle, svakom od točnih prijedloga objekta pridružena je njegova referentna klasa i njegov referentni granični okvir. Također je svakoj točki ulaznog oblaka točaka pridružena referentna klasa. Gubitak klasifikacije  $L_{cls}$  je logistički gubitak koji se računa za referentnu klasu  $u$  na sljedeći način.

$$L_{cls} = -\log p_u,$$

gdje je  $p_u$  izlaz iz pod-mreže za klasifikaciju.

Gubitak graničnog okvira  $L_{box}$  je „glatka“ L1-norma razlike graničnih okvira  $v_i$  i izlaza pod-mreže za određivanje preciznih graničnih okvira  $t_i$ .

$$L_{box} = \sum_i \text{smooth}_L(t_i^u - v_i)$$

Svakoj točki unutar područja interesa pod-mreža za segmentaciju pridružuje za svaku klasu vrijednost koja predstavlja pripadnost točke objektu te klase. Veća vrijednost odgovara većoj vjerojatnosti da ta točka pripada razmatranoj klasi. Razmatra se referentna klasa pridružena tom

području interesa. Pošto je podatkovni skup za učenje sadrži podatke o referentnoj klasi za svaku točku ulaznog oblaka točaka, točke područja interesa kojima je pridružena referentna klasa ista kao i referentna klasa pridružena samom području interesa dobivaju vrijednost 1, dok ostale dobivaju vrijednost 0. Tako se dobiva referentna maska za područje interesa. Za svaku se točku područja interesa pripadnost referentnoj klasi pridružena od strane pod-mreže za segmentaciju uspoređuje s vrijednošću referentne maske za tu točku prema mjeri binarne unakrsne entropije. Srednja vrijednost te mjere za sve točke područja interesa predstavlja  $L_{mask}$ .

$$L = L_{cls} + L_{box} + L_{mask}$$

### 3.3. Učenje mreže

Kao što je i uobičajeno pri učenju neuronskih mreža, iz skupa za učenje se izdvajaju podskupovi (engl. mini-batch) te se svaki korak učenja izvodi na jednom od tih podskupova. Pri treniranju GSPN-a nasumično se odabire 512 ključnih točaka (engl. seed points) za svaku scenu u svakom podskupu što daje 512 pretpostavki oblika. Rezolucija svake pretpostavke oblika je 512 točaka. Konteksti  $c_k$ , opisani u poglavlju 3.1., se dobivaju izdvajanjem podskupova ulaznog oblaka točaka u okolini ključnih točaka u obliku kugle. Radijusi korišteni pri učenju iznose 0.5, 1.0 i 1.5 te ugrubo odgovaraju veličini najmanjeg i najvećeg objekta u sceni, tj. 0.5 najmanjem, a 1.5 najvećem. Središte objekta za pojedinu ključnu točku dobije se regresijom jediničnog vektora iz ključnih točaka u središte objekta. Također se koristi i „KL-annealing“ radi stabilizacije GSPN učenja mreže.

Pri učenju R-PointNet-a se koristi uklanjanje ne-maksimuma na svakom predlaganju objekta te se održava maksimalni broj od 128 predlaganja pri treniranju. Uklanjanje ne-maksimuma je postupak koji filtrira prijedloge objekata tako da ostane samo jedan prijedlog, koji ima najveću mjeru uvjerenosti.

### **3.4. Inferencija**

Tokom inferencije, nasumično se odabire 2048 ključnih primjera točaka za svaku ispitnu scenu i radi se na maksimalno 512 područja interesa nakon uklanjanja ne-maksimuma. Nakon određivanja binarne segmentacije za svako područje interesa, segmentacijska maska se projicira u početni oblak točaka korištenjem metode najbližeg susjeda. Točke oblaka točaka izvan područja interesa se ne projiciraju.



## 4. ISPITNI PODATKOVNI SKUPOVI

Budući da je GSPN mreža po strukturi i načinu rada poprilično općenita te primjenjiva na više problema segmentacije na instance objekata, tim koji je kreirao GSPN ispitivao ga je na ispitnim podatkovnim skupovima ScanNet-u, PartNet-u te NYUv2. U ovom poglavlju su opisani rezultati dobiveni od strane autora metode za svaki pojedinačni problem, a u sljedećem će biti prikazani rezultati dobiveni na podatkovnom skupu ScanNet u okviru ovog rada.

### 4.1. ScanNet

Ispitni podatkovni skup ScanNet je projekt proizašao iz suradnje Google Tanga, Intela, NSF-a te Stanforda. ScanNet je veliki podatkovni skup koji se sastoji od 1613 3D modela unutrašnjih prostorija. Svaka rekonstrukcija prostorije nastaje spajanjem više snimaka iz različitih kutova gledanja RGB-D kamerom. Scene su označene instancama objekata i semantičkim segmentacijskim maskama. Podatkovni skup je podijeljen na 1201 scena za učenje, 312 scena za potvrdu (engl. validation) te 100 ispitnih scena. Za ispitne scene ne postoje referentni podatci.

