

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Denis Čaušević

**Primjena uvjetnih slučajnih polja s Gausovim
potencijalima za semantičku segmentaciju**

Zagreb, 2016.

Ovaj rad izrađen je na Zavodu za elektroniku, mikroelektroniku, računalne i inteligentne sustave na Fakultetu elektrotehnike i računarstva, pod vodstvom izv. prof. dr. sc. Siniše Šegvića i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2014./ 2015.

Sadržaj

1. Uvod	1
2. Pregled literature	5
3. Uvjetna slučajna polja.....	8
3.1. Osnovni pojmovi	8
3.2. Grafički modeli	9
3.2.1. Usmjereni grafički modeli.....	9
3.2.2. Neusmjereni grafički modeli.....	10
3.2.3. Zaključivanje	13
3.2.4. Učenje.....	15
3.3. Uvjetna slučajna polja.....	16
3.3.1. Zaključivanje	17
3.3.2. Učenje.....	21
4. Semantička segmentacija.....	22
4.1. Unarni potencijali	23
4.1.1. Konvolucijska mreža Farabeta i LeCuna [9][1]	23
4.1.2. Konvolucijska mreža Kreše et al. [22].....	25
4.1.3. Funkcija gubitka.....	26
4.2. Gaussovi binarni potencijali	27
5. Eksperimenti.....	33
5.1. Skupovi podataka	33
5.2. Metrike.....	34
5.3. Učenje parametara uvjetnog slučajnog polja	35
5.4. Uvjetno slučajno polje s konvolucijskim unarnim potencijalima	35
5.5. Ostali eksperimenti	38
5.5.1. Otpornost na šum i diskretizaciju	38
5.5.2. Eksperimenti s konvolucijskom mrežom Farabeta i LeCuna [9][1].....	42

5.5.3.	Uvjetna slučajna polja kao rekurzivne neuronske mreže	54
5.5.4.	Primjena naučenih značajki pri definiciji binarnih potencijala	55
5.5.5.	SegNet.....	58
5.5.6.	Konvolucijske značajke u binarnim potencijalima	60
5.5.7.	Dubinska informacija u binarnim potencijalima	63
6.	Zaključak	70
	Zahvale.....	72
	Literatura	73
	Sažetak.....	76
	Summary	77

1. Uvod

Temeljni problem u računalnom vidu jest kako na temelju dostupnih vizualnih podataka dobivenih kamerom, zaključiti nešto o stanju stvarnoga svijeta. Dostupni podaci su najčešće pojedinačne slike ili sekvenca slika, odnosno video isječci, dok stanje svijeta može predstavljati raspodjelu interesantnih područja na slici, prisustvo određenih objekata, raspored, položaj i strukturu objekata od interesa i slično. Općenito stanje svijeta je uski segment stvarnosti o kojem se želi nešto zaključiti na temelju vizualnih podataka.

Problemu se u pojedinim situacijama može pristupiti na klasičan način, tako da se uzme u obzir znanje o procesu stvaranja slike (perspektivna projekcija, zakon o refleksiji svjetlosti i sl.), te na temelju toga izvede zaključak o stanju stvarnoga svijeta. Takav pristup, zasnovan na razvoju složenih matematičkih modela koji opisuju kompleksne prirodne pojave, presložen je i vremenski prezahtjevan za širu uporabu.

Drugi, i danas mnogo rasprostranjeniji pristup temelji se na strojnom učenju, gdje se treniraju odgovarajući modeli za rješavanje specifičnih problema, poput detekcije i raspoznavanja pješaka, znakova i slično. Treniranje modela sastoji se u estimaciji parametara koji opisuju vezu između skupa mjerenja (vidljivih varijabli) i stanja svijeta. Broj i tip parametara ovisi o konkretnom modelu. Kod stroja s potpornim vektorima (*engl. Support Vector Machines*) to bi bili potporni vektori i odgovarajući koeficijenti, a kod probabilističkih modela gustoće vjerojatnosti.

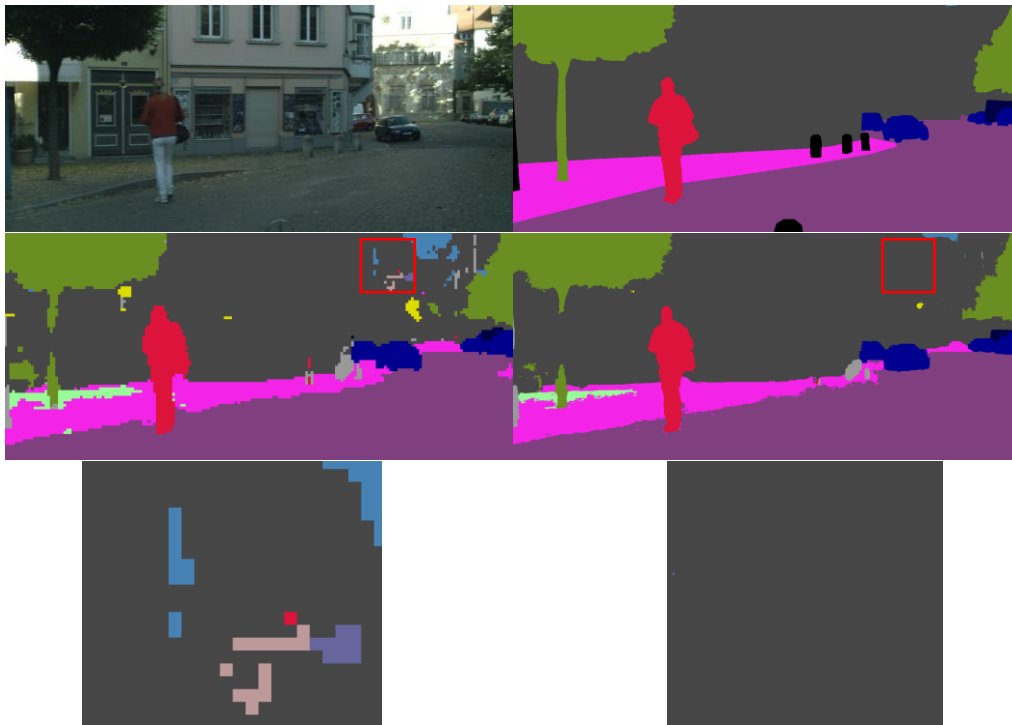
Tipična primjena klasičnih modela strojnog učenja svodi se na opis ulazne slike ili pojedinačnih piksela pomoću vektora značajki, koji se zatim nastoji preslikati u željeni izlaz. Pritom izlaz modela može predstavljati informaciju o prisustvu objekta u slici, informaciju o pripadnosti piksela objektu i slično. Ovakav pristup prikladan je za mnoštvo problema, međutim u nekim situacijama potrebni su modeli koji su u stanju iskoristiti strukturne informacije u slici, kako bi mogli donositi bolje zaključke. Razmotrimo jednostavan problem segmentacije slike, gdje je potrebno za svaki piksel odrediti, predstavlja li pozadinu ili traženi objekt. Naivni pristup uključivao bi opis svakog piksela pomoću vektora svjetlosnih intenziteta, te primjenu nekog od prethodno opisanih modela

strojnog učenja (primjerice stroja s potpornim vektorima) na svaki piksel u slici. Nedostatak navedenog pristupa jest zanemarivanje ovisnosti između oznaka pojedinih piksela. Naime, ovako primijenjeni modeli donose neovisne odluke o klasifikaciji za svaki pojedini piksel, što će rezultirati šumom u segmentiranoj slici (primjerice, pojedini pikseli unutar objekta mogu biti klasificirani kao pozadina). Naravno, moguće je izgraditi modele koji će uzimati u obzir i oznake susjednih piksela, međutim takvi pristupi nastoje pronaći lokalno najizgledniju raspodjelu oznaka na temelju dostupnih mjerenja, koja ne mora nužno odgovarati i globalnom optimumu. Iz tog razloga, primjenjuju se grafički modeli koji omogućavaju predviđanje sekvence međusobno ovisnih oznaka, tako da izgrađuju niz lokalnih modela koji su međusobno povezani. Grafički modeli su pogodni za takav tip problema budući da modeliraju kompleksnu distribuciju velikog broja varijabli kao produkt lokalnih faktora temeljenih na manjim podskupovima varijabli.

Uvjetna slučajna polja (*engl. Conditional Random Fields - CRF*) vrsta su diskriminativnih probabilističkih grafičkih modela, koji su pronašli široku primjenu u računalnom vidu i obradi prirodnog jezika, u područjima koja zahtijevaju strukturiranu predikciju. U okviru obrade prirodnog jezika, uvjetna slučajna polja uspješno su primijenjena za rješavanje problema raspoznavanja imenovanih entiteta [10], pronalaska semantičkih uloga u tekstu [12], integraciji informacija [11], morfološku analizu u japanskom jeziku [13] i sl. Uvjetna slučajna polja primjenjuju se u računalnom vidu za rješavanje problema oduzimanja pozadine, gdje je cilj odvojiti piksele objekta od piksela pozadine. Uklanjanje šuma u slici također se može izvesti primjenom uvjetnih slučajnih polja. Raspoznavanje objekata uspješno je izvedeno primjenom uvjetnog slučajnog polja s latentnim varijablama koje opisuju pojedine dijelove objekta od interesa [14]. Problem stereo podudaranja, gdje je potrebno odrediti parove piksela, koji predstavljaju istu točku u 3D prostoru, u slikama snimljenim pomoću para kamera moguće je riješiti uvjetnim slučajnim poljima.

Međutim najčešća primjena uvjetnih slučajnih polja u okviru računalnog vida jest semantička segmentacija. Problem semantičke segmentacije predstavlja nadogradnju prethodno opisanog problema segmentacije slike na pozadinu i željeni objekt. Cilj je svakom pikselu slike pridružiti oznaku semantičke klase kojoj taj piksel pripada (npr. nebo, cesta, pješak itd.). Uvjetna slučajna polja omogućavaju postizanje boljih rezultata

za problem semantičke segmentacije u odnosu na klasične modele budući da u obzir uzimaju informaciju iz širokog susjedstva prilikom određivanja oznake svakog piksela. Klasični modeli, poput neuronskih mreža koje klasificiraju piksele u slici isključivo na temelju njihovog susjedstva, svoje odluke baziraju na unarnim potencijalima, dok uvjetna slučajna polja kroz binarne potencijale (ili čak potencijale višeg reda) uključuju strukturne informacije u proces donošenja odluke (Slika 1). Semantička segmentacija predstavlja vrlo zanimljivo područje računalnog vida budući da se segmentirane slike mogu koristiti u robotici, autonomnom upravljanju vozilima, te nizu drugih inteligentnih transportnih sustava.



Slika 1 Usporedba klasičnih modela (unarni potencijali) i uvjetnih slučajnih polja kod problema semantičke segmentacije. Prvi redak sadrži izvornu ručno označenu sliku. U sljedećem retku su s lijeva na desno prikazani izlazi klasičnog modela (unarnih potencijala) i uvjetnog slučajnog polja. Zadnji redak prikazuje uvećane odsječke prethodnih slika u kojima je vidljivo kako uvjetno slučajno polja nadjačava odluku unarnih potencijala te eliminira piksele neba (i drugih objekata) unutar piksela zgrade.

U okviru ovog rada istražuje se primjena uvjetnih slučajnih polja za rješavanje problema semantičke segmentacije. Razmatraju se unarni potencijali definirani putem

konvolucijskih neuronskih mreža, te binarni potencijali u obliku Gaussovih funkcija. Performanse uvjetnog slučajnog polja ispituju se na javno dostupnim skupovima podataka za područje razumijevanja urbanih prometnih scena.

2. Pregled literature

Većina današnjih pristupa rješavanju problema semantičke segmentacije temelji se na konvolucijskim neuronskim mrežama. Razvojem grafičkih kartica omogućeno je učenje veoma dubokih neuronskih mreža. S povećanjem dubine mreža, povećao se i njihov kapacitet, koji je iskorišten za obradu velikih susjedstava piksela prilikom semantičke segmentacije. Prve duboke neuronske mreže korištene su za klasifikaciju slika [15]. Navedene mreže sadrže milione parametara, koji su učeni na ogromnom skupu slika. Pritom njihova receptivna polja (skup piksela koje pojedini neuron u mreži vidi prilikom klasifikacije) nadmašuju 200x200 piksela. Pritom se nepoznata slika na ulazu mreže povezuje s jednom od 1000 mogućih klasa.

Navedene duboke neuronske mreže prilagođene su rješavanju problema klasifikacije. Prvi korak k semantičkoj segmentaciji bilo je njihova pretvorba u potpuno konvolucijske neuronske mreže [3]. Spomenuta procedura svodi se na pretvorbu potpuno povezanih slojeva u ekvivalentne konvolucijske slojeve s istim težinama. Posljedica navedene transformacije jest promjena dimenzionalnosti izlaza neuronske mreže. Izlaz potpuno konvolucijske neuronske mreže predstavlja tenzor koji za svaki piksel i svaku klasu definira mjeru sigurnosti. Pritom je izlaz manje dimenzionalnosti od ulazne slike zbog sažimajućih slojeva u konvolucijskoj neuronskoj mreži.

Kako bi se na izlazu mreže dobila informacija o pripadnosti piksela klasama iste rezolucije kao i ulazna slika, potrebno je provesti naduzorkovanje. Tipično se primjenjuje bilinearna interpolacija ili interpolacija najbližim susjedom. Međutim pritom se ne može očekivati preveliki povratak informacije izgubljene tijekom sažimanja. Tom problemu nastoji se doskočiti učenjem naduzorkovanja (dekonvolucijskih slojeva) [3], te iskorištavanjem indeksa piksela korištenih u sažimajućim slojevima [6]. Primjenom zapamćenih indeksa generira se rijetka mapa većih dimenzija od ulazne, te se zatim vrši njezina konvolucija s naučenim dekoderskim filtrom. Oba postupka rezultiraju povećanjem performanse u odnosu na bazni model zbog preciznijeg naduzorkovanja semantičkih mapa.

Kako se pojedine klase objekata (primjerice pješaci) mogu na istoj slici pojaviti u različitim veličinama zbog dubine scene, ispostavlja se da prezentiranje različitih varijanti ulaze

slike doprinosi poboljšanju konačne performanse neuronske mreže [9]. Umjesto jedne rezolucije, mreža na ulazu dobiva više rezolucija iste slike, te se zatim aktivacije neurona unutar mreže u određenom trenutku spajaju kako bi se izgradio tenzor značajki koji predstavlja ulaznu sliku na različitim rezolucijama. Navedeni tenzor se potom koristi za generiranje semantičke mape.

Budući da semantička mapa na izlazu neuronske mreže opisuje njezinu sigurnost u klasu svakog piksela, segmentirana slika dobiva se pridruživanjem svakom pikselu one klase u koju je mreža najsigurnija. U kontekstu uvjetnih slučajnih polja, spomenuta semantička mapa predstavlja unarne potencijale. Iako prethodne mreže imaju veoma velika receptivna polja, klasifikacija na razini piksela ne mora u svim slučajevima rezultirati korektnom segmentacijom. Primjerice, pojedini pikseli unutar nekog objekta mogu biti označeni drugom klasom. Kako bi se taj problem riješio potrebno je zahtijevati slaganje između klasa pojedinih piksela. U tu svrhu koriste se binarni potencijali uvjetnih slučajnih polja (Slika 1). Budući da neuronske mreže ne uzimaju u obzir strukturne informacije, moguće je da pojedini dijelovi istog objekta budu označeni različitim klasama (primjerice auto i guma). U takvim situacijama uvjetna slučajna polja mogu iskoristiti informacije dostupne kroz binarne potencijale te piksele gume pridružiti ostatku auta.