Evaluacija metode GSPN na ScanNet-u se vrši mjerenjem uspješnosti segmentacije na instance usporedbom s referentnim podacima koji su službeno priloženi uz podatkovni skup ScanNet. Evaluacija se vrši i uspoređuje na 18 klasa objekata. Mreži se predaju oblaci točaka u boji uz semantičke oznake. Tih 18 prisutnih klasa objekata sadrži različite vrste objekata uobičajenih za unutrašnje prostore; od objekata manjih dimenzija kao što su slike do objekata većih dimenzija kao što su stolovi i hladnjaci, što znatno otežava njihovu segmentaciju i klasifikaciju. U svrhu evaluacije kao mjerilo preciznosti se koristi prosječna preciznost (engl. average precision - AP) s IoU (engl. Intersection over Union = presjek unije) pragom od 0.5. Za razliku od zadataka prepoznavanja, IoU se, u ovom slučaju, računa na segmentacijama, a ne na graničnim okvirima što naglašava važnost detalja.

Rezultati evaluacije se uspoređuju s 2 druga rada; SGPN i PMRCNN. GSPN daje znatno bolje rezultate za sve klase objekata osim za zidne slike. U usporedbi s SGPN-om se jasno vidi prednost GSPN metode koja ne samo da uspješno koristi informaciju o 3D geometriji, već i informaciju o boji, što je posebno važno za prepoznavanje objekata kao što su vrata i prozori za čije se prepoznavanje mora koristiti informacija o teksturi. PMRCNN koristi 2D pristup za razliku od ostala 2 te ne iskorištava potencijal 3D-a, ali daje bolje rezultate za zidne slike koje ne ovise o

geometriji. 2D pristupi i dalje imaju bolje mogućnosti učenja na površinskom izgledu dok su 3D pristupi korisniji na oblicima. Usporedba spomenutih mreža nalazi se na slici 4.1.

	Mean	cabi- net	bed	chair	sofa	table	door	win- dow	book- shelf	pic- ture	coun- ter	desk	cur- tain	fri- dge	shower curtain	toilet	sink	bath- tub	other
PMRCNN	5.3	4.7	0.2	0.2	10.7	2.0	3.1	0.4	0.0	<b>18.4</b>	0.1	0.0	2.0	6.5	0.0	10.9	1.4	33.3	2.1
SGPN	13.3	6.0	36.1	25.7	33.5	16.1	7.9	12.2	14.9	1.3	2.6	0.0	6.2	2.6	0.0	16.1	10.4	19.4	3.8
Ours	<b>30.6</b>	<b>34.8</b>	<b>40.5</b>	<b>58.9</b>	<b>39.6</b>	<b>27.5</b>	<b>28.3</b>	<b>24.5</b>	<b>31.1</b>	2.8	<b>5.4</b>	<b>12.6</b>	<b>6.8</b>	<b>21.9</b>	<b>21.4</b>	<b>82.1</b>	<b>33.1</b>	<b>50.0</b>	<b>29.0</b>

*Slika 4.1. Rezultati službene evaluacije ScanNet podatkovnog skupa od strane autora GSPN-a, izraženi prosječnom preciznošću pri IoU pragu od 0.5*

## 4.2. PartNet

PartNet ispitni skup se sastoji od pokućstva koje je podijeljeno na konstrukcijske dijelove, npr. stolica se sastoji od naslonjača, držača naslonjača, plohe za sjedenje, nogu, prečke za nogu, itd. Za razliku od prethodnog, ScanNet ispitnog skupa, PartNet se sastoji od zrnastijih (detaljnijih) objekata na kojima se bolje vidi primjena segmentacije na instance objekata. Za evaluaciju su korištene 4 najveće klase objekata; stol, stolica, lampa i bife. Kao i u prethodnom ispitnom skupu korišten je IoU prag od 0.5 te su rezultati evaluacije uspoređeni s SGPN metodom. U usporedbi s SGPN metodom rezultati evaluacije su znatno na strani GSPN metode koja daje manje „zašumljene“ segmente.

## 4.3. NYUv2

NYUv2, jedan od najrasprostranjenijih ispitnih skupova za segmentaciju, se za razliku od ScanNet podatkovnog skupa sastoji od RGB-D snimaka snimanih iz samo jednog kuta te nedostaje puno veći dio podataka u oblaku točaka. IoU prag korišten na ovom podatkovnom skupu je 0.25. U ovom slučaju GSPN je uspoređen s pristupima SGPN-CNN, „Mask R-CNN“ [6] [7] te „Mask R-CNN“ uz neke preinake. Iako je GSPN metoda pružila bolje rezultate za 12 od 19 klasa, opet je znatna razlika u klasama koje su geometrijski izraženije, kao što su kade i toaleti, od klasa koje ovise o površinskom izgledu, kao što su monitori te „Mask R-CNN“ pruža bolje rezultate za takve primjere. Za usporedbu s „Mask R-CNN“ mrežom koja puno bolje obrađuje boje od GSPN metode,

ali nema značajnije koristi od 3D ulaza, dolazimo do zaključka da postoji potreba za razvijanjem 3D pristupa.

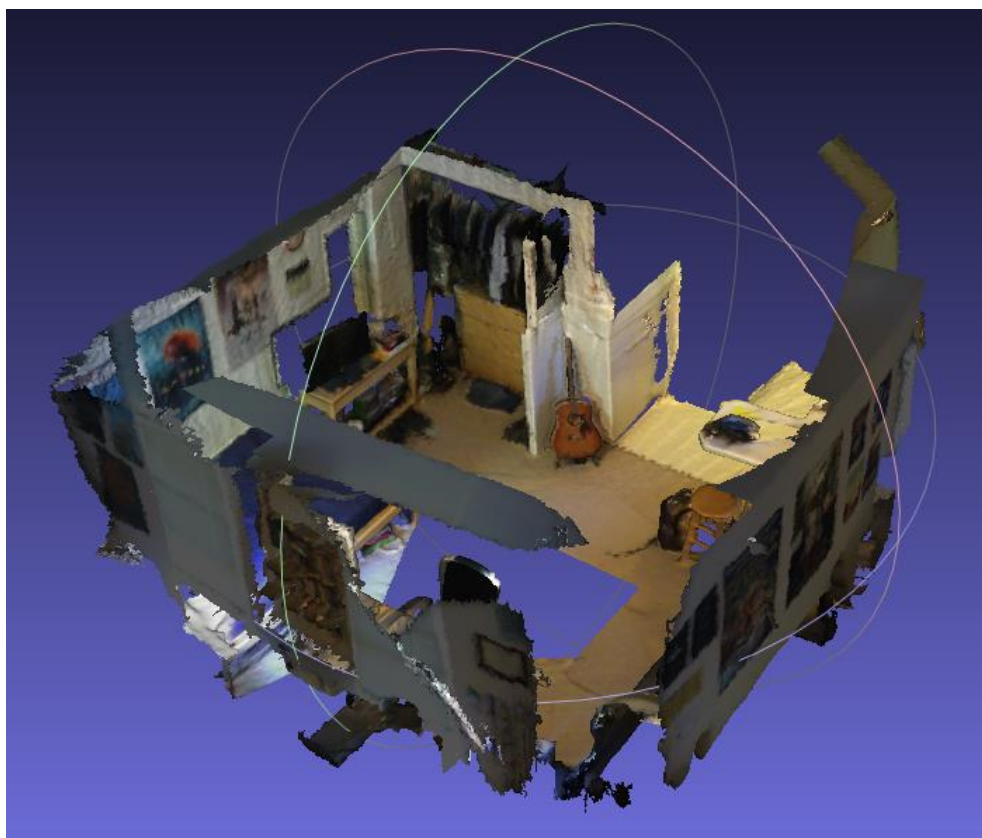
## 5. EKSPERIMENTALNA EVALUACIJA

### 5.1. Sklopovlje

U ovom potpoglavlju opisano je sklopovlje korišteno za eksperimentalnu evaluaciju neuronske mreže R-PointNet provedenu u okviru ovog rada. Pokusi su provedeni korištenjem osobnog računala koje se sastoji od procesora Intel i7-4700k, grafičke kartice Nvidia Gtx 1660Ti, 8GB DDR3 RAM-a te 200 GB tvrdog diska alociranog za „dual-boot“ Linux verzije Ubuntu. Korišteni su 18.04 verzija Ubuntu Linuxa, Tensorflow 1.4 na Python-u verzije 2.7 te CUDA verzija 8.0. Ograničenje ove konfiguracije je bio nedostatak GDDR RAM-a, tj. radne memorije grafičke kartice. Grafička kartica Nvidia GTX 1660Ti sadrži 6 GB GDDR6 radne memorije, ali to nije bilo dovoljno za odraditi učenje zbog zahtjevnosti 3D podataka koji su obrađivani, iako je prednost oblaka točaka efikasnija obrada 3D podataka. Za normalno izvođenje bilo je potrebno minimalno 12 GB GDDR radne memorije. Problem nedostatka radne memorije riješen je uređivanjem postavki pri učenju neuronske mreže, tj. umjesto prvotnog „Batch size“-a vrijednosti 2, promijenjen je na 1 te na račun preciznosti dobilo na „preformansama“. U originalnom radu [2] ne postoje informacije koliko je vremena ukupno bilo potrebno za učenje neuronske mreže na Titan XP grafičkoj kartici, ali je znatno manje vremena bilo potrebno nego u ovom pokusu, te su stoga u potpoglavlju 5.2. opisane postavke samog modela korištene pri učenju neuronske mreže.

## 5.2. Učenje neuronske mreže

Budući da je kompletni ScanNet [51] podatkovni skup, uz sve formate skeniranja, veličine 1.2TB za ovaj pokus preuzete su samo nužne datoteke. Među tim datotekama se nalaze mreže (engl. mesh) .ply formata koji se kasnije preobrađuju u oblake točaka te oznake (engl. labels) u JSON formatu, koje su se također kasnije obrađivale. Na slici 5.1. prikazana je unutrašnja prostorija iz podatkovnog skupa ScanNet u izvornom obliku.