Prvi pristupi primjeni uvjetnih slučajnih polja za semantičku segmentaciju omogućavali su definiranje binarnih potencijala isključivo između susjednih piksela zbog računski zahtjevnog zaključivanja. Kasnije je pokazano kako se definiranjem binarnih potencijala pomoću Gaussovih funkcija, navedeno ograničenje može ublažiti [5] budući da se korak slanja poruka u aproksimaciji srednjeg polja može efikasno izvesti kao konvolucija s Gausovim jezgrama. Chen i ostali [16] prikazali su postupak u kojem uvjetno slučajno polje primjenjuju na izlaz potpuno konvolucijske neuronske mreže pri čemu su parametri mreže i uvjetnog slučajnog polja ućeni odvojeno. Navedeni pristup rezultirao je značajnim poboljšanjem rezultata semantičke segmentacije, međutim nedostatak je bilo odvojeno ućenje mreže i uvjetnog slučajnog polja. Taj problem je riješen formuliranjem iteracije srednjeg polja u kontekstu slojeva neuronske mreže [2]. Spomenuti pristup predstavlja aproksimaciju srednjeg polja kao povratnu neuronsku mrežu što omogućava zajedničko treniranje neuronske mreže i uvjetnog slučajnog polja. Opis binarnih potencijala

Gaussovim funkcijama nije uvijek dovoljan kako bi se izrazile složene veze između klasa piksela. Potencijalna mana je simetričnost promatranih funkcija, što se nastoji riješiti učenjem binarnih potencijala [17]. Pritom se definiraju dodatni slojevi neuronske mreže koji djeluju nad vektorima značajki parova piksela dobivenih putem klasične potpuno konvolucijske neuronske mreže. Navedeni slojevi nastoje naučiti asimetrične binarne potencijale, pomoću kojih je moguće modelirati da čovjek sjedi na stolici, a ne obratno. Dodatni pristupi proširuju klasičnu definiciju uvjetnog slučajnog polja sa potencijalima višeg reda, koji uzimaju u obzir detekcije pojedinih objekata, te superpiksele [18]. Pritom su dodatni potencijali prikladno definirani kako bi se omogućilo zajedničko treniranje neuronske mreže i uvjetnog slučajnog polja po uzoru na [2]. U okviru ovog rada razmatra se osnovna definicija uvjetnog slučajnog polja s Gaussovim binarnim potencijalima, te se promatra utjecaj unarnih i binarnih potencijala na performanse semantičke segmentacije.

3. Uvjetna slučajna polja

Uvjetna slučajna polja su diskriminativni neusmjereni probabilistički grafički modeli. Grafički modeli koriste grafove kako bi na sažet način prikazali zajedničku distribuciju vjerojatnosti. U ovom poglavlju najprije se opisuju osnovni pojmovi iz teorije vjerojatnosti potrebni za razumijevanje teoretske pozadine grafičkih modela. Nakon toga detaljnije se razrađuju dva osnovna tipa grafičkih modela: usmjereni i neusmjereni grafički modeli, te se na kraju detaljnije predstavljaju uvjetna slučajna polja

3.1. Osnovni pojmovi

Slučajna varijabla je funkcija $X: \Omega \rightarrow E$ definirana na skupu elementarnih događaja. E predstavlja skup mogućih realizacija te slučajne varijable. Ovisno o tome da li slučajna varijabla može poprimiti samo diskretne ili kontinuirane vrijednosti, razlikujemo diskretne i slučajne varijable. Slučajna varijabla poprima vrijednosti s određenom vjerojatnošću. Vjerojatnost je mjera koja zadovoljava sljedeće aksiome:

- $P(A) \geq 0, \forall A$
 - $P(S) = 1, S$ je siguran događaj
 - $P(A \cup B) = P(A) + P(B),$
A i B su međusobno disjunktne događaji
- (1)

Pretpostavimo da je realizacija slučajne varijable B poznata, $B = b$. Tada je uvjetna vjerojatnost događaja A definirana izrazom (2).

$$P(A = a|B = b) = \frac{P(A = a, B = b)}{P(B = b)} \quad (2)$$

Aposteriorne vjerojatnosti događaja A i B povezane su preko Bayesovog pravila (3).

$$P(A = a|B = b) = \frac{P(B = b|A = a)P(A = a)}{\sum_j P(B = b|A = a_j)P(A = a_j)} \quad (3)$$

Za dvije slučajne varijable X i Y , kažemo da su nezavisne ukoliko poznavanje realizacije jedne varijable nema nikakvog utjecaja na vjerojatnost druge (4).

$$\begin{aligned} P(X = x, Y = y) &= P(X = x) * P(Y = y) \\ P(X = x|Y = y) &= P(X = x) \quad P(Y = y|X = x) = P(Y = y) \end{aligned} \quad (4)$$

Analogno za slučajnu varijablu X kažemo da je uvjetno nezavisna od varijable Y za zadani Z ako su X i Y nezavisni za zadani Z (5).

$$\begin{aligned} P(X = x|Y = y, Z = z) &= P(X = x|Z = z) \\ P(Y = y|X = x, Z = z) &= P(Y = y|Z = z) \end{aligned} \quad (5)$$

Pravilo lanca omogućava faktorizaciju zajedničke vjerojatnosti na sljedeći način:

$$P(x_1, x_2, x_3, \dots, x_k) = \prod_{i=1}^k P(x_i|x_1, x_2, \dots, x_{i-1}) \quad (6)$$

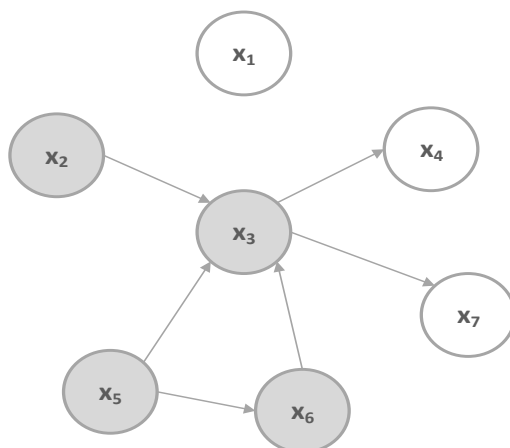
3.2. Grafički modeli

3.2.1. Usmjereni grafički modeli

Usmjereni grafički model je aciklički usmjereni graf, koji predstavlja faktorizaciju zajedničke vjerojatosne distribucije u produkt uvjetnih distribucija vjerojatnosti (7).

$$P(x_1, x_2, \dots, x_N) = \prod_{i=1}^N P(x_i|pa(x_i)) \quad (7)$$

Pritom $pa(x_i)$ vraća sve roditelje slučajne varijable x_i . Grafički modeli definirani na ovakav način nazivaju se Bayesovim mrežama. Usmjereni grafički modeli mogu se promatrati kao grafovi čiji čvorovi predstavljaju slučajne varijable, dok bridovi ukazuju na ovisnosti između varijabli. Ukoliko slučajna varijabla X ovisi o varijabli Y , onda će između njihovih čvorova postojati brid usmjeren od Y prema X . Ako je slučajna varijabla X neovisna o svim ostalim varijablama onda joj u faktoriziranom izrazu za zajedničku vjerojatnost odgovara apriorna vjerojatnost $P(X)$. Primjer Bayesove mreže prikazan je na slici (Slika 2), dok je pripadna faktorizacija zajedničke vjerojatnosti dana sa (8).



Slika 2 Primjer usmjerenog grafičkog modela

$$P(x_1 \dots x_7) = P(x_1)P(x_2)P(x_3|x_2, x_5, x_6)P(x_4|x_3)P(x_5)P(x_6|x_5)P(x_7|x_3) \quad (8)$$

Grafički prikaz Bayesove mreže olakšava zaključivanje o nezavisnosti slučajnih varijabli. Dvije slučajne varijable su nezavisne ako ne postoji usmjereni put od jedne prema drugoj, te ako nemaju zajedničkih predaka. U prethodnom primjeru (Slika 2), slučajna varijabla x_1 neovisna je o svim ostalim varijablama. Slučajne varijable x_3 i x_6 su međusobno ovisne budući da postoji usmjereni put od jedne prema drugoj, te budući da imaju zajedničkog pretka x_5 . Svaka slučajna varijabla uvjetno je neovisna o bilo kojoj drugoj varijabli, koja se ne nalazi u njezinom Markovljevom omotaču, uz zadane varijable iz Markovljevog omotača. Markovljev omotač nekog čvora čine njegovi roditelji, djeca i ostali roditelji djece. U primjeru (Slika 2), sivom bojom prikazani su čvorovi, koji čine Markovljev omotač čvora x_6 . Varijabla x_6 je u tom slučaju uvjetno neovisna o x_4 za zadane varijable x_2, x_3, x_5 .

3.2.2. Neusmjereni grafički modeli

Neusmjereni grafički modeli predstavljaju vjerojatnosnu distribuciju varijabli $\{x_i\}_{i=1}^N$, koja poprima oblik umnoška potencijalnih funkcija $\phi[x_{1\dots N}]$ tako da je

$$P(x_{1\dots N}) = \frac{1}{Z} \prod_{c=1}^C \phi_c[x_{1\dots N}] \quad (9)$$

Potencijalne funkcije ϕ_c uvijek vraćaju pozitivnu vrijednost. Svaka od tih funkcija opisuje tendenciju slučajnih varijabli $x_{1\dots N}$ da poprime određene vrijednosti. U većini problema

računalnog vida, potencijalne funkcije neće biti definirane nad svim varijablama $x_{1\dots N}$, nego nad njihovim podskupovima S_c , koji se nazivaju klike (engl. *cliques*). Faktor Z naziva se partijska funkcija i predstavlja normalizacijski član, koji se koristi kako bi konačni izraz predstavljao vjerojatnosnu distribuciju. U diskretnom slučaju računa se prema (10). U praksi je njegovo izračunavanje dosta skupo te se uglavnom izostavlja.

$$Z = \sum_{x_1} \sum_{x_2} \dots \sum_{x_N} \prod_{c=1}^C \phi_c[x_{1\dots N}] \quad (10)$$

Izraz (9) može se koristiti i u obliku (11)

$$P(x_{1\dots N}) = \frac{1}{Z} \exp \left[- \sum_{c=1}^C \psi_c[x_{1\dots N}] \right], \quad (11)$$

gdje je

$$\psi_c[x_{1\dots N}] = -\log[\phi_c[x_{1\dots N}]] \quad (12)$$

Vjerojatnost napisana u obliku (11) naziva se Gibbsovom distribucijom [29]. U statističkoj mehanici i termodinamici navedena distribucija poznata je kao Boltzmannova distribucija, te se koristi za opis vjerojatnosti određenog stanja sustava kao funkcije energije stanja i temperature sustava na koji se distribucija primjenjuje. Funkcije ψ_c mogu vraćati i pozitivne i negativne vrijednosti, te se mogu promatrati kao cijena svake kombinacije $x_{1\dots N}$. U tom smislu, suma u izrazu (11) ponekad se naziva energija, a proces učenja modela minimizacija energije. Potencijalne funkcije mogu se odnositi na sve slučajne varijable, međutim kao što je prethodno već rečeno, češća je situacija da se odnose samo na podskupove slučajnih varijabli $S \subset \{x_i\}_{i=1}^N$. Izraz (9) u tom slučaju prelazi u oblik (13).

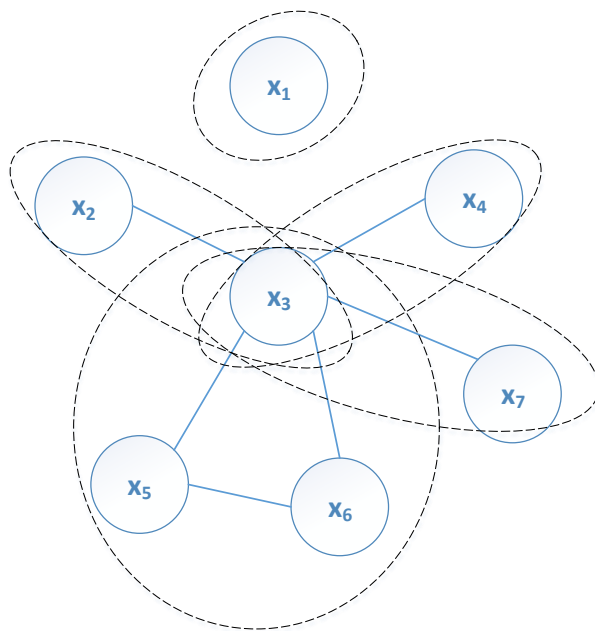
$$P(x_{1\dots N}) = \frac{1}{Z} \prod_{c=1}^C \phi_c[S_c] \quad (13)$$

S_c predstavlja c -ti podskup slučajnih varijabli (kliku). Drugim riječima, vjerojatnosna distribucija faktorizirana je u umnožak izraza koji ovise samo o podskupovima varijabli. Ovakav se model još naziva i Markovljevo slučajno polje. Način vizualizacije ovakvih modela ekvivalentan je vizualizaciji usmjerenih grafičkih modela uz razliku da se umjesto usmjerenih bridova koriste neusmjereni bridovi. Svi čvorovi iz svakog podskupa S_c tvore u grafičkom prikazu potpuno povezani podgraf.

U neusmjerenom grafičkom modelu skup čvorova je uvjetno neovisan o drugom skupu za zadani treći skup ako treći skup razdvaja prva dva tako da prekida svaki put od čvorova iz prvog do čvorova drugog skupa. Na sljedećem grafu (Slika 3), varijabla x_2 je uvjetno neovisna o varijablama x_1, x_4, x_5, x_6, x_7 uz zadanu varijablu x_3 , koja predstavlja njenog neposrednog susjeda (Markovljev omotač) te prekida svaki put od varijable x_2 do svih ostalih. Neusmjereni grafički model koji opisuje faktorizaciju zajedničke vjerojatnosti (14) prikazan je na slici (Slika 3).

$$P(x_1, x_2, \dots, x_7) = \frac{1}{Z} \phi_1[x_1] \phi_2[x_2, x_3] \phi_3[x_3, x_5, x_6] \phi_4[x_4, x_3] \phi_5[x_7, x_3] \quad (14)$$

Zaokruženi su podskupovi varijabli (klike) o kojima ovise potencijalne funkcije.



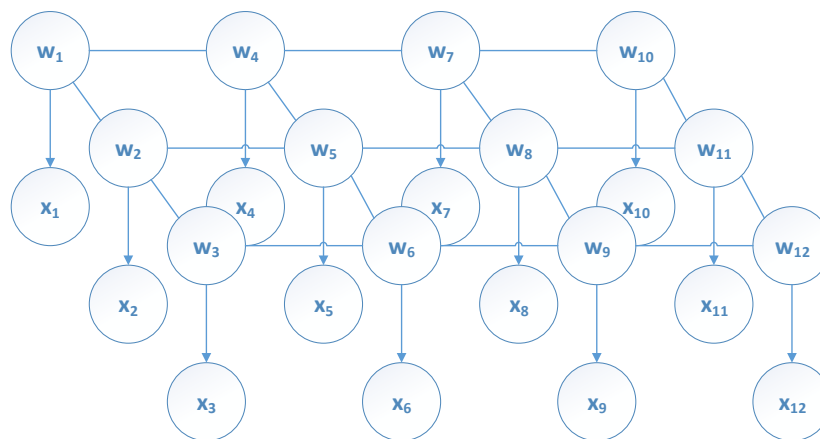
Slika 3 Primjer neusmjerenog grafičkog modela

Bitno je uočiti kako definicija Markovljevog omotača čvora u neusmjerenom grafičkom modelu uključuje samo susjedne čvorove, za razliku od Markovljevog omotača u usmjerenom grafičkom modelu. Razlika nastaje zbog načina definicije uvjetne nezavisnosti u usmjerenom odnosno neusmjerenom grafu budući da Markovljev omotač nekog čvora S sadrži minimalan skup čvorova različitih od S uz koje je S uvjetno nezavisan od svih preostalih čvorova u grafu. Karakterističan predstavnik neusmjerenih grafičkih

modela jesu uvjetna slučajna polja, te je nastavak rada fokusiran prvenstveno na neusmjerene grafičke modele.