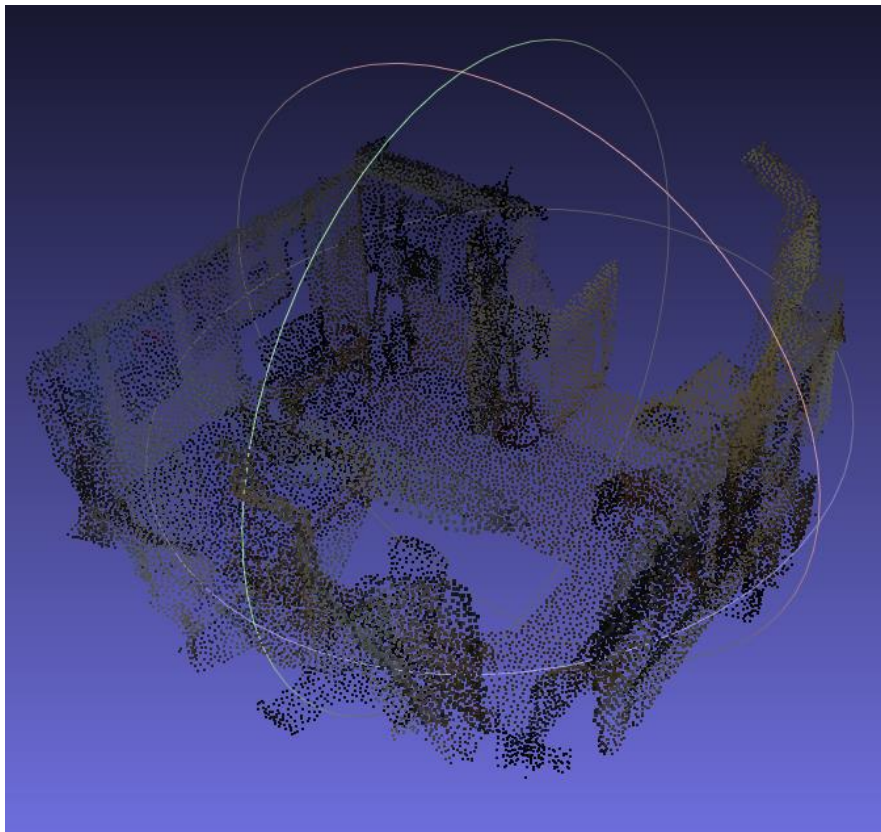


*Slika 5.1. Prikaz scene u izvornom, „mesh“, obliku*

Za provođenje svih daljnjih koraka pri učenju i evaluaciji mreže potrebno je prvo kompilirati Tensorflow operatore koji su napisani u C++ radi bržeg izvođenja. Četiri operatora korištena za ovaj model su 3D interpolacija, grupiranje, metoda najbližeg susjeda te uzorkovanje.

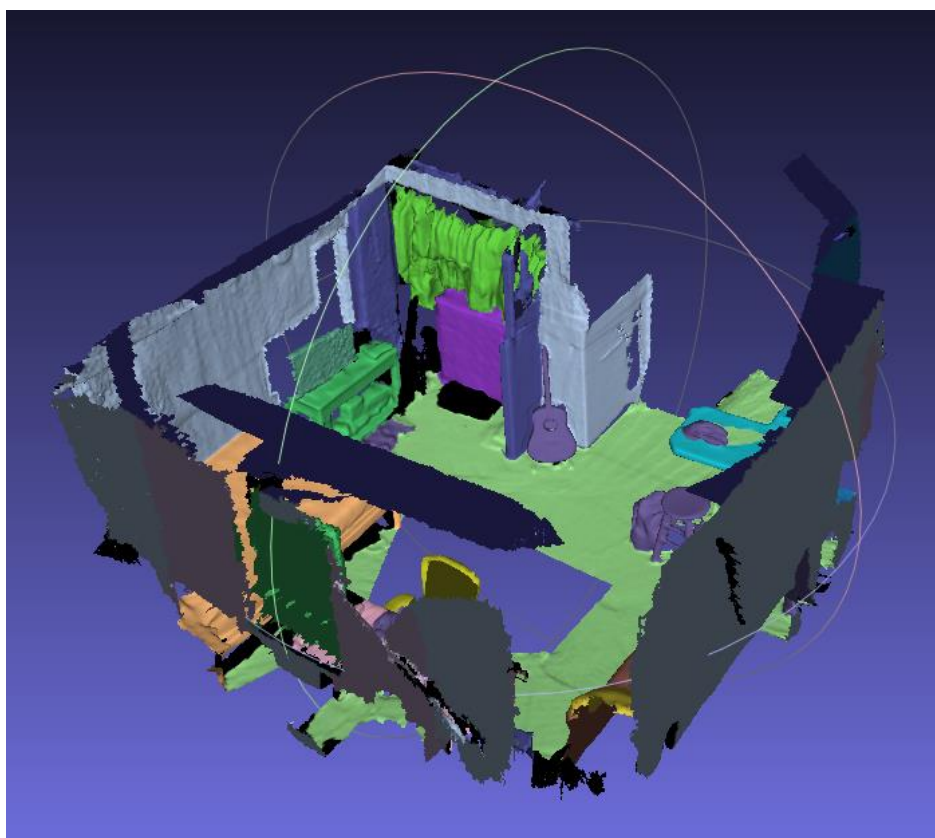
Sama predobrada podataka se odvija u dva koraka. Prvi korak predobrade je „downsample“ scena iz mreže u oblake točaka te prilagođavanje oznaka oblaku točaka, a zatim njihovo zapisivanje u tekstualne .txt datoteke. Drugi korak predobrade je ponovno uzorkovanje svake točke scene oblaka

točaka na konačan broj te ponovno uzorkovanje svake instance objekta na konačan broj točaka radi jednostavnijeg provođenja učenja i evaluacije neuronske mreže. Konačan broj točaka kojima su prikazane scene je 18000, a konačan broj točaka kojim su prikazane instance objekata je 512. U ovom koraku se izvodi i augmentacija na skupu podataka za učenje dok se augmentacija ne provodi na skupu podataka za potvrđivanje (engl. validation). Na slici 5.2. prikazan je rezultat predobrade snimke sa slike 5.1.



*Slika 5.2. Prikaz scene u, predobrađenom, oblaku točaka*

Budući da se cijeli podatkovni skup sastoji od oko 1500 snimaka 700 scena, podatkovni skup se dijeli na dva skupa, skupa za učenje te skup za evaluaciju. Skup za učenje se sastoji od 1201 snimaka, dok se skup za evaluaciju sastoji od 311 snimaka. Snimke su nasumično podijeljene između ta dva skupa, vrijednosti pripadnosti snimke skupu nalaze se u tekstualnim dokumentima „scannet\_train.txt“ i „scannet\_val.txt“. Na slici 5.3. nalazi se primjer scene segmentirane bojama ovisno o pripadnosti određenoj klasi.



*Slika 5.3. Prikaz scene označene različitim bojama ovisno o klasi objekta*

Proces učenja neuronske mreže se sastoji od dva dijela u kojima se odvojeno uče SPN i R-PointNet. S obzirom da se parametri i postavke u ovom pokusu razlikuju od originalnih, opisani su u sljedećem odlomku.

Glavna datoteka koja sadrži većinu općih postavki je „config.py“. Ostale datoteke uglavnom koriste zadane postavke iz te „config.py“ datoteke. Neki od bitnijih parametara definiranih u

glavnoj datoteci postavki su broj kategorija objekata, broj točaka scene, broj točaka instance, „batch size“ te razne vrijednosti u definiranju područja interesa.

Postavke direktno vezane uz treniranje su definirane unutar „train.py“ datoteke, ali se pri svakom pokretanju učenja neuronske mreže mogu predati kao argumenti u terminalu. Takve postavke ili parametri se nazivaju hiper-parametri (engl. hyperparameters) te se oni definiraju prije učenja neuronske mreže dok se ostali parametri definiraju u procesu učenja, kao što su težine čvorova (engl. node weights). Neki od hiper-parametara su „batch size“, „momentum“ (moment), „optimizer“ (metoda optimiziranja), „learning rate“ (stopa učenja) te „decay step & rate“ (korak i stopa raspadanja).

Od bitnijih preddefiniranih hiper-parametara su stopa učenja koja iznosi 0.001, metoda optimiranja je „adam“, korak i stopa raspadanja iznose 55000 i 0.7.

Prilikom učenja neuronske mreže prvo se odvija učenje SPN dijela mreže te se pri pokretanju „train.py“ datoteke predaje argument „SPN“ kao modul treniranja. Budući da je ovaj pokus bio ograničen mogućnostima sklopovlja umjesto originalnih 800 epoha koje se izvode prilikom učenja, postavljeno je 170 epoha. Iako se 170 epoha čini znatno manjim od zadanog broja, rezultati su relevantni te nam daju mogućnost usporedbe s rezultatima dobivenim od strane autora.

Prilikom iteracije svake epohe, u terminalu se ispisuju vrijednosti koje nam opisuju uspjeh iteracije te se naučene vrijednosti spremaju svaku desetu iteraciju u datoteku „model.ckpt“. Te vrijednosti su: „mean loss“ (srednja vrijednost gubitka), „mean reconstruction loss“ (srednja vrijednost gubitka rekonstrukcije), „mean kl-divergence loss“ (srednja vrijednost gubitka Kullback–Leibler divergencije), „mean spn class loss“ (srednja vrijednost gubitka SPN klase). Nakon svake epohe vrši se i evaluacija na evaluacijskom dijelu podatkovnog skupa te ako su rezultati bolji od prethodno najbolje epohe, vrijednosti modela se spremaju u datoteku naziva „best\_model\_epoch\_XXX.ckpt“. Sve vrijednosti koje ispisuje terminal logiraju se u log datoteku te se one mogu koristiti u svrhu analize učenja neuronske mreže.

Nakon uspješno odrađenog učenja SPN dijela neuronske mreže, potrebno je predati SPN model s naučenim vrijednostima u proces učenja R-PointNet neuronske mreže.



Za razliku od učenja SPN dijela neuronske mreže kada se kao argument predavao samo modul treniranja, u tom slučaju „SPN“, sada je bilo potrebno predati određite datoteke prethodnog modela „model.ckpt“ te kao „restore\_scope“ predati ime prethodnog procesa kako bi preuzeo globalne varijable korištene pri učenju SPN dijela.

Proces učenja R-PointNet neuronske mreže se odvija istim principom kao i prethodni proces. Također se sastoji od 170 epoha umjesto originalno korištenih 800. Na ispisu vrijednosti koje opisuju uspjeh učenja SPN dijela, opisanih u prethodnim odlomcima, nadodaju se i vrijednosti vezane uz R-PointNet neuronsku mrežu, a to su „mean rpointnet class loss“ (srednja vrijednost gubitka RPointNet klase), „mean rpointnet bbox loss“ (srednja vrijednost gubitka graničnih okvira RPointNeta) te „mean rpointnet mask loss“ (srednja vrijednost gubitka RPointNet maski).

Za evaluaciju R-PointNet neuronske mreže koriste se vrijednosti gubitka vezani samo za R-PointNet te su one naznačene u prethodnom odlomku.

Nakon uspješno odrađenog treniranja R-PointNet neuronske mreže, vrši se evaluacija pomoću službene skripte ScanNet podatkovnog skupa za validaciju.

Za evaluaciju pomoću službene skripte ScanNet podatkovnog skupa potrebno je prvo generirati predloške na podskupu za evaluaciju.

Generiranje predložaka provodi se pokretanjem „test.py“ skripte uz predaju prethodno istrenirane neuronske mreže R-PointNet. U ovom procesu se preuzimaju referentne vrijednosti iz podatkovnog skupa te se vrši generiranje predložka segmentacije za scenu. Referentne vrijednosti za svaki snimak spremaju se u tekstualni dokument, unutar „gt“ direktorija, dok se predlošci spremaju unutar „pred“ direktorija u istom formatu.