3.2.3. Zaključivanje

Markovljeva slučajna polja kao i uvjetna slučajna polja su grafički modeli u obliku mreže. Zbog takve strukture prikladna su za obradu slike, budući da se svakom pikselu u slici može pridružiti prikladna slučajna varijabla (čvor u grafu) koja opisuje informaciju od interesa za taj piksel (Slika 4).



Slika 4 Grafički model u obliku mreže. U kontekstu semantičke segmentacije, čvorovi x bi predstavljali piksele, a w varijable čije su vrijednosti iz skupa oznaka semantičkih razreda

Markovljeva slučajna polja predstavljaju generativni model, u kojem se izravno modelira zajednička vjerojatnost $P(w, x)$, za razliku od uvjetnih slučajnih polja koja izravno modeliraju uvjetnu vjerojatnost $P(w|x)$. Zbog toga se uvjetna slučajna polja smatraju diskriminativnim modelima. Pritom w predstavljaju slučajne varijable koje označavaju stanje stvarnog svijeta koje nas zanima. U kontekstu semantičke segmentacije, to bi bile semantičke oznake piksela.

Svrha uporabe bilo kojeg od prethodnih modela jest donošenje zaključaka o stanju stvarnog svijeta. To u idealnom slučaju uključuje izračunavanje aposteriorne vjerojatnosti $P(w_{1...N}|x_{1...N})$ pomoću Bayesovog pravila. Broj mogućih stanja svijeta u praktičnim je primjenama prevelik da bi se aposteriorna vjerojatnost mogla izravno odrediti. Mogući pristupi zaključivanju uključuju:

- MAP (engl. *maximum a posterior probability*) procjenitelj – cilj je pronaći ona stanja svijeta koja maksimiziraju aposteriornu vjerojatnost (15).

$$\hat{w}_{1\dots N} = \underset{w_{1\dots N}}{\operatorname{argmax}}\{P(w_{1\dots N}|x_{1\dots N})\} = \underset{w_{1\dots N}}{\operatorname{argmax}}\{P(x_{1\dots N}|w_{1\dots N}) * P(w_{1\dots N})\} \quad (15)$$

- Maksimalne marginalne vjerojatnosti – cilj je estimirati stanje svijeta pomoću maksimalnih marginalnih distribucija.

$$P(w_n|x_{1\dots N}) = \iint P(w_{1\dots N}|x_{1\dots N}) dw_{1\dots n-1} dw_{n+1\dots N} \quad (16)$$

$$\hat{w}_n = \underset{w_n}{\operatorname{argmax}}\{P(w_n|x_{1\dots N})\} \quad (17)$$

Nedostatak primjene ovog pristupa jest odsustvo garancije da optimalna procjena pojedinih varijabli stanja \hat{w}_i rezultira optimalnom globalnom procjenom stanja svijeta $\hat{w}_{1\dots N}$.

- Uzorkovanje aposteriorne distribucije – cilj je aproksimirati aposteriornu distribuciju pomoću njezinih uzoraka budući da je izravni izračun preskup. Ova metoda spada u kategoriju aproksimativnih metoda zaključivanja budući da djeluje na aproksimaciji aposteriorne vjerojatnosti. Osnovna ideja sastoji se u tome da se aposteriorna distribucija aproksimira kao niz delta funkcija od kojih svaka odgovara po jednom uzorku. Na temelju takve aproksimacije može se odrediti srednja vrijednost i varijanca aposteriorne distribucije, kao i MAP procjena, kojoj bi odgovarao uzorak najveće vjerojatnosti. Drugi bi pristup uključivao procjenu marginalnih vjerojatnosnih distribucija na temelju promatranja marginalnih statističkih svojstava uzoraka.

Ključna razlika između MAP procjenitelja i maksimalne marginalne vjerojatnosti jest što se u slučaju marginalne vjerojatnosti, varijable stanja w_i promatraju u izolaciji, neovisno jedna o drugoj, što neizbježno dovodi do očekivanog smanjenja performansi u odnosu na MAP procjenu. Najslabija performansa očekivala bi se u slučaju uzorkovanja aposteriorne distribucije zbog aproksimiranja vjerojatnosne distribucije skupom uzoraka.

3.2.4. Učenje

Pod učenjem grafičkog modela podrazumijevamo učenje njegovih parametara, a ne same strukture modela. U slučaju usmjerenih modela to bi bile uvjetne vjerojatnosti $P(x_n|pa(x_n))$, a u slučaju neusmjerenih modela potencijalne funkcije $\phi_c[x_{1...N}]$.

U nastavku se promatra općeniti opis učenja neusmjerenog grafičkog modela (markovljevog slučajnog polja). Neusmjereni grafički model može se prikazati u faktoriziranom obliku

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{c=1}^C \phi_c[\mathbf{x}, \boldsymbol{\theta}] \quad (18)$$

gdje je $\mathbf{x} = x_{1...N}$, te $\boldsymbol{\theta}$ parametri modela. Uz pretpostavku nezavisnosti uzoraka, primjena procjene najveće izglednosti zahtijevala bi uzimanje u obzir ograničenja o nenegativnosti potencijalnih funkcija ϕ_c . Kako bi se to izbjeglo koristi se prikaz (11) neusmjerenog grafičkog modela. Za optimalne parametre odabiru se oni parametri koji maksimiziraju funkciju (logaritamske) izglednosti L .

$$L = \frac{1}{Z(\boldsymbol{\theta})^I} \exp \left[- \sum_{i=1}^I \sum_{c=1}^C \psi_c[\mathbf{x}^i, \boldsymbol{\theta}] \right] \quad (19)$$

$$\log L = -I * \log[Z(\boldsymbol{\theta})] - \sum_{i=1}^I \sum_{c=1}^C \psi_c[\mathbf{x}^i, \boldsymbol{\theta}] \quad (20)$$

\mathbf{x}^i predstavlja i -ti podatak iz skupa podataka I . U kontekstu segmentacije slike to bi bile slike za učenje i njihove oznake, ovisno o definiciji potencijalnih funkcija.

Deriviranjem izraza (20) dobiva se

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} = -I * \frac{\partial \log[\sum_{\mathbf{x}_i} \exp[-\sum_{c=1}^C \psi_c[\mathbf{x}^i, \boldsymbol{\theta}]]]}{\partial \boldsymbol{\theta}} - \sum_{i=1}^I \sum_{c=1}^C \frac{\partial \psi_c[\mathbf{x}^i, \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \quad (21)$$

Problem kod primjene izraza (21) predstavlja prvi izraz na desnoj strani jednakosti budući da sadrži sumu preko svih mogućih vrijednosti \mathbf{x}_i , što je preskupo za izračunati. Najbolje što se u tom slučaju može napraviti jest aproksimirati gradijent. Metoda koja omogućava aproksimaciju gradijenta logaritamske izglednosti s obzirom na parametre $\boldsymbol{\theta}$ za funkcije oblika

$$P(\mathbf{x}) = \frac{1}{Z(\boldsymbol{\theta})} f[\mathbf{x}, \boldsymbol{\theta}] \quad (22)$$

naziva se kontrastivna divergencija (engl. *contrastive divergence*). Prva komponenta derivacije izraza (22) može se zapisati u obliku (23).

$$\begin{aligned} \frac{\partial \log[Z(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} &= \frac{1}{Z(\boldsymbol{\theta})} * \frac{\partial}{\partial \boldsymbol{\theta}} \left(\sum_{\mathbf{x}} f[\mathbf{x}, \boldsymbol{\theta}] \right) \\ &= \frac{1}{Z(\boldsymbol{\theta})} * \sum_{\mathbf{x}} \frac{\partial}{\partial \boldsymbol{\theta}} f[\mathbf{x}, \boldsymbol{\theta}] \\ &= \frac{1}{Z(\boldsymbol{\theta})} * \sum_{\mathbf{x}} f[\mathbf{x}, \boldsymbol{\theta}] \frac{\partial}{\partial \boldsymbol{\theta}} (\log[f[\mathbf{x}, \boldsymbol{\theta}]]) \\ &= \sum_{\mathbf{x}} P(\mathbf{x}) * \frac{\partial}{\partial \boldsymbol{\theta}} (\log[f[\mathbf{x}, \boldsymbol{\theta}]]) \end{aligned} \quad (23)$$

Posljednji izraz u jednadžbi (23) predstavlja očekivanje derivacije od $\log[f[\mathbf{x}, \boldsymbol{\theta}]]$, koje se može aproksimirati na temelju J neovisnih uzoraka \mathbf{x}^* :

$$\frac{\partial \log[Z(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} = \sum_{\mathbf{x}} P(\mathbf{x}) * \frac{\partial}{\partial \boldsymbol{\theta}} (\log[f[\mathbf{x}, \boldsymbol{\theta}]]) \approx \frac{1}{J} \sum_{j=1}^J \frac{\partial}{\partial \boldsymbol{\theta}} (\log[f[\mathbf{x}_j^*, \boldsymbol{\theta}]]) \quad (24)$$

U tom slučaju gradijent funkcije izglednosti (21) poprimio bi oblik (25).

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} \approx -\frac{I}{J} * \sum_{j=1}^J \frac{\partial}{\partial \boldsymbol{\theta}} (\log[f[\mathbf{x}_j^*, \boldsymbol{\theta}]]) - \sum_{i=1}^I \sum_{c=1}^c \frac{\partial \psi_c[\mathbf{x}^i, \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \quad (25)$$

Primjena ove metode zahtijeva uzimanje uzoraka u svakom koraku iterativnog optimizacijskog postupka što je dosta vremenski zahtjevno.

3.3. Uvjetna slučajna polja

Uvjetno slučajno polje može se formalno definirati na sljedeći način [19]:

Neka je $G = (V, E)$ graf definiran nad $Y = (Y_v)_{v \in V}$, tako da je Y indeksiran vrhovima grafa G . Onda je (X, Y) uvjetno slučajno polje ako slučajne varijable Y_v zadovoljavaju Markovljevo svojstvo s obzirom na graf kada su uvjetovane s X : $P(Y_v | X, Y_w, w \neq v) = P(Y_v | X, Y_w, w \sim v)$, gdje $w \sim v$ označava da su w i v susjedi u grafu G

Markovljevo svojstvo za slučaj semantičke segmentacije osigurava da oznaka nekog piksela v ovisi o oznakama samo onih piksela w s kojima je navedeni piksel povezan u grafu. Pritom su veze između piksela definirane vezama između čvorova koji ih predstavljaju u grafu.

3.3.1. Zaključivanje

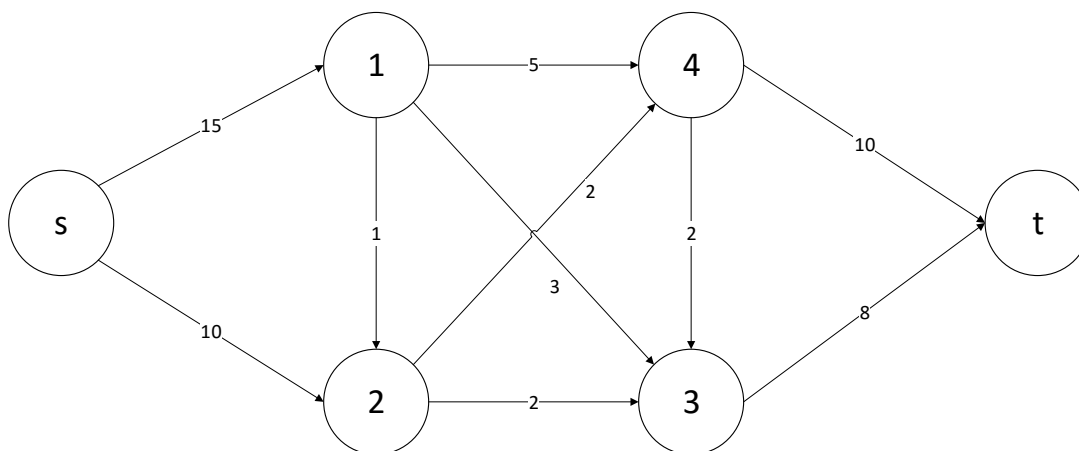
Promotrimo neusmjereni graf $G = (V, E)$ u kojem su potencijalne funkcije definirane samo nad dvije varijable tako da je svaki čvor u graf povezan samo s neposredna četiri susjeda. Cilj MAP zaključivanja je pronaći vrijednosti varijabli stanja $\{w_n\}_{n=1}^N$, koje maksimiziraju aposteriornu vjerojatnost $P(w_{1...N}|x_{1...N})$ tako da je

$$\begin{aligned}
\hat{w}_{1...N} &= \operatorname{argmax}_{w_{1...N}} P(w_{1...N}|x_{1...N}) = \operatorname{argmax}_{w_{1...N}} \left[\prod_{n=1}^N P(x_n|w_n)P(w_{1...N}) \right] \\
&= \operatorname{argmax}_{w_{1...N}} \left[\sum_{n=1}^N \log(P(x_n|w_n)) + \log(P(w_{1...N})) \right] \\
&= \operatorname{argmax}_{w_{1...N}} \left[\sum_{n=1}^N \log(P(x_n|w_n)) - \sum_{(m,n) \in C} \psi[w_m, w_n, \theta] \right] \quad (26) \\
&= \operatorname{argmin}_{w_{1...N}} \left[\sum_{n=1}^N -\log(P(x_n|w_n)) + \sum_{(m,n) \in C} \psi[w_m, w_n, \theta] \right] \\
&= \operatorname{argmin}_{w_{1...N}} \left[\sum_{n=1}^N U_n(w_n, x_n) + \sum_{(m,n) \in C} P_{mn}(w_m, w_n) \right]
\end{aligned}$$

U izrazu (26) iskorištena je pretpostavka da se radi o poljima, čije potencijalne funkcije ovise samo o dvije varijable, što je iskorišteno pri prijelazu iz drugog u treći redak zajedno sa izrazom (11). Promatraju se diskretne slučajne varijable, ali isto razmatranje vrijedi i za kontinuirani slučaj. U_n predstavlja cijenu promatranja podatka na pikselu n uz zadano stanje w_n , a $P_{mn}(w_m, w_n)$ cijenu odabiranja susjednih stanja w_m i w_n . Iz izraza je izbačena normalizacijska konstanta Z budući da nema nikakvog utjecaja na postupak minimizacije. Binarni potencijali $P_{mn}(w_m, w_n)$ mogu se definirati kao $P_{mn}(w_m, w_n) = |w_m - w_n|$ čime se potiče zaglađivanje objekata u slici budući da će susjedni pikseli označeni istom oznakom imati manju cijenu nego kada su označeni različitim oznakama [21]. Izraz (26) predstavlja

pojednostavljenu formulaciju uvjetnog slučajnog polja, gdje su binarni potencijali P_{mn} neovisni o ulaznim podacima x .

Metode koje omogućavaju optimizaciju izraza (26) nazivaju se rezovi grafa (engl. *graph cuts*). Optimizacijski problem nastoji se formulirati u obliku grafa s bridovima kojima su pridružene težine (kapaciteti ili cijene), te se nastoji pronaći takav rez grafa, koji će presjeći one bridove, čija je suma težina optimalna. Osnovna ideja MAP zaključivanja svodi se na transformaciju problema u problem maksimalnog toka (engl. *max-flow*). Primjer problema maksimalnog toka prikazan je na sljedećoj slici (Slika 5). Brojevi na bridovima predstavljaju njihove kapacitete. Cilj je „preusmjeriti“ maksimalni mogući tok kroz graf od izvorišta s do ponora t tako da kapacitet niti jednog brida u grafu ne bude premašen.



Slika 5 Problem maksimalnog toka. Težine na bridovima predstavljaju njihov kapacitet. Potrebno je pronaći maksimalni tok koji se može prenijeti od s do t bez probijanja kapaciteta bridova na putu između s i t .

Navedeni problem formalno se definira na sljedeći način:

- Zadan je usmjereni graf $G = (V, E)$, pri čemu je svakom bridu, koji povezuje vrhove m i n , pridružen nenegativni kapacitet c_{mn} . U skupu V , jedan vrh predstavlja izvor, a drugi ponor.
- Potrebno je odrediti maksimalnu količinu (tok) koji se može prenijeti od izvora do ponora bez probijanja kapaciteta bridova na putu.

- Kada je prethodni zahtjev ostvaren (pronađeno rješenje maksimalnog toka), svaki put od izvora do ponora sadrži zasićeni brid (brid kroz koji „protječe“ tok jednak njegovom kapacitetu)

Navedeni problem ekvivalentan je problemu minimalnog reza, gdje je potrebno pronaći minimalni skup bridova koji razdvajaju izvor od ponora. Cijena pojedinog reza definira se kao zbroj kapaciteta svih bridova u tom rezu. Rez najmanje cijene sastoji se od zasićenih bridova u problemu maksimalnog toka. Problem maksimalnog toka može se riješiti algoritmom rastućeg puta (engl. *augmenting paths*), čiji je pseudokod dan u nastavku (Tablica 1). Složenost navedenog algoritma iznosi $O(|E|^2|V|)$.

Tablica 1 *Augmenting paths* algoritam

```

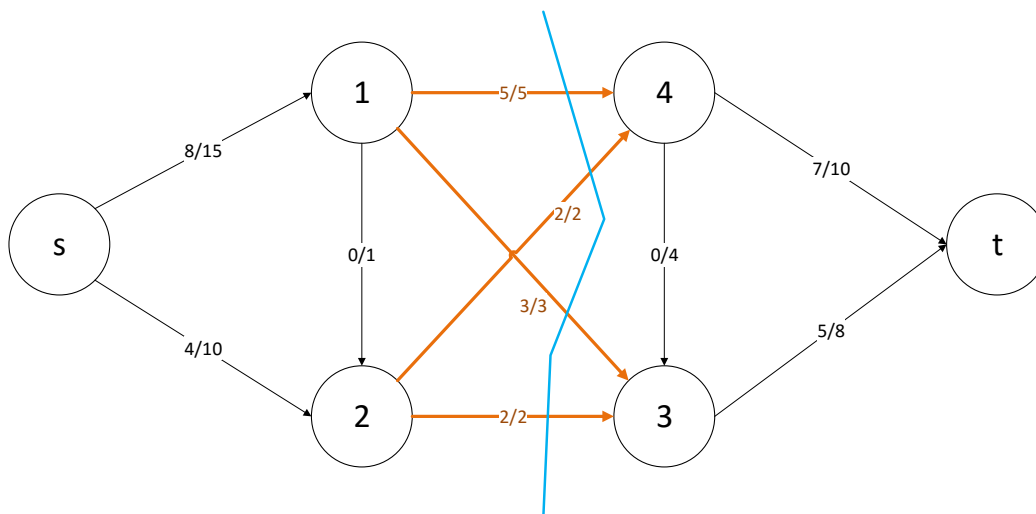
1 Zadan je graf  $G = (V, E)$ , izvor  $s$  i ponor  $t$ 
2 Neka  $f_e$  označava tok kroz brid  $e \in E$ 
3  $f_e = 0$  za  $\forall e \in E$ 
4 Ponavljaj dok postoji put  $p$  od  $s$  do  $t$  takav da  $c_{mn} > 0$  za
   $\forall (m, n) \in p$ 
  a.  $c_p = \min_{(m,n) \in p} \{c_{mn}\}$ 
  b. Za svaki brid  $e = (m, n) \in p$ 
    i.  $f_e = f_e + c_p$ 
    ii.  $f_{e'} = f_{e'} - c_p$  ako postoji  $e' = (b, a) \in E$ , gdje je  $e = (a, b) \in E$ 

```

Rezultat primjene algoritma rastućeg puta na prethodnom primjeru (Slika 5) prikazan je u nastavku (Slika 6). Oznake na bridovima predstavljaju omjer tok/kapacitet. Izvorište je označeno sa s , a ponor sa t . Tijek „slanja“ toka kroz graf je sljedeći:

- Šalje se tok veličine 2 putem $s \rightarrow 2 \rightarrow 3 \rightarrow t$
- Šalje se tok veličine 2 putem $s \rightarrow 2 \rightarrow 4 \rightarrow t$
- Šalje se tok veličine 3 putem $s \rightarrow 1 \rightarrow 3 \rightarrow t$
- Šalje se tok veličine 5 putem $s \rightarrow 1 \rightarrow 4 \rightarrow t$

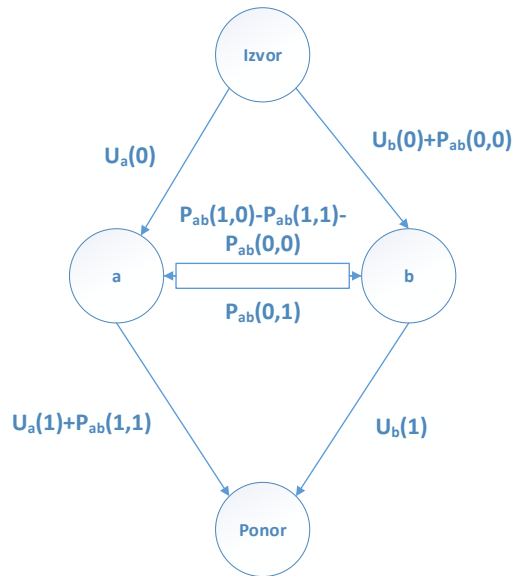
Maksimalna količina toka koja se može poslati kroz graf iznosi 12. Na slici su bojom označeni bridovi koji čine minimalni rez.



Slika 6 Algoritam rastućeg puta

Da bi se MAP zaključivanje moglo provesti potrebno je originalni problem transformirati na način da se može primijeniti prethodno opisani algoritam. Potrebno je definirati usmjereni graf, te pridružiti njegovim bridovima težine tako da minimalni rez na takvom grafu odgovara MAP rješenju.

Za slučaj binarnih varijabli stanja, graf se konstruira na način da se svakom pikselu pridruži po jedan vrh, te se pikseli međusobno povežu dvosmjernim bridovima. Zatim se u graf dodaju ponor i izvor, te se izvor poveže usmjerenim bridom sa svakim pikselom, te se svaki piksel poveže usmjerenim bridom s ponorom. Težine se pridružuju bridovima tako da cijena svakog reza odgovara cijeni odgovarajućih oznaka stanja u izrazu (26). Pritom bridovi moraju imati pridružene nenegativne težine. U suprotnom je potrebno izvršiti reparametrizaciju (dodati bridovima konstantni iznos) kako bi se problem sveo na problem maksimalnog toka. Primjer generiranog grafa za slučaj dva piksela prikazan je na sljedećoj slici (Slika 7). Težine bridova određene su kako bi se zadovoljila relacija (26) za svaku sekvencu oznaka stanja. Primjerice, ako je rješenje $(a = 0, b = 0)$, minimalni rez će prerezati bridove od izvora prema pikselima, te će ukupna cijena biti $U_a(0) + U_b(0) + P_{ab}(0,0)$. Za rješenje $(a = 1, b = 0)$, prerezat će se bridovi $izvor \rightarrow b, a \rightarrow ponor, a \rightarrow b$, te će ukupna cijena biti $U_b(0) + P_{ab}(0,0) + P_{ab}(1,0) - P_{ab}(0,0) - P_{ab}(1,1) + U_a(1) + P_{ab}(1,1) = U_b(0) + U_a(1) + P_{ab}(1,0)$.



Slika 7 Primjer prilagodbe problema zaključivanja problemu maksimalnog toka

Minimizacija izraza (26) može se izvesti i primjenom metoda temeljenih na gradijntnom spustu. Međutim u tom slučaju ne postoji garancija da pronađeno rješenje predstavlja globalni minimum (moguće je zaglavljivanje u lokalnom minimumu). Problem se može shvatiti i kao minimizacija energije gdje lijevi dio minimizacijskog izraza u (26) predstavlja energiju podataka, dok desni dio predstavlja apriornu energiju [20].

Problem zaključivanja ovisi o strukturi uvjetnog slučajnog polja. Složeniji grafovi, koji uključuju širi raspon veza između čvorova, onemogućavaju jednostavnu i efikasnu izvedbu zaključivanja. Međutim u posebnom slučaju kada su klike u grafu ograničene na dvije varijable, te binarni potencijali definirani Gausovim funkcijama moguće je efikasno izvesti zaključivanje primjenom metode srednjeg polja (engl. *mean field*) [5].

3.3.2. Učenje

Učenje parametara uvjetnog slučajnog polja najčešće se ostvaruje pomoću metode najveće izglednosti uz primjenu odgovarajućih tehnika temeljenih na gradijntnom spustu. Konkretna izvedba učenja ovisi o strukturi binarnih potencijala uvjetnog slučajnog polja.

4. Semantička segmentacija

Neka je X polje slučajnih varijabli $\{X_1, X_2, \dots, X_N\}$ koje mogu poprimiti vrijednosti iz skupa oznaka $L = \{l_1, l_2, \dots, l_k\}$. Također, neka je I polje slučajnih varijabli $\{I_1, I_2, \dots, I_N\}$, koje predstavlja piksele u slici. U kontekstu semantičke segmentacije I predstavlja ulaznu sliku, dok X predstavlja oznake pridijeljene svim pikselima u slici I .

Uvjetno slučajno polje nastoji pikselima iz slike I pridijeliti one oznake x^* koje maksimiziraju aposteriornu vjerojatnost $P(x|I)$. Drugim riječima, vrijedi izraz:

$$x^* = \operatorname{argmax}_{x \in L^N} P(x|I) \quad (27)$$

Pritom je uvjetna vjerojatnost definirana izrazom (28).

$$P(X|I) = \frac{1}{Z(I)} e^{-\sum_{c \in C_G} \phi_c(X_c|I)} = \frac{1}{Z(I)} \prod_{c \in C_G} e^{-\phi_c(X_c|I)} \quad (28)$$

Faktor $Z(I)$ naziva se particijska funkcija i predstavlja normalizacijski član, koji se koristi kako bi konačni izraz predstavljao vjerojatnosnu distribuciju. X_c predstavlja skup oznaka pridruženih slučajnim varijablama (pikselima) iz klike $c \in C_G$.

Neka $G = (v, e)$ predstavlja graf definiran nad X , pri čemu svaka klika $c \in C_G$ inducira potencijal ϕ_c . Uvjetno slučajno polje (I, X) opisuje Gibbsovu distribuciju (28). Za potrebe zaključivanja u uvjetnom slučajnom polju, koristi se negativna log izglednost, definirana kao negativni logaritam aposteriorne vjerojatnosti $P(X|I)$. Problem zaključivanja na taj se način formulira kao problem minimizacije negativne log izglednosti, koja se u skladu s optimizacijskim metodama često naziva i energijom. Kako je aposteriorna vjerojatnost $P(X|I)$ definirana preko Gibbsove distribucije (28), pripadna (Gibbsova) energija polja oznaka $x \in L^N$ definirana je izrazom (29). Potencijalne funkcije ϕ_c pritom ne moraju biti nužno nenegativne.

$$E(x|I) = \sum_{c \in C_G} \phi_c(x_c|I) \quad (29)$$

U klasičnim uvjetnim slučajnim poljima klike su sadržavale samo bliske susjede pojedinog čvora zbog prevelikog računskog troška zaključivanja. Međutim, pojava metode aproksimacije srednjeg polja (*engl. mean field approximation*), omogućila je izvedbu zaključivanja u potpuno povezanom polju u veoma kratkom vremenu.

U potpuno povezanom binarnom CRF-u, C_G je skup svih unarnih i binarnih klika u potpunom grafu G . Klike su potpuno povezani podgrafovi grafa G . Binarne klike sadrže samo dva čvora grafa G , te uzimaju u obzir značajke oba piksela, predstavljena prethodno spomenutim čvorovima. Pripadna Gibbsova energija može se prikazati kao (30).

$$E(\mathbf{x}) = \sum_i \phi_u(x_i|\mathbf{I}) + \sum_{i<j} \phi_p(x_i, x_j|\mathbf{I}) \quad (30)$$

4.1. Unarni potencijali

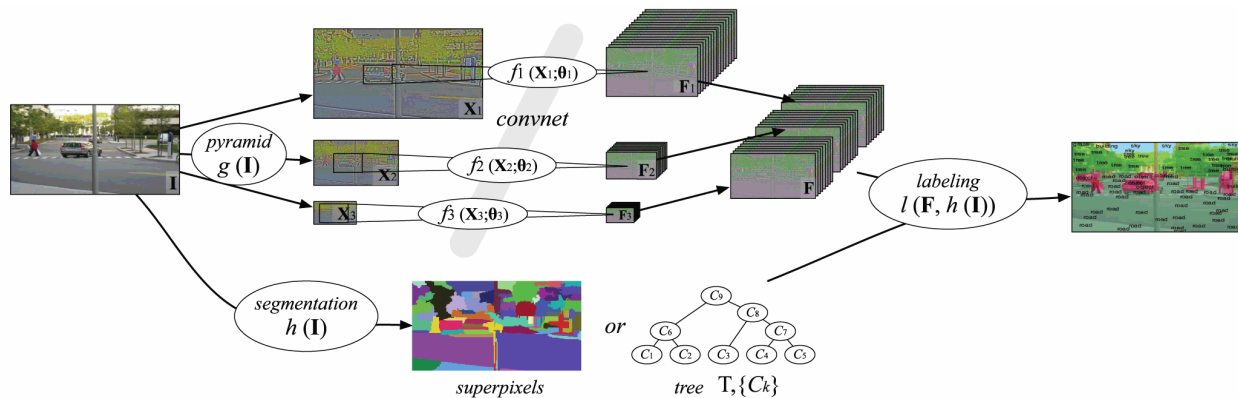
Unarni potencijali (prvi pribrojnik u (30)) modeliraju cijenu oznake pridružene nekom pikselu. Za određivanje cijene promatra se samo taj piksel (čvor u grafu G), odnosno njegove značajke. Promotrimo to na primjeru. Neka se promatra segmentacija slike uz dvije oznake (*nebo, trava*). Pritom neka piksel x_i ima sljedeće RGB vrijednosti (0,255,0). Unarni potencijal bi tom pikselu pridružio vrijednost ∞ za oznaku *trava*, budući da ne sadrži zelenu komponentu, čime bi jako kaznio navedeno označavanje. Suprotno tome, za oznaku *nebo*, vrijednost unarnog potencijala iznosila bi 0, budući da je takvo označavanje poželjno, pa mu je pripadna cijena zanemariva.

Za potrebe generiranja kvalitetnih unarnih potencijala, nije dovoljno promatrati samo jednostavne značajke poput svjetlosnih intenziteta, pa se zbog toga koriste konvolucijske neuronske mreže za generiranje značajki koje bolje opisuju piksele slike. Razmatraju se dvije osnovne arhitekture konvolucijskih mreža koje su opisane u nastavku.