Nakon generiranih datoteka za evaluaciju, potrebno je pokrenuti službenu ScanNet skriptu za evaluaciju segmentacije na semantičke instance. Kao argument predaju se direktoriji prethodno generiranih predložaka i referentnih vrijednosti.

### **5.3. Rezultati**

U sljedećem potpoglavlju opisani su rezultati uspješnosti GSPN metode dobivenih pokusom na osobnom računalu. S obzirom da su autori R-PointNet mreže stavili na raspolaganje već naučeni model, prikazani su i rezultati pokusa dobiveni tim modelom.

Rezultati dobiveni evaluacijom nad modelom izučenim sa samo 170 epoha i „batch“ veličinom 1 znatno se razlikuju od onog izučenog s 800 epoha i dvostruko većom „batch“ veličinom. Na prvi pogled se primjećuju rezultati za kategorije objekata koje imaju AP 0, neovisno o postotku IoU praga. Jedan od obrazloženja za takve rezultate je jednostavno premal uzorak pri treniranju u kojima su se te kategorije pojavljivale. U rezultatima s manjim brojem epoha također se može vidjeti razlika između kategorija objekata koje imaju izraženije geometrijske oblike te takve kategorije objekata pružaju bolje rezultate u odnosu na kategorije objekata koje bolje definira površinski izgled. Iako se smanjenjem praga na IoU na 25% u odnosu na 75% ili 50% AP-a povećava i do 4 puta, svejedno se ne mogu usporediti s rezultatima modela izučenog od strane autora GSPN metode.

U nastavku su na slikama 5.4. i 5.5. prikazani rezultati evaluacije ScanNet podatkovnog skupa na osobnom računalu. Rezultati dobiveni već naučenim modelom mogu se poistovjetiti s rezultatima već opisanim u potpoglavlju 4.1.

Vizualizacija rezultata na pojedinim scenama prikazane su na slikama 5.6.-5.9. Prikazane vizualizacije scena su dobivene od strane autora GSPN metode. Na vizualizacijama je pozadina prikazana ljubičasto-tamno plavom bojom, tj. sve što ne pripada nekoj od klasa iz podatkovnog skupa. Svi objekti su prikazani nasumično odabranim bojama te se različiti objekti koji su iste klase također razlikuju u boji kojom su prikazani.

Slike 5.8. i 5.9. prikazuju problem GSPN metode opisan u potpoglavlju 4.1., a to je precizno prepoznavanje slika. Na slici 5.8. u gornjem lijevom kutu označena su 3 objekta na zidu koja predstavljaju slike, dok na referentnoj vizualizaciji vidimo da je to samo zid.

```
#####
what      :          AP      AP_50%      AP_25%
#####
cabinet   :          0.004      0.021      0.188
bed       :          0.000      0.000      0.000
chair     :          0.072      0.230      0.508
sofa      :          0.000      0.000      0.000
table     :          0.010      0.031      0.163
door      :          0.056      0.183      0.419
window    :          0.010      0.040      0.235
bookshelf :          0.000      0.000      0.000
picture   :          0.037      0.073      0.144
counter   :          0.000      0.000      0.000
desk      :          0.001      0.013      0.137
curtain   :          0.000      0.000      0.000
refrigerator :        0.000      0.000      0.000
shower curtain :        0.082      0.217      0.397
toilet    :          0.088      0.290      0.478
sink      :          0.053      0.202      0.444
bathtub   :          0.014      0.095      0.412
otherfurniture :        0.008      0.027      0.070
-----
average   :          0.024      0.079      0.200
```

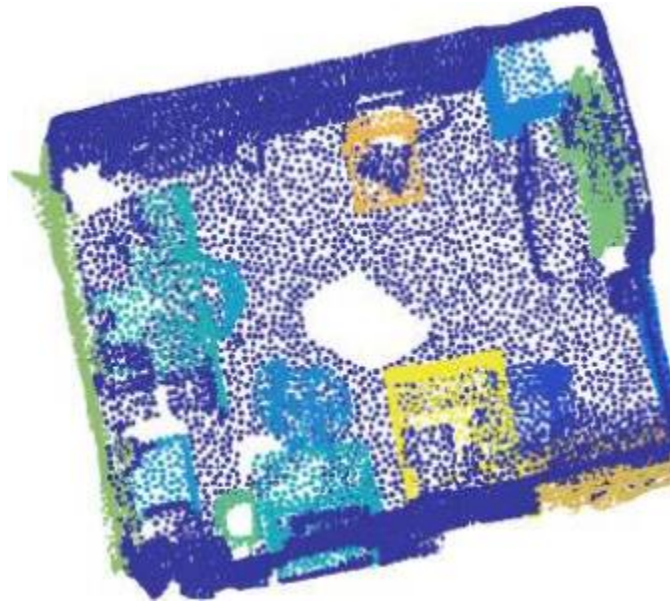
*Slika 5.4. Rezultati evaluacije ScanNet podatkovnog skupa dobivenih učenjem SPN i R-PointNet mreža na osobnom računalu*

```
#####
what      :          AP      AP_50%      AP_25%
#####
cabinet   :          0.106      0.281      0.479
bed       :          0.281      0.677      0.797
chair     :          0.510      0.751      0.819
sofa      :          0.267      0.620      0.771
table     :          0.257      0.492      0.639
door      :          0.191      0.380      0.650
window    :          0.102      0.274      0.538
bookshelf :          0.128      0.264      0.574
picture   :          0.153      0.346      0.511
counter   :          0.056      0.169      0.472
desk      :          0.066      0.229      0.646
curtain   :          0.062      0.113      0.362
refrigerator :        0.102      0.240      0.277
shower curtain :        0.186      0.363      0.560
toilet    :          0.595      0.813      0.870
sink      :          0.232      0.429      0.675
bathtub   :          0.378      0.687      0.756
otherfurniture :        0.116      0.245      0.334
-----
average   :          0.210      0.410      0.596
```