4.1.1. Konvolucijska mreža Farabeta i LeCuna [9][1]

Konvolucijska mreža Farabeta i Lecuna [9][1] jedan je od prvih pristupa segmentaciji temeljenih na učenju s kraja na kraj. Neuronska mreža sastoji se od dvije logičke cjeline: detektora značajki i klasifikatora. Detektor nastoji iz ulazne slike izlučiti informativne značajke primjenom niza konvolucijskih i slojeva sažimanja (*engl. pooling*). Klasifikator,

kojeg čini potpuno povezani sloj neurona i sloj multinomijalne regresije, nastoji na temelju izlučenih značajki označiti piksele pripadnim oznakama razreda.



Slika 8 Arhitektura konvolucijske mreže definirana u [9]. Slika je preuzeta iz [9].

Navedena mreža radi istovremeno sa tri različite rezolucije ulazne slike (faktor poduzorkovanja iznosi 1, 0.5, 0.25). Svaka razina neuronske mreže obrađuje jednu rezoluciju ulazne slike. Pritom svaka razina sadrži po tri konvolucijska sloja. Međutim, isti slojevi u različitim razinama mreže dijele težine usljed čega se smanjuje broj parametara mreže koje je potrebno naučiti. U svakoj razini mreže nakon prva dva konvolucijska sloja nalazi se po jedan sloj sažimanja. Izlazi konvolucijskih slojeva provode se kroz *lReLU* aktivacijsku funkciju. Budući da su izlazi svake razine mreže drugačijih dimenzija, moraju su najprije naduzorkovati na jednaku dimenziju prije nego što se dovedu na ulaz potpuno povezanog sloja.

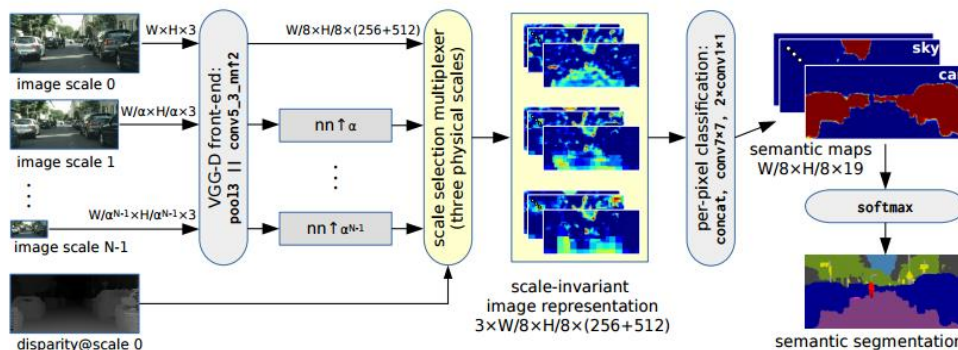
Prethodno opisani slojevi na izlazu generiraju semantičku mapu, koja prolazi kroz softmax sloj generirajući vjerojatnosnu distribuciju po klasama za svaki piksel. Dobivena semantička mapa je manje rezolucije u odnosu na izvornu sliku, te se vrši njezino naduzorkovanje na originalnu rezoluciju. Generirana vjerojatnosna mapa može se zatim iskoristiti za definiranje unarnih potencijala uvjetnog slučajnog polja. Pritom, ako se vrijednost unarnih potencijala dobiva kao $U(x, w) = -\log(P(w|x))$, gdje je $P(w|x)$ vjerojatnost klase w za zadani piksel x , pročitana iz prethodno opisane vjerojatnosne mape.

Izvorni rad za potrebe zaglađivanja semantičke segmentacije, koristi metodu superpiksela, gdje se slični pikseli grupiraju u superpiksele, koji se zatim označavaju

najzastupljenijom klasom. Navedeni pristup postprocesiranju, alternativa je grafičkim modelima, te nije potreban ukoliko se promatrana mreža koristi samo za dobivanje unarnih potencijala. Skica arhitekture mreže prikazana je na slici (Slika 8).

4.1.2. Konvolucijska mreža Kreše et al. [22]

U ovom odlomku razmatra se neuronska mreža čija je arhitektura prikazana na slici (Slika 9). Koristi se analogna ideja kao i u prethodnom poglavlju, gdje se razmatra ulazna slika pri različitim rezolucijama. Međutim, za razliku od prethodno opisane mreže, broj rezolucijskih razina nije fiksiran (3 u prethodnom slučaju). Dio mreže zadužen za detekciju značajki dublji je u odnosu na mrežu iz poglavlja 4.1.1, te se temelji na VGG-D mreži sa 16 slojeva [23]. Pritom je navedena mreža modificirana dodavanjem normalizacijskih slojeva za grupe slika (engl. *batch normalization*) nakon aktivacijskih funkcija u zadnjim konvolucijskim slojevima. Bitna razlika u odnosu na mrežu iz poglavlja 4.1.1 jest što se značajke dobivene pri različitim rezolucijama ne konkatenuiraju prije prosljeđivanja klasifikacijskom dijelu mreže. Umjesto toga definira se posebna komponenta, tzv. multipleksor za odabir skale (engl. *scale selection multiplexer*) čiji je zadatak predstaviti svaki piksel sa konvolucijskim značajkama $M = 3$ od N razina rezolucijske piramide. Odabir konkretne rezolucije formulira se kao minimizacijski problem uzimajući u obzir omjer između referentne skale i projekcije receptivnog polja na temelju informacije o dubini piksela.



Slika 9 Arhitektura konvolucijske mreže definirana u radu [22]

Klasifikacijski dio mreže sastoji se od tri konvolucijska sloja. Prvi sloj ima filtre veličine 7×7 , a preostala dva filtre veličine 1×1 . Analogno mreži opisanoj u poglavlju 4.1.1 izlaz

mreže prolazi kroz dodatni softmax slo, kako bi se dobila vjerojatnosna distribucija po klasama koja se može iskoristiti za definiranje unarnih potencijala uvjetnog slučajnog polja.

4.1.3. Funkcija gubitka

Učenje neuronskih mreža predstavlja postupak određivanja njihovih parametara. Učenje se tipično provodi u nadziranoj konfiguraciji, gdje se mreži predočavaju označeni uzorci za učenje. Drugim riječima, za svaku ulaznu sliku, poznat je željeni izlaz, tj. njezina semantička segmentacija. Budući da se problem optimizacije parametara mreže ne može riješiti u zatvorenom obliku, primjenjuju se gradijentni postupci. Spomenuti postupci zahtijevaju derivabilnu funkciju cilja. Kod neuronskih mreža tipično se definira funkcija gubitka, te se nastoje pronaći oni parametri koji ju minimiziraju.

U eksperimentima provedenim u okviru ovog rada, korištene su sljedeće funkcije gubitka:

- Negativna log izglednost:

$$nll = - \sum_{i=0}^{|D|} \ln P(\mathbf{Y} = \mathbf{y}^{(i)} | \mathbf{x}^{(i)}, \boldsymbol{\theta})$$

pri čemu D predstavlja skup slika za učenje, $\mathbf{y}^{(i)}$ je stvarna segmentacija i -te slike $\mathbf{x}^{(i)}$. $\boldsymbol{\theta}$ su parametri modela, dok je $P(\mathbf{Y} = \mathbf{y}^{(i)} | \mathbf{x}^{(i)}, \boldsymbol{\theta})$ vjerojatnost da je i -ta slika označena labelama $\mathbf{y}^{(i)}$.

- Uravnoteženi gubitak [30] definiran nad cijelim skupom slika:

$$bayesian_{nll} = - \frac{1}{N} \sum_{i=1}^N \ln P(Y_i = c_i) \cdot \frac{1}{P_{apr}(c_i)} \quad (31)$$

pri čemu N predstavlja broj piksela koji se u jednom trenutku razmatraju (cijeli skup ili grupa slika). $P(Y_i = c_i)$ je vjerojatnost da i -ti piksel pripada razredu c_i . c_i je razred kojem i -ti piksel zaista pripada. $P_{apr}(c_i)$ je apriorna vjerojatnost razreda c_i . Apriorne vjerojatnosti mogu se izračunati za cijeli skup slika ili pak samo za pojedinu grupu slika. U prvom slučaju, u izrazu (31) apriorne vjerojatnosti računaju se na temelju svih slika iz skupa za učenje, dok se u drugom slučaju za to koriste samo slike iz trenutne grupe slika.

- Uravnoteženi gubitak definiran nad pojedinom grupom slika (engl. *batch*) – izraz (31)

4.2. Gaussovi binarni potencijali

Binarni potencijali (drugi pribrojnik u (30)) mogu se definirati kao otežana suma Gaussovih jezgri (32) [5].

$$\phi_p(x_i, x_j | \mathbf{I}) = \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \quad (32)$$

$\mu(x_i, x_j)$ je funkcija prikladnosti koja kažnjava susjedne piksele koji su označeni različitim oznakama. Definirana je Pottsovim modelom kao $\mu(x_i, x_j) = [x_i \neq x_j]$, gdje je $[\cdot]$ indikatorska funkcija koja poprima vrijednost 1 ako je $x_i \neq x_j$. Opisana funkcija prikladnosti definira kaznu za susjedne piksele koji su označeni različitim semantičkim oznakama. $w^{(m)}$ su težine pridružene pojedinim jezgrama. Omogućavaju prilagodbu jakosti utjecaja pojedinih jezgara na konačan rezultat semantičke segmentacije. Svaka jezgra ima oblik $k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) = \exp\left(-\frac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)^T \Sigma^{(m)}(\mathbf{f}_i - \mathbf{f}_j)\right)$, gdje je \mathbf{f}_i vektor značajki koji opisuje i -ti piksel, a $\Sigma^{(m)}$ pozitivno definitna matrica koja definira oblik jezgre $k^{(m)}$. Za potrebe semantičke segmentacije binarni potencijali mogu se definirati pomoću dvije jezgre:

- jezgre zaglađivanja (engl. *smoothness kernel*) i
- jezgre izgleda (engl. *appearance kernel*).

Jezgra zaglađivanja opisana je desnom, a jezgra izgleda lijevom eksponencijalnom funkcijom u izrazu (33) [5].

$$\phi_p(x_i, x_j | \mathbf{I}) = \mu(x_i, x_j) \left[w_1 e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}} + w_2 e^{-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}} \right] \quad (33)$$

p_i predstavlja poziciju i -tog piksela, dok I_i predstavlja svjetlosne intenzitete i -tog piksela.

Metoda srednjeg polja omogućava aproksimaciju vjerojatnosti $P(\mathbf{X}|\mathbf{I})$ izračunom distribucije $Q(\mathbf{X})$ koja minimizira KL-divergenciju između svih distribucija Q koje su definirane kao produkt marginalnih distribucija $Q(\mathbf{X}) = \prod_i Q_i(X_i)$ [5]. Marginalne distribucije su vjerojatnosne distribucije definirane nad jednom komponentom vektora slučajnih varijabli. U kontekstu semantičke segmentacije, to bi bile vjerojatnosne distribucije koje opisuju vjerojatnost dodjele oznake klase pojedinom pikselu. Problem kod egzaktnog zaključivanja predstavlja složenost izvorne distribucije $P(\mathbf{X}|\mathbf{I})$, pa se zbog toga nastoji pronaći jednostavnija distribucija $Q(\mathbf{X})$, koja je što bliža (sličnija) distribuciji $P(\mathbf{X}|\mathbf{I})$. Za definiranje udaljenosti distribucija koristi se relativna entropija, tj. KL-divergencija $D(Q||P)$, koja se može zapisati kao:

$$D(Q||P) = E_Q \left[\ln \frac{Q(\mathbf{X})}{P(\mathbf{X}|\mathbf{I})} \right] = E_Q[\ln Q(\mathbf{X})] - E_Q[\ln P(\mathbf{X}|\mathbf{I})] \quad (34)$$

Uvrštavajući u (34) izraz za uvjetnu vjerojatnost (28), dobivamo:

$$\begin{aligned} D(Q||P) &= E_Q[\ln Q(\mathbf{X})] + \ln Z(\mathbf{I}) - E_Q \left[\sum_{c \in C_G} -\phi_c(\mathbf{X}_c|\mathbf{I}) \right] \\ &= \ln Z(\mathbf{I}) - F(P, Q) \end{aligned} \quad (35)$$

$$\begin{aligned} F(P, Q) &= E_Q \left[\sum_{c \in C_G} -\phi_c(\mathbf{X}_c|\mathbf{I}) \right] - E_Q[\ln Q(\mathbf{X})] \\ &= E_Q \left[\sum_{c \in C_G} -\phi_c(\mathbf{X}_c|\mathbf{I}) \right] + H_Q[\mathbf{X}] \\ &= - \sum_{c \in C_G} E_Q[\phi_c(\mathbf{X}_c|\mathbf{I})] + H_Q[\mathbf{X}] \end{aligned} \quad (36)$$

$H_Q[\mathbf{X}]$ predstavlja entropiju od Q , dok je E_Q očekivana vrijednost distribucije Q . Iz (35) vidimo da je minimizacija KL-divergencije ekvivalentna maksimizaciji člana $F(P, Q)$. Uzimajući u obzir da je $Q(\mathbf{X}) = \prod_i Q_i(X_i)$, slijedi:

$$F(P, Q) = - \sum_{c \in C_G} \sum_{X_c} \prod_{X_i \in X_c} Q_i(X_i) \phi_c(\mathbf{X}_c|\mathbf{I}) + \sum_i H_{Q_i}[X_i] \quad (37)$$

Ograničimo $F(P, Q)$ samo na onaj dio koji sadrži $Q_i(X_i)$:

$$F_i[Q] = - \sum_{c \in C_G} \sum_{X_c} \prod_{X_j \in X_c} Q_j(X_j) \phi_c(\mathbf{X}_c | \mathbf{I}) + H_{Q_i}[X_i] \quad (38)$$

Indeks i u prethodnom izrazu označava da se promatra samo i -ta marginalna distribucija dok se ostale distribucije smatraju nepromjenjivim. Definiramo Lagrangeovu funkciju imajući u vidu da Q_i moraju biti distribucije:

$$L_i[Q] = - \sum_{c \in C_G} \sum_{X_c} \prod_{X_j \in X_c} Q_j(X_j) \phi_c(\mathbf{X}_c | \mathbf{I}) + H_{Q_i}[X_i] + \lambda \left(\sum_i Q_i(x_i) - 1 \right) \quad (39)$$

Derivacijom pripadne Lagrangeove funkcije po Q_i dobiva se:

$$\frac{\partial L_i}{\partial Q_i(x_i)} = - \sum_{c \in C_G} \sum_{X_c} \prod_{X_j \in X_c, j \neq i} Q_j(X_j) \phi_c(\mathbf{X}_c | \mathbf{I}, x_i) - \ln Q_i(x_i) - 1 + \lambda \quad (40)$$

Izjednačavanjem (40) s nulom dobiva se:

$$\ln Q_i(x_i) = \lambda - 1 - \sum_{c \in C_G} \sum_{X_c} \prod_{X_j \in X_c, j \neq i} Q_j(X_j) \phi_c(\mathbf{X}_c | \mathbf{I}, x_i) \quad (41)$$

Eksponciranjem izraza (41) i renormalizacijom dobiva se konačni izraz:

$$Q_i(x_i) = \frac{1}{Z_i} \exp \left(- \sum_{c \in C_G} \sum_{X_c} \prod_{X_j \in X_c, j \neq i} Q_j(X_j) \phi_c(\mathbf{X}_c | \mathbf{I}, x_i) \right) \quad (42)$$

λ se izgubi prilikom normalizacije budući da je konstantna u odnosu na x_i . Izraz (42) predstavlja maksimum od $F_i[Q]$ za zadane $Q_j(X_j)_{j \neq i}$ budući da je $F_i[Q]$ konkavna funkcija (prvi pribrojnik u (38) je linearna funkcija po $Q_i(X_i)$, dok je drugi pribrojnik konkavna funkcija po $Q_i(X_i)$).

Ako su sve klike velike jedan ili dva čvora, onda se izraz (42) može prikazati preko unarnih i binarnih potencijala:

$$\begin{aligned}
Q_i(x_i) &= \frac{1}{Z_i} \exp \left(-\phi_u(x_i|\mathbf{I}) - \sum_{j>i} E_{Q_j}[\phi_p(x_i, x_j|\mathbf{I})] \right) \\
&= \frac{1}{Z_i} \exp \left(-\phi_u(x_i|\mathbf{I}) - \sum_{j>i} \sum_{l' \in L} Q_j(l') \phi_p(x_i, x_j|\mathbf{I}) \right) \\
&= \frac{1}{Z_i} \exp \left(-\phi_u(x_i|\mathbf{I}) \right. \\
&\quad \left. - \sum_{j>i} \sum_{l' \in L} Q_j(l') \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \right)
\end{aligned} \tag{43}$$

U konačnici, minimizacija KL-divergencije svodi se na sljedeći iterativni izraz (44)

$$Q_i(x_i = l) = \frac{1}{Z_i} \exp \left\{ -\phi_u(x_i|\mathbf{I}) - \sum_{l' \in L} \mu(l, l') \sum_{m=1}^K w^{(m)} \sum_{j>i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l') \right\} \tag{44}$$

Prethodni izraz vodi do sljedećeg iterativnog algoritma zaključivanja (Tablica 2),

Tablica 2 Metoda srednjeg polja u potpuno povezanom CRF-u

<p>Inicijaliziraj Q: $Q_i(x_i) \leftarrow \frac{1}{Z_i} \exp(-\phi_u(x_i \mathbf{I}))$</p> <p>Ponavljaj do konvergencije:</p> $\hat{Q}_i^{(m)}(l) \leftarrow \sum_{j>i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l) \text{ za } \forall m$ $\hat{Q}_i(x_i) \leftarrow \sum_{l \in L} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \hat{Q}_i^{(m)}(l)$ $Q_i(x_i) \leftarrow \exp\{-\phi_u(x_i \mathbf{I}) - \hat{Q}_i(x_i)\}$ <p>Normaliziraj $Q_i(x_i)$</p>

Cjelokupni algoritam prikazan u prethodnoj tablici (Tablica 2) prikazuje iterativnu izvedbu aproksimacije srednjeg polja (44) rastavljenju na glavne korake. Prva dva koraka unutar petlje predstavljaju utjecaj svih susjeda po jezgrenoju funkciji m . Zadnja dva koraka predstavljaju lokalno osvježavanje i normalizaciju distribucija pomoću unarnih potencijala, koji su tijekom cijelog iterativnog postupka konstantni. Efikasna izvedba opisanog algoritma moguća je uz Gaussove binarne potencijale, koji omogućavaju opis i efikasnu

izvedbu prvog koraka unutar petlje prethodnog algoritma pomoću konvolucija s Gausovim jezgrama.

Kao što je već naglašeno, u prethodnom algoritmu problematičan je prvi korak unutar petlje, koji predstavlja slanje poruka od svih X_j do svih X_i . Navedeni korak moguće je efikasno riješiti pomoću visoko dimenzionalnog filtriranja. Navedeni korak može se opisati pomoću konvolucije sa Gausovom jezgrom u prostoru značajki (45).

$$\begin{aligned}\hat{Q}_i^{(m)}(l) &= \sum_{j \in v} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l) - Q_i(l) = [G_{\Sigma^{(m)}} * \mathbf{Q}(l)](\mathbf{f}_i) - Q_i(l) \\ &= \bar{Q}_i^{(m)}(l) - Q_i(l)\end{aligned}\quad (45)$$

Navedena konvolucija obavlja nisko-propusno filtriranje, te pritom pojasno ograničava $\bar{Q}_i^{(m)}(l)$ budući da se koristi filter konačne veličine, kod kojeg su svi elementi izvan dvije standardne devijacije postavljeni u nulu. Konvolucija se može izvesti pod-uzorkovanjem $Q(l)$, konvolucijom uzoraka sa Gausovom jezgrom te naknadnim nad-uzorkovanjem rezultata (Tablica 3). Pod-uzorkovanje se koristi kako bi se smanjila vremenska složenost potrebna za izvedbu konvolucije, budući da se konvolucija u nekoj točki u tom slučaju može aproksimirati agregacijom konstantnog broja susjednih vrijednosti.

Tablica 3 Efikasna izvedba slanja poruka $\bar{Q}_i^{(m)}(l) = \sum_{j \in v} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l)$

$Q_{\downarrow}(l) \leftarrow \text{pod-uzorkuj}(Q(l))$ $\forall i \in v_{\downarrow} \bar{Q}_{\downarrow i}^{(m)}(l) \leftarrow \sum_{j \in v_{\downarrow}} k^{(m)}(\mathbf{f}_{\downarrow i}, \mathbf{f}_{\downarrow j}) Q_{\downarrow j}(l)$ $\bar{Q}^{(m)}(l) \leftarrow \text{nad-uzorkuj}(\bar{Q}_{\downarrow}^{(m)}(l))$

Aproksimacija koraka slanja poruka može se efikasno izvesti uporabom permutoedarske rešetke (engl. *permutohedral lattice*), koja omogućava učinkovitu izvedbu filtriranja u visoko dimenzionalnom prostoru (dimenzionalnost tog prostora odgovara dimenziji vektora značajki, koji opisuje svaki piksel, uvećanoj za jedan). Postupak filtriranja svodi se na to da se svakoj vrijednosti koja se želi filtrirati pridijeli jedna pozicija u novom vektorskom prostoru, te se zatim sve vrijednosti zamijene sa težinskom kombinacijom svih ostalih vrijednosti, pri čemu su težine Gaussove funkcije pozicija pridijeljenih promatranim vrijednostima. Permutoedarska rešetka pritom dijeli novi vektorski prostor uniformnim simpleksima, što omogućava jednostavno preslikavanje izvornog vektora značajki (signala) u točke rešetke (engl. *lattice*) primjenom baricentrične interpolacije. Zatim se izvodi konvolucija sa jednodimenzionalnim jezgrama duž svake osi rešetke, što u konačnici rezultira konvolucijom s Gaussovom jezgrom u visoko dimenzionalnom prostoru [8].

5. Eksperimenti

U ovom poglavlju detaljno se opisuju provedeni eksperimenti s konvolucijskim unarnim i Gausovim binarnim potencijalima. Najprije se opisuju korišteni skupovi podataka, te se utvrđuju korištene evaluacijske metrike. Eksperimentalni rezultati se izlažu prema kvaliteti dobivenih rezultata. U poglavlju se razmatraju ograničenja uvjetnih slučajnih polja kroz zašumljivanje ulaza kao i unarnih potencijala, utjecaj unarnih, a potom i binarnih potencijala na performanse semantičke segmentacije. Pritom se dobiveni rezultati detaljno komentiraju.

5.1. Skupovi podataka

Za potrebe ispitivanja performansi naučenih modela, korišteni su javno dostupni skupovi podataka za semantičku segmentaciju. Pritom je većina eksperimenata rađena na skupovima koji sadrže slike urbanih prometnih scena: Cityscapes [24] i KITTI [25].

Skup podataka KITTI [25] sadrži veliku kolekciju video isječaka rezolucije 1241x376. Izvorni skup podataka nije označen za semantičku segmentaciju, pa se koristi podskup od 146 slika označenih s 11 klasa [26]. Za potrebe učenja neuronske mreže navedeni skup je proširen s ručno anotiranih 299 dodatnih slika, pri čemu je za učenje mreže korišteno 399 slika, dok je 46 slika korišteno za testiranje. Za potrebe određivanja parametara uvjetnog slučajnog polja korišten je validacijski skup sa nasumično odabranih 10 slika iz izvornog skupa za učenje.

Skup podataka Cityscapes [24] sadrži 19 klasa objekata sadržanih u slikama snimljenim u 50 gradova u različitim vremenskim uslovima. Sadrži 5000 označenih slika, od čega je 2975 korišteno za učenje, 500 za validaciju, te ostatak za testiranje. Slike imaju rezoluciju 2048x1024, te su dostupne i stereo slike za rekonstrukciju dubine. Za potrebe određivanja parametara uvjetnog slučajnog polja korišten je podskup validacijskog skupa (200 nasumično odabranih slika).

Stanford Background [27] sadrži 715 slika vanjskih scena dimenzija 320x240 piksela. Pritom svaka slika sadrži barem jedan objekt od interesa, te se linija horizonta nalazi

unutar slike. Skup je podijeljen na skup za učenje (548 slika), skup za validaciju (60 slika), te skup za ispitivanje (107 slika). Na slikama je označeno 8 semantičkih klasa.

Skup podataka Pascal VOC 2010 [28] sadrži 1928 slika približne veličine 500x400 piksela, te ukupno 21 klasu (20 klasa objekata i 1 klasa pozadine). Slike navedenog skupa korištene su prilikom početnih razmatranja utjecaja zašumljenja na performanse uvjetnog slučajnog polja. Pritom su parametri uvjetnog slučajnog polja određeni na temelju nasumično odabranog podskupa slika (200 slika).

5.2. Metrike

Kako bi bila moguća usporedba performansi pojedinih modela, koriste se sljedeće metrike:

- Točnost piksela (engl. *pixel accuracy*) – omjer broja točno označenih piksela naspram ukupnog broja piksela:
- Točnost razreda – omjer broja točno označenih piksela pojedinog razreda naspram ukupnog broja piksela tog razreda.
- Srednja točnost razreda – prosjek točnosti razreda:

$$accuracy_{class}^{(mean)} = \frac{1}{|C|} \sum_{c=1}^{|C|} accuracy_{class}(c)$$

Pritom je C skup razreda, dok $accuracy_{class}(c)$ predstavlja točnost razreda c .

- IOU (engl. *intersection over union*) – često korištena mjera za semantičku segmentaciju:

$$IOU_c = \frac{TP}{TP + FP + FN}$$

Pritom TP predstavlja broj ispravno označenih piksela, FP broj piksela koji su pogrešno označeni promatranom klasom, te FN broj piksela koji su označeni drugom klasom iako u stvarnosti predstavljaju promatranu klasu. IOU se definira za pojedinu klasu c . Srednji IOU dobiva se kao aritmetička sredina IOU -a pojedinih razreda.

Točnost piksela u većini slučajeva nije dovoljno izražajna mjera budući da se kod problema semantičke segmentacije pojavljuju slabo zastupljene klase, te se može desiti

da model počne zanemarivati navedene klase, što se pak ne očituje u točnosti piksela. Kako bi se moglo pratiti ponašanje modela i na slabo zastupljenim razredima, definira se točnost razreda, odnosno srednja točnost razreda i IOU.

5.3. Učenje parametara uvjetnog slučajnog polja

U svim narednim eksperimentima, ukoliko nije drugačije navedeno, parametri uvjetnog slučajnog polja određuju se pretragom po rešetci. Pritom se pretraga vrši na validacijskom dijelu skupa podataka. Vrši se gruba pretraga, kod koje se fiksiraju parametri jezgre izgladivanja na vrijednosti ($w_2 = 3, \theta_\gamma = 3$), te se pretražuje diskretna rešetka kako bi se pronašli najbolji parametri jezgre izgleda iz skupa $w_1 \in \{5,10\}, \theta_\alpha \in \{50,60,70,80,90\}, \theta_\beta \in \{3,4,5,6,7,8,9\}$. Prilikom zaključivanja pomoću uvjetnog slučajnog polja, broj iteracija srednjeg polja fiksiran je na 10 ukoliko nije drugačije navedeno.

5.4. Uvjetno slučajno polje s konvolucijskim unarnim potencijalima

Za određivanje unarnih potencijala naučena je konvolucijska neuronska mreža prema radu [22]. Pritom je korišten skup podataka Cityscapes, te se koristi IOU kao temeljna metrika. Neuronska mreža učena je stohastičkim gradijentnim spustom uz jednu sliku u grupi (engl. *batch*). Mreža je učena 15 epoha uz početnu stopu učenja od 10^{-5} , koja se duplo smanji nakon druge i desete epohe. Prvih trinaest slojeva mreže inicijalizirano je primjenom javno dostupnog modela VGG-D mreže [23], dok su parametri ostalih slojeva nasumično inicijalizirani. Za potrebe određivanja parametara uvjetnog slučajnog polja vrši se pretraga po rešetci na dijelu validacijskog skupa (200 slika). Prilikom zaključivanja broj iteracija srednjeg polja fiksiran je na 10.

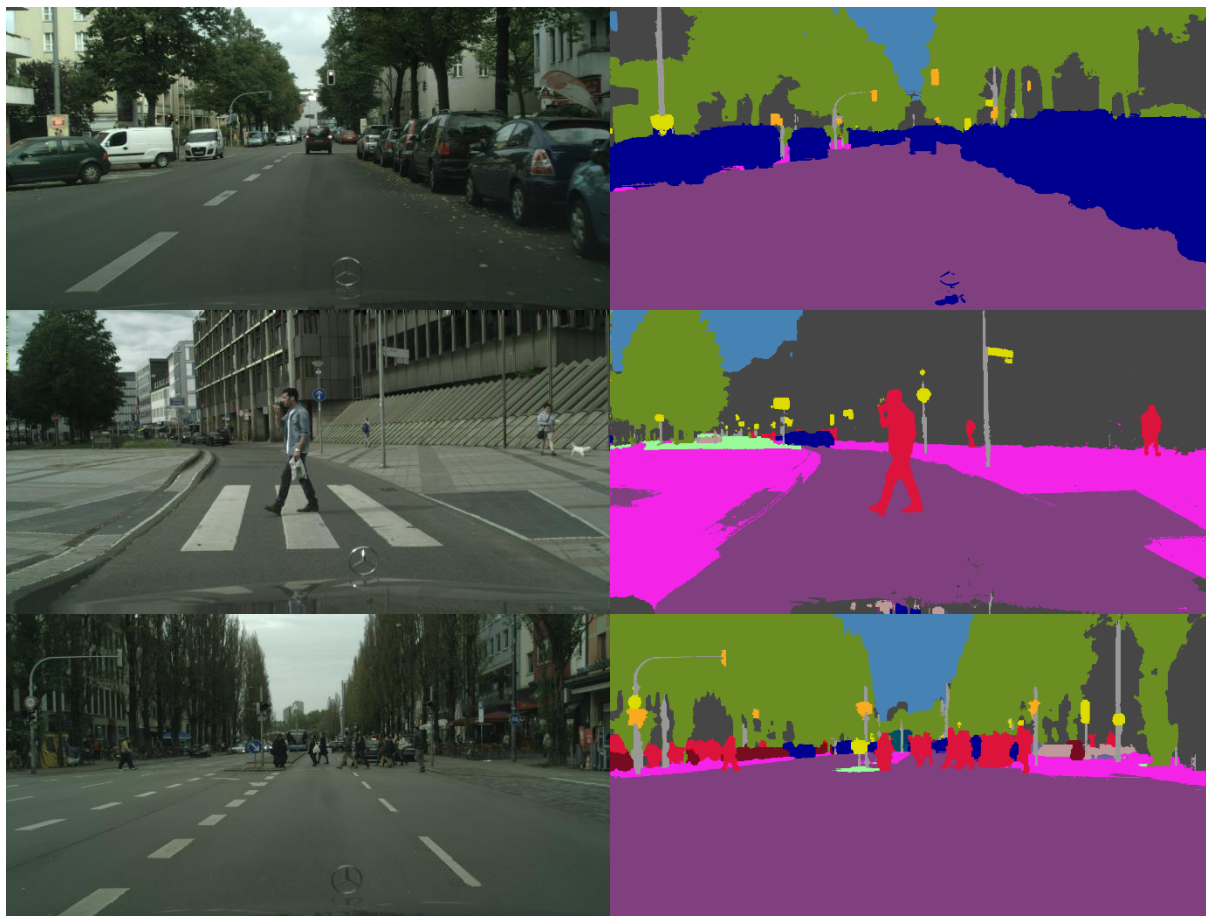
Tablica 4 prikazuje dobivene rezultate na skupu podataka Cityscapes.