*Slika 5.5. Rezultati evaluacije ScanNet podatkovnog skupa dobivenih s već naučenim modelom R-PointNet-a od strane autora metode*

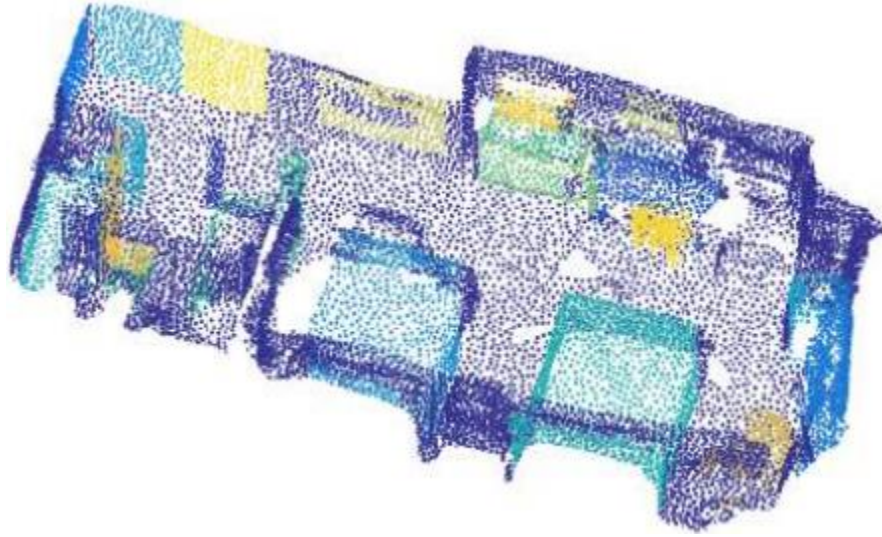


*Slika 5.6. Vizualizacija rezultata semantičke segmentacije na instance objekata scene GSPN metodom*

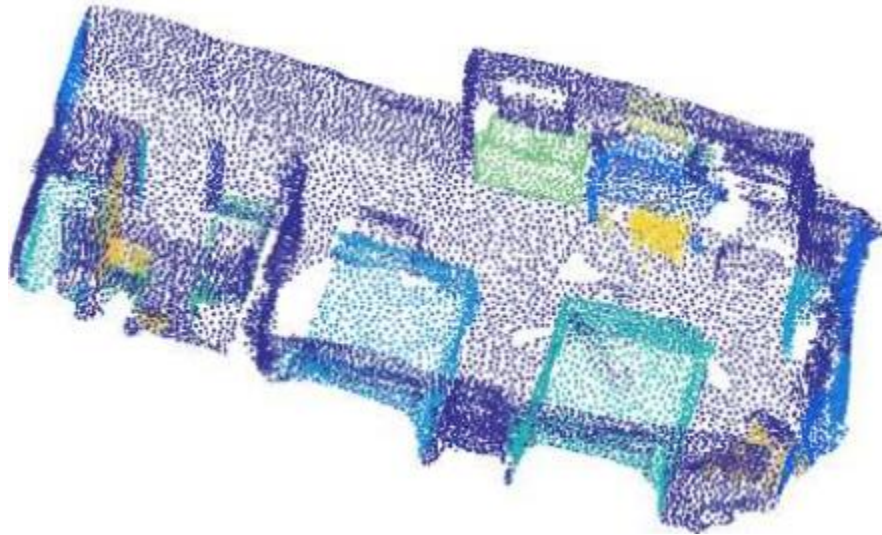


*Slika 5.7. Vizualizacija referentnih vrijednosti scene*





*Slika 5.8. Vizualizacija rezultata semantičke segmentacije na instance objekata scene GSPN metodom*



*Slika 5.9. Vizualizacija referentnih vrijednosti scene*

## 6. ZAKLJUČAK

U ovom radu prikazane su metode semantičke segmentacije 3D snimki unutrašnjih prostorija primjenom umjetnih neuronskih mreža. Prikazano je trenutno stanje u tom području te njegove mogućnosti i nedostaci. Kompletnom analizom jedne od metoda i njezinom provedbom na osobnom računalu dobivena je ideja razvoja i primjene takvih i sličnih metoda. Opisom metoda koje prate ovo područje umjetne inteligencije kao što su formati ulaza neuronske mreže, ispitni podatkovni skupovi, programska podrška, alati za prikupljanje podatka za obradu i sličnih dobivena je i pozadinska slika računalnog vida i semantičke segmentacije. Osobnom primjenom može se doći do zaključka o tehnološkim ograničenjima s kojima se susreće područje umjetne inteligencije. Konkretni primjer ograničenja raspoloživog sklopovlja doveo je do smanjenja efikasnosti procesa učenja neuronske mreže te samim time do znatno lošijih rezultata uspješnosti opisane, GSPN, metode. Iako su smanjenje broja mini-batcha i epoha doveli do lošijih rezultata u odnosu na rezultate dobivene od strane autora metode, jasno se mogu vidjeti pozitivne i negativne strane opisane metode. Procesom obrade 3D snimki u formatu oblaka točaka postigli su se odlični rezultati s objektima izraženih geometrijskih oblika, dok je osjetan nedostatak uspješnosti prepoznavanja geometrijski neizraženih objekata kao što su slike i zavjese. Saznanjima o mogućnostima ove i sličnih metoda dolazimo do ideja za primjenu semantičke segmentacije u unutrašnjim prostorijama. Mogućnost ove metode da jasno razaznaje i instance objekata istih i različitih klasa pronašla bi uporabu u robotima za kućanstvo. Već postojećim robotima za održavanje domova koristila bi informacija koju bi dobili o objektima u okruženju, npr. je li objekt moguće pomaknuti (stolica), proći kroz njega (vrata), spustiti se ili popeti uz objekt (stepenice, rampa) i sl. Nadalje, primjena ne mora nužno biti vezana uz pokušstvo već može pronaći primjenu u kuhinjama pri sortiranju i pranju suđa, kupaonicama pri sušenju i glačanju odjeće, spremanju objekata kao što su obuća i igračke u odgovarajuće ormariće, a sve to bi olakšala informacija o pripadnosti objekata određenoj klasi i razlikovanje instanci objekata.

## Literatura

- [1] – Li Yi, Wang Zhao, He Wang, Minhyuk Sung, Leonidas Guibas, GSPN: Generative Shape Proposal Network for 3D Instance Segmentation in Point Cloud, <https://arxiv.org/pdf/1812.03320.pdf>
- [2] – GSPN GitHub repozitorij, <https://github.com/ericyi/GSPN/>
- [3] – Fahimeh Fooladgar, Shohreh Kasaei, A survey on indoor RGB-D semantic segmentation: from hand-crafted features to deep convolutional neural networks, *Multimed Tools Appl*, br. 79, str. 4499–4524, 2020. <https://doi.org/10.1007/s11042-019-7684-3>
- [4] – Carsten Moenning, Neil A. Dodgson, Fast Marching farthest point sampling, <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-562.pdf>
- [5] – ScanNet GitHub repozitorij, <https://github.com/ScanNet/ScanNet/>
- [6] – Ross Girshick, Microsoft Research, Fast R-CNN, <https://arxiv.org/pdf/1504.08083.pdf>
- [7] – Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, Mask R-CNN, <https://arxiv.org/pdf/1703.06870.pdf>
- [8] - Silberman N, Fergus R (2011), Indoor scene segmentation using a structured light sensor, 2011 IEEE international conference on computer vision workshops (ICCV Workshops), IEEE, str. 601–608, 2011.
- [9] - Ren X, Bo L, Fox D (2012), Rgb-(d) scene labeling: features and algorithms, 2012 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, str. 2759–2766, 2012.
- [10] - Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas, PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, <https://arxiv.org/pdf/1612.00593.pdf>
- [11] - W. Wang, R. Yu, Q. Huang, U. Neumann, Sgpn: Similarity group proposal network for 3d point cloud instance segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, str. 2569–2578, 2018.

[12] – A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1, str. 1097-1105, 10.2012.



## **Sažetak**

U ovom radu istražena je i implementirana jedna od metoda za semantičku segmentaciju 3D snimaka unutrašnjih prostora umjetnom neuronskom mrežom. Opisane su neke od prethodno korištenih te „State of the Art“ metode za ovaj problem. Detaljno je proučena arhitektura R-PointNet neuronske mreže te svi procesi korišteni u Generative Shape Proposal Network (GSPN) mreži, koja predstavlja sastavni dio R-PointNet mreže. Navedeni su i ukratko opisani neki od najpoznatijih 3D ispitnih podatkovnih skupova unutrašnjih prostorija i njihova primjena za učenje i ispitivanje GSPN metode. Provedena je eksperimentalna evaluacija s učenjem neuronske mreže na osobnom računalu te analizirani rezultati uspješnosti. Istaknute su mogućnosti i nedostaci semantičke segmentacije u unutrašnjim prostorima te njihova eventualna primjena.

## **Abstract**

Title: Semantic segmentation of 3D in-door scenes using artificial neural network

This paper investigates and implements one of the methods for semantic segmentation of 3D in-door scenes using artificial neural network. Some State of the Art solutions for this problem are described. Architecture of the neural network R-PointNet and all of the processes used in Generative Shape Proposal Network (GSPN) network, representing the main part of the R-PointNet, are studied in detail. Some of the most popular datasets of 3D in-door scenes are listed and briefly explained as well as their use for training and testing of the GSPN method. An experimental evaluation was performed where the considered neural network was trained on a personal computer and the resulting efficiency scores were analyzed. Some advantages and disadvantages of the in-door semantic segmentation using the considered neural network are discussed and its possible real-world applications are suggested.

## **Životopis**

Nikola Stipić rođen je 09.07.1998. godine u Osijeku. U Gašincima i Satnici Đakovačkoj pohađao je OŠ Josipa Antuna Čolnića te Gimnaziju Antuna Gustava Matoša u Đakovu. 2017. godine upisuje Fakultet elektrotehnike, računarstva i informacijskih tehnologija u Osijeku, smjer računarstvo, gdje je trenutno student treće godine.