Tablica 4 IOU razreda na Cityscapes skupu za ispitivanje

IOU	cesta	pločnik	zgrada	zid	ograda	stup	semafor	znak	vegetacija	teren	nebo	osoba	vozač	auto	kamion	bus	vlak	motor	bicikl	Srednji IOU
Unarni	95.3	73.5	86.4	36.8	42.7	45.5	56.6	57.8	89.5	63.6	90.3	73.5	53.1	90.3	30.8	48.2	39.6	52.2	62.7	62.5
Unarni + CRF	96.3	76.8	88.8	40.0	45.4	50.1	63.3	69.6	90.6	67.1	92.2	77.6	55.9	90.1	39.2	51.3	44.4	54.4	66.1	66.3

Može se uočiti kako primjena duboke arhitekture, te multipleksora za kombiniranje informacija za različite rezolucije, doprinosi izvrsnim unarnim potencijalima, pogotovo za manje zastupljene klase poput motora i vozača. Uvjetno slučajno polje u stanju je iskoristiti kontekstnu informaciju i kvalitetne unarne potencijale kako bi se dodatno poboljšale performanse segmentacije.

Slika 10 prikazuje rezultate semantičke segmentacije na skupu za ispitivanje. Pritom označene slike nisu javno dostupne, te su prikazane samo ulazna slika i izlaz uvjetnog slučajnog polja. Međutim, vidljivo je kako se veoma dobro uspijevaju segmentirati inače problematični razredi poput stupova i prometnih znakova.



Slika 10 Izlaz uvjetnog slučajnog polja - Cityscapes

Navedena mreža primijenjena je i na skup podataka KITTI, pri čemu su dobiveni sljedeći rezultati (Tablica 5). Očito postojanje multipleksora i dublja arhitektura mreže doprinose drastičnom povećanju IOU mjere.

Tablica 5 IOU razreda na KITTI skupu – usporedba s mrežom iz poglavlja 4.1.1 (uravnoteženi gubitak)

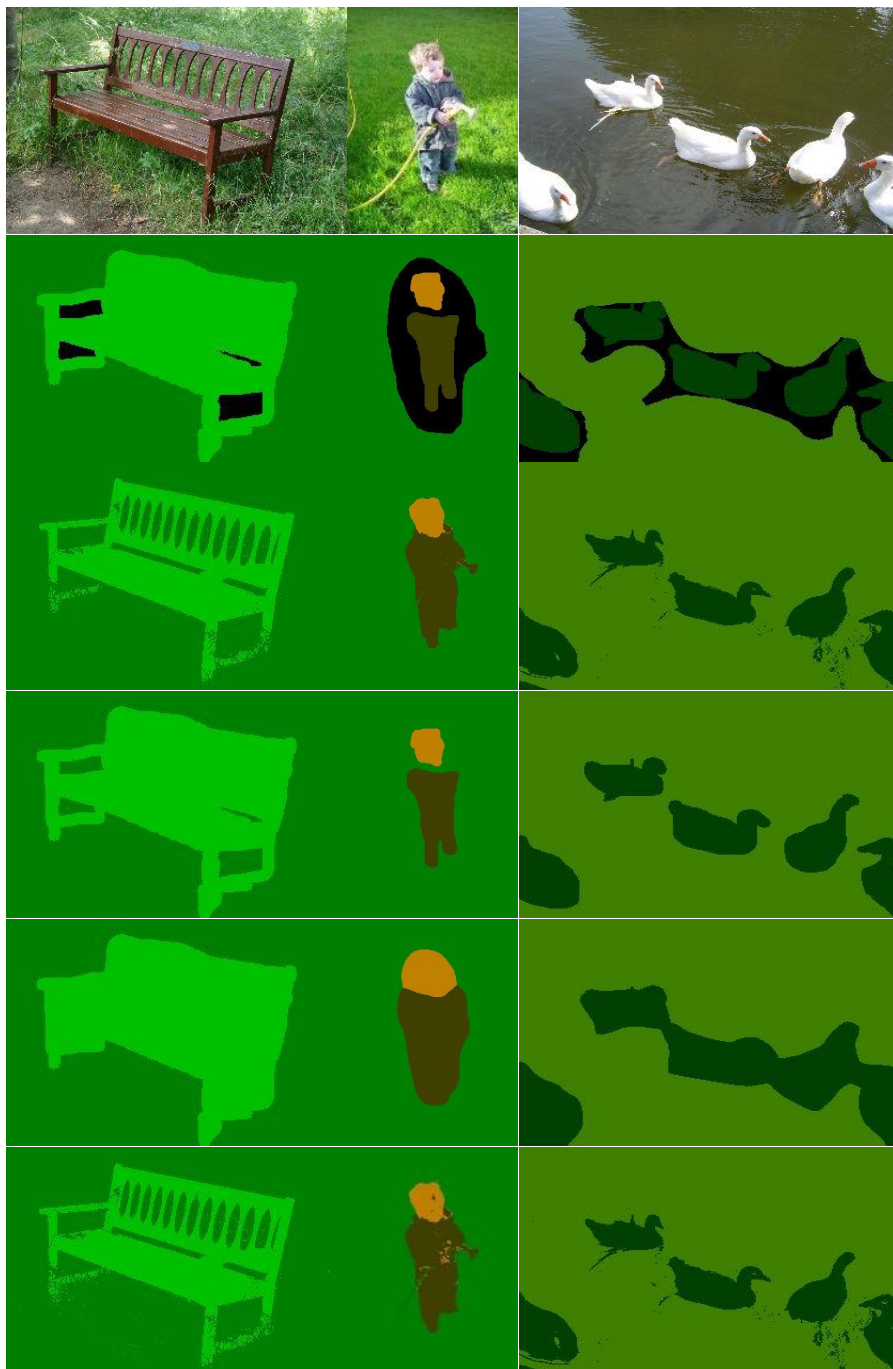
IOU	nebo	zgrada	cesta	pločnik	ograda	vegetacija	stup	vozilo	znak	pješak	biciklist	Srednji IOU
Mreža [1]	79.5	67.4	80.0	51.2	10.6	66.9	9.2	72.8	0.5	0.2	19.9	41.6
Mreža [22]	84.1	83.9	91.4	73.1	43.2	79.2	33.6	82.0	44.7	57.3	12.8	62.3

5.5. Ostali eksperimenti

5.5.1. Otpornost na šum i diskretizaciju

U ovom poglavlju razmatra se otpornost uvjetnog slučajnog polja na šum u ulaznoj slici i unarnim potencijalima. Uvjetno slučajno polje koristi Gaussove binarne potencijale, te aproksimaciju unarnih potencijala. Unarni potencijali određeni su zašumljivanjem anotirane slike. Na temelju oznake za svaki piksel u anotiranoj slici, postavi se odgovarajuća ćelija (koja odgovara toj oznaci) u vjerojatnosnoj mapi na veću vjerojatnost, dok se sve ostale ćelije (isti pikseli, ali druge oznake) postavljaju na manju vjerojatnost. Ako oznaka nekog piksela u anotiranoj slici nije poznata, koristi se uniformna razdioba nad oznakama.

Slika 11 prikazuje rezultate segmentacije za tri slike iz Pascal VOC skupa podataka uz primjenu aproksimiranih unarnih potencijala. Odozgo prema dole prikazane su: izvorne slike, anotirane slike, izlazi dobiveni za sve potencijale, izlazi samo za unarne potencijale, izlazi za unarne potencijale i jezgru zaglađivanja, izlazi za unarne potencijale i jezgru izgleda.



Slika 11 Rezultati segmentacije uz zašumljene unarne potencijale – VOC Pascal. Odozgo prema dole prikazane su: izvorne slike, anotirane slike, izlazi dobiveni za sve potencijale, izlazi samo za unarne potencijale, izlazi za unarne potencijale i jezgru zaglađivanja, izlazi za unarne potencijale i jezgru izgleda

Vidljivo je da CRF radi dosta dobro za kvalitetne unarne potencijale. Naravno, problem je dobiti takve potencijale. Binarni potencijali omogućavaju uzimanje u obzir šireg konteksta svakog piksela te zaglađivanje segmentiranih područja na temelju sličnosti piksela. Konkretno, koristi se jezgra zaglađivanja (definirana nad pozicijama piksela), te jezgra izgleda (definirana nad svjetlosnim intenzitetima i pozicijama piksela). Budući da se koriste navedene jezgre, u nekim situacijama dolazi do pogrešne segmentacije budući da su razmatrani pikseli veoma slični (npr. pikseli crijeva koje drži dijete u usporedbi sa pikselima djeteta).

Slika 12 prikazuje rezultate segmentacije za slučaj aproksimacije unarnih potencijala na temelju zašumljenih anotiranih slika. Odozgo prema dole prikazane su: izvorne slike, zašumljene anotirane slike, izlazi dobiveni za sve potencijale, zašumljene izvorne slike i pripadni izlazi. Na temelju dobivenih rezultata vidljivo je kako zašumljeni unarni potencijali ne predstavljaju preveliki problem za konačnu segmentaciju slika, za razliku od zašumljenih ulaznih slika, budući da šum u ulaznoj slici zbunjuje Gaussove jezgre definirane nad intenzitetima piksela.



Slika 12 Rezultati segmentacije – slike uključene u javno dostupnu implementaciju. Odozgo prema dole prikazane su: izvorne slike, zašumljene anotirane slike, izlazi dobiveni za sve potencijale, zašumljene izvorne slike i pripadni izlazi.

U praktičnoj primjeni uvjetnih slučajnih polja, unarni potencijali generirat će se primjenom konvolucijskih neuronskih mreža. Navedene mreže pritom obavljaju sažimanje ulazne slike pri čemu je neizbježan gubitak informacije. Kako bi se moglo primijeniti uvjetno slučajno polje potrebno je vjerojatnosnu mapu naduzorkovati do izvorne rezolucije. Kako bi se mogla procijeniti sposobnost uvjetnog slučajnog polja da rekonstruira izgubljeni dio informacija, izvodi se eksperiment u kojem se označena slika (engl. *ground truth*) poduzorkuje, a zatim naduzorkuje postupkom najbližeg susjeda. Dobivene

naduzorkovane oznake zatim se ublaže tako da na granicama objekata postoji linearni prijelaz iz jedne klase u drugu, te se zatim primjeni uvjetno slučajno polje s unarnim potencijalima koji odgovaraju ublaženim oznakama. Dobiveni rezultati prikazani su u tablici (Tablica 6). Eksperimenti su provedeni nad KITTI skupom podataka.

Tablica 6 Eksperimentalni rezultati uz diskretizirane unarne potencijale - KITTI

	Točnost	Točnost razreda
Naduzorkovane slike (najbliži susjed)	95.554920 %	85.787123 %
CRF (bez zaglađivanja rubova)	95.776139 %	85.998540 %
CRF (sigurnost u labelu = 0.5)	96.419804 %	84.195245 %
CRF (4-susjedstvo)	96.408836 %	83.703668 %
CRF (8-susjedstvo)	96.463359 %	83.832028 %

Prilikom modeliranja prijelaza između objekata razmatrane su različite vrst susjedstava. Pritom se najveća točnost ostvaruje uz 8-susjedstvo što je i očekivano budući da je u tom slučaju učinak diskretizacije najviše potisnut. Na temelju rezultata vidljivo je kako se ostvaruje kompromis između točnosti i točnosti razreda, zbog čega se za diskretne unarne potencijale ostvaruje najveća točnost razreda, budući da binarni potencijali nisu uspjeli potisnuti slabije zastupljenije klase iz segmentirane slike.

5.5.2. Eksperimenti s konvolucijskom mrežom Farabeta i LeCuna [9][1]

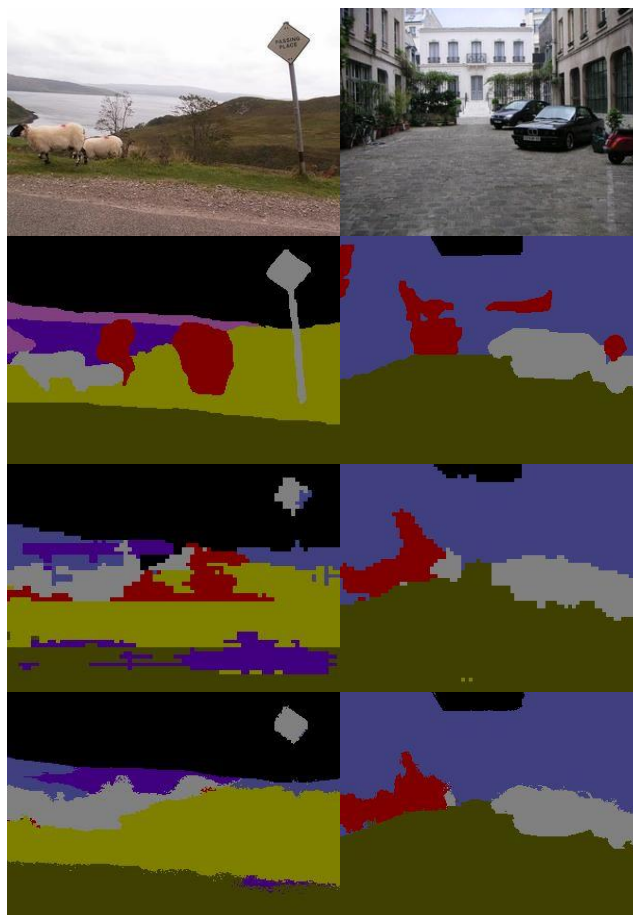
Za određivanje unarnih potencijala naučena je konvolucijska neuronska mreža prema radu [1]. Korištena su dva skupa slika: Stanford Background i KITTI.

Naredni rezultati (Tablica 7) dobiveni su za skup slika Stanford Background.

Tablica 7 Eksperimentalni rezultati uz unarne potencijale iz rada [1]

	Točnost	Točnost razreda
<i>Unarni potencijali</i>	76.237244 %	64.301310 %
<i>CRF</i>	76.516417 %	64.583723 %

Slika 13 prikazuje primjere segmentiranih slika iz skupa Stanford Background. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, izlaz samo uz unarne potencijale (ekvivalentno izlazu konvolucijske neuronske mreže), izlaz uvjetnog slučajnog polja.



Slika 13 Rezultati segmentacije uz unarne potencijale iz rada [1]. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, izlaz samo uz unarne potencijale (ekvivalentno izlazu konvolucijske neuronske mreže), izlaz uvjetnog slučajnog polja.

Vidljivo je kako uvjetno slučajno polje uspijeva iskoristiti strukturnu informaciju i poboljšati performanse semantičke segmentacije.

U nastavku je prikazan utjecaj unarnih potencijala na semantičku segmentaciju skupa podataka KITTI. Prilikom učenja neuronske mreže korištene su tri različite varijante funkcije gubitka opisane u poglavlju 4.1.3:

- Negativna log izglednost
- Uravnoteženi gubitak definiran nad cijelim skupom slika (31)
- Uravnoteženi gubitak definiran nad pojedinom grupom (*engl. batch*) slika za učenje (31)

Dobiveni rezultati prikazani su u tablici (Tablica 8). Optimalni parametri određeni su pretragom po rešetci za slučaj uravnoteženog gubitka definiranog nad grupom slika (Tablica 9).

Tablica 8 Eksperimentalni rezultati uz unarne potencijale iz rada [1]

	Točnost	Točnost razreda
Negativna log izglednost		
<i>Unarni potencijali</i>	74.236634 %	42.176689 %
<i>Optimalni parametri</i>	75.761430 %	42.460751 %
Uravnoteženi gubitak definiran nad grupom slika		
<i>Unarni potencijali</i>	68.493831 %	57.326605 %
<i>Optimalni parametri</i>	75.638651 %	56.505104 %
Uravnoteženi gubitak definiran nad cijelim skupom slika		
<i>Unarni potencijali</i>	51.391618 %	51.986308 %
<i>Optimalni parametri</i>	55.093914 %	51.613221 %

Tablica 9 Optimalni parametri – Uravnoteženi gubitak nad grupom slika

```
// jezgra zaglađivanja
x_stddev = 3
y_stddev = 3
weight = 5

// jezgra izgleda
x_stddev = 78
y_stddev = 78
r_stddev = g_stddev = b_stddev = 3
weight = 5
```

Navedeni optimalni parametri određeni su uz Pottsov model za funkciju kompatibilnosti oznaka piksela. Vidljivo je kako uravnoteženi gubitak definiran nad grupom slika podiže točnost razreda budući da se manje zastupljenim razredima umjetno povećavaju

gradijenti tijekom učenja, tj. daje im se veći značaj. Prednost navedenog gubitka u odnosu na istovjetni gubitak definiran nad cijelim skupom slika jest u tome što se važnost pojedinih razreda povećava samo u slučaju da se taj razred javlja u slici koja se koristi tijekom učenja (točnije u jednoj grupi slika - *batchu*). Budući da je, primjerice razred pješaka dosta slabo zastupljen u skupu slika za učenje, tijekom učenja mreže prema uravnoteženom gubitku definiranom nad cijelim skupom, velika težina bit će pridružena razredu pješaka čak i kada se pješaci uopće ne pojavljuju na slikama za učenje. To rezultira time da mreža „svugdje vidi pješake“. Negativna log izglednost ne uzima u obzir zastupljenost pojedinih razreda na slikama, što očekivano rezultira manjom točnošću razreda.

U slučaju primjene uravnoteženog gubitka, odgovarajućim odabirom parametara uvjetnog slučajnog polja može se u određenoj mjeri utjecati na točnost, odnosno točnost razreda (u općenitom slučaju smanjenjem točnosti razreda povećava se ukupna točnost).

U narednim tablicama (Tablica 10 - Tablica 15) prikazani su segmentacijski rezultati konvolucijske neuronske mreže (ekvivalent unarnih potencijala kod uvjetnog slučajnog polja) i uvjetnog slučajnog polja po razredima za različite funkcije gubitka

Retci predstavljaju stvarne razrede (ručno anotirane) dok stupci predstavljaju predikcije uvjetnog slučajnog polja. Tablice su po retcima normalizirane. Dijagonalni elementi predstavljaju točnost za pojedine razrede, dok preostali elementi predstavljaju udio pogrešno označenih piksela.

Tablica 10 Segmentacijski rezultati po razredima - negativna log izglednost – unarni potencijali

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	8,41	0,00	51,88	12,98	0,00	11,30	6,72	0,00	8,71	0,00	0,00
zgrada	0,00	79,07	0,00	16,87	2,42	0,00	0,14	0,82	0,00	0,68	0,00	0,00
pješak	0,00	22,17	0,00	18,97	58,75	0,00	0,00	0,00	0,00	0,11	0,00	0,00
vegetacija	0,00	4,17	0,00	94,68	0,05	0,00	0,07	0,85	0,00	0,17	0,00	0,00
vozilo	0,00	8,72	0,00	14,88	74,09	0,00	1,38	0,00	0,00	0,93	0,00	0,00
ograda	0,00	15,41	0,00	74,11	5,47	0,00	0,44	0,00	0,00	4,57	0,00	0,00
cesta	0,00	0,71	0,00	2,82	4,79	0,00	86,19	0,00	0,00	5,49	0,00	0,00
nebo	0,00	7,33	0,00	5,72	0,03	0,00	0,00	86,92	0,00	0,00	0,00	0,00
stup	0,00	31,75	0,00	60,41	4,42	0,00	0,67	0,60	0,00	2,15	0,00	0,00
pločnik	0,00	2,53	0,00	12,32	13,47	0,00	28,69	0,00	0,00	42,98	0,00	0,00
biciklist	0,00	57,28	0,00	37,64	5,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00
znak	0,00	0,11	0,00	40,98	55,09	0,00	3,22	0,00	0,00	0,58	0,00	0,00

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	8434	0	52005	13008	0	11324	6734	0	8729	0	0
zgrada	1370405	0	292407	41877	0	2444	14232	0	11721	0	0
pješak	2547	0	2180	6751	0	0	0	0	13	0	0
vegetacija	41367	0	938930	524	0	728	8451	0	1687	0	0
vozilo	76294	0	130114	647996	0	12057	0	0	8137	0	0
ograda	49603	0	238611	17613	0	1426	4	0	14722	0	0
cesta	5097	0	20270	34471	0	620456	0	0	39543	0	0
nebo	14239	0	11107	62	0	0	168886	0	0	0	0
stup	6825	0	12987	951	0	144	129	0	462	0	0
pločnik	9747	0	47400	51809	0	110339	0	0	165317	0	0
biciklist	3470	0	2280	308	0	0	0	0	0	0	0
znak	12	0	4296	5775	0	338	0	0	61	0	0

Tablica 11 Segmentacijski rezultati po razredima - negativna log izglednost – optimalni parametri

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	6,58	0,00	57,51	10,64	0,00	12,42	6,67	0,00	6,18	0,00	0,00
zgrada	0,00	81,41	0,00	15,71	1,97	0,00	0,09	0,46	0,00	0,37	0,00	0,00
pješak	0,00	18,14	0,00	27,28	53,88	0,00	0,69	0,00	0,00	0,01	0,00	0,00
vegetacija	0,00	2,20	0,00	97,02	0,03	0,00	0,05	0,62	0,00	0,09	0,00	0,00
vozilo	0,00	8,34	0,00	14,08	76,58	0,00	0,73	0,00	0,00	0,27	0,00	0,00
ograda	0,00	11,66	0,00	82,01	2,91	0,00	0,75	0,00	0,00	2,67	0,00	0,00
cesta	0,00	0,62	0,00	2,35	4,20	0,00	90,09	0,00	0,00	2,74	0,00	0,00
nebo	0,00	5,81	0,00	5,92	0,00	0,00	0,00	88,28	0,00	0,00	0,00	0,00
stup	0,00	31,39	0,00	63,18	2,02	0,00	0,33	0,67	0,00	2,40	0,00	0,00
pločnik	0,00	2,58	0,00	14,78	10,90	0,00	38,06	0,00	0,00	33,69	0,00	0,00
biciklist	0,00	59,46	0,00	36,99	3,55	0,00	0,00	0,00	0,00	0,00	0,00	0,00
znak	0,00	0,49	0,00	43,23	52,17	0,00	4,12	0,00	0,00	0,00	0,00	0,00

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	6592	0	57642	10667	0	12452	6683	0	6198	0	0
zgrada	1410820	0	272194	34145	0	1608	7896	0	6423	0	0
pješak	2085	0	3135	6191	0	79	0	0	1	0	0
vegetacija	21812	0	962096	281	0	452	6198	0	848	0	0
vozilo	72899	0	123146	669795	0	6399	1	0	2358	0	0
ograda	37549	0	264039	9374	0	2411	0	0	8606	0	0
cesta	4442	0	16916	30264	0	648524	0	0	19691	0	0
nebo	11281	0	11495	0	0	0	171518	0	0	0	0
stup	6749	0	13583	434	0	72	144	0	516	0	0
pločnik	9906	0	56827	41915	0	146376	0	0	129588	0	0
biciklist	3602	0	2241	215	0	0	0	0	0	0	0
znak	51	0	4531	5468	0	432	0	0	0	0	0

Tablica 12 Segmentacijski rezultati po razredima - Uravnoteženi gubitak – unarni potencijali

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	2,77	12,03	26,25	9,15	16,46	5,80	7,26	4,35	14,10	0,28	1,54
zgrada	0,00	60,58	14,65	6,38	1,90	4,89	0,05	2,55	4,82	1,16	1,95	1,07
pješak	0,00	4,09	10,23	6,16	43,91	9,54	0,00	0,00	8,93	0,30	1,20	15,65
vegetacija	0,00	2,39	7,66	84,15	0,20	1,09	0,00	1,35	1,86	0,27	0,44	0,58
vozilo	0,00	3,24	5,25	1,99	75,65	5,45	0,50	0,05	2,61	2,90	0,38	1,98
ograda	0,00	4,56	8,62	37,94	8,17	22,99	0,01	0,02	5,20	6,93	0,19	5,39
cesta	0,00	0,03	0,12	0,23	3,70	4,75	76,43	0,00	1,82	12,85	0,00	0,08
nebo	0,00	2,21	3,57	1,80	0,00	0,02	0,00	91,66	0,29	0,00	0,35	0,10
stup	0,00	2,15	10,83	16,73	0,24	4,76	0,00	0,96	54,83	0,85	7,58	1,06
pločnik	0,00	0,14	0,51	2,29	7,74	9,09	11,89	0,08	4,38	62,91	0,00	0,96
biciklist	0,00	3,60	4,64	2,67	4,01	0,00	0,00	0,00	4,24	0,00	76,71	4,13
znak	0,00	0,00	5,30	9,60	48,61	16,99	1,16	0,00	1,11	2,77	0,00	14,46

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	2780	12062	26314	9172	16503	5814	7272	4360	14132	284	1541
zgrada	1049839	253965	110614	33009	84738	812	44158	83532	20025	33780	18614
pješak	470	1175	708	5046	1096	0	0	1026	34	138	1798
vegetacija	23700	76008	834552	1982	10765	0	13404	18422	2723	4365	5766
vozilo	28316	45953	17387	661621	47642	4362	422	22861	25385	3300	17349
ograda	14683	27739	122147	26306	74016	17	62	16746	22300	623	17340
cesta	190	829	1635	26628	34190	550172	6	13120	92524	0	543
nebo	4285	6946	3507	5	34	0	178081	569	0	682	185
stup	463	2328	3597	52	1023	0	207	11788	183	1630	227
pločnik	550	1975	8805	29785	34976	45713	305	16862	241959	0	3682
biciklist	218	281	162	243	0	0	0	257	0	4647	250
znak	0	556	1006	5095	1781	122	0	116	290	0	1516

Tablica 13 Segmentacijski rezultati po razredima - Uravnoteženi gubitak - optimalni parametri

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	3,19	6,50	37,35	12,83	9,03	8,78	7,08	1,42	13,30	0,08	0,44
zgrada	0,00	70,41	9,53	8,44	1,81	4,75	0,02	1,31	1,88	1,04	0,59	0,22
pješak	0,00	5,74	4,29	15,03	44,42	11,17	0,06	0,00	2,22	2,68	0,39	13,99
vegetacija	0,00	1,48	2,65	92,07	0,10	1,00	0,00	1,00	1,06	0,33	0,06	0,25
vozilo	0,00	3,66	2,74	2,68	83,69	1,75	0,37	0,00	1,79	2,30	0,18	0,85
ograda	0,00	5,79	4,30	53,16	7,55	15,38	0,38	0,00	2,68	8,29	0,09	2,37
cesta	0,00	0,09	0,14	0,35	4,24	1,40	85,50	0,00	0,30	7,96	0,00	0,02
nebo	0,00	2,17	2,01	2,72	0,00	0,00	0,00	92,93	0,06	0,00	0,08	0,04
stup	0,00	8,22	6,29	30,98	0,24	3,14	0,03	0,87	40,48	2,70	6,60	0,44
pločnik	0,00	0,62	0,38	3,77	8,72	5,84	11,38	0,00	0,84	68,05	0,00	0,41
biciklist	0,00	14,96	2,23	7,56	4,28	0,02	0,00	0,00	2,26	0,00	67,02	1,68
znak	0,00	6,33	0,28	8,91	47,87	32,41	0,75	0,00	0,15	1,56	0,00	1,75

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	3202	6520	37436	12865	9050	8796	7092	1419	13331	79	444
zgrada	1220205	165199	146329	31394	82243	429	22621	32555	18043	10310	3758
pješak	660	493	1727	5104	1284	7	0	255	308	45	1608
vegetacija	14641	26306	913042	981	9903	37	9871	10552	3288	565	2501
vozilo	31983	23968	23413	731933	15326	3230	26	15652	20086	1554	7427
ograda	18637	13855	171171	24312	49506	1239	9	8640	26689	292	7629
cesta	631	1021	2488	30507	10105	615456	2	2157	57291	0	179
nebo	4218	3897	5280	5	3	0	180551	122	0	146	72
stup	1767	1353	6661	51	675	7	187	8703	580	1419	95
pločnik	2380	1452	14492	33541	22462	43758	2	3226	261739	0	1560
biciklist	906	135	458	259	1	0	0	137	0	4060	102
znak	663	29	934	5018	3397	79	0	16	163	0	183

Tablica 14 Segmentacijski rezultati po razredima - Uravnoteženi gubitak na cijelom skupu
– unarni potencijali

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	0,64	47,41	21,03	5,91	0,77	5,80	6,62	5,08	6,19	0,09	0,46
zgrada	0,00	29,69	52,61	7,00	0,84	0,56	0,04	1,47	4,11	0,38	2,54	0,77
pješak	0,00	0,06	51,28	0,14	32,29	0,00	0,00	0,00	5,03	0,00	5,28	5,92
vegetacija	0,00	0,27	29,96	67,63	0,03	0,10	0,00	0,47	0,89	0,00	0,55	0,09
vozilo	0,00	1,40	24,05	2,39	64,95	1,15	0,54	0,09	2,49	0,65	0,83	1,47
ograda	0,00	0,88	48,33	30,78	3,57	5,63	0,01	0,03	3,92	4,10	0,25	2,49
cesta	0,00	0,01	8,45	0,66	2,66	0,24	78,45	0,00	4,01	5,32	0,00	0,19
nebo	0,00	1,56	12,24	0,75	0,00	0,00	0,00	84,79	0,25	0,00	0,41	0,00
stup	0,00	0,28	27,19	9,88	0,00	0,09	0,00	0,10	53,07	0,25	8,39	0,75
pločnik	0,00	0,00	20,99	1,64	5,70	0,64	14,36	0,01	7,98	47,75	0,00	0,93
biciklist	0,00	0,00	23,51	0,50	2,91	0,00	0,00	0,00	3,76	0,00	69,25	0,08
znak	0,00	0,00	45,16	3,35	26,94	1,00	2,23	0,00	0,47	1,50	0,00	19,35

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	638	47520	21081	5921	773	5817	6639	5091	6204	90	460
zgrada	514580	911778	121234	14618	9634	672	25453	71197	6625	43984	13311
pješak	7	5893	16	3710	0	0	0	578	0	607	680
vegetacija	2712	297103	670682	317	978	0	4667	8854	49	5429	896
vozilo	12206	210348	20877	568072	10035	4756	756	21755	5667	7259	12867
ograda	2844	155605	99113	11506	18136	48	100	12607	13186	819	8015
cesta	48	60827	4782	19178	1751	564739	8	28854	38298	0	1352
nebo	3030	23779	1448	0	0	0	164744	490	0	802	1
stup	60	5845	2124	1	20	0	22	11408	54	1803	161
pločnik	18	80734	6311	21909	2476	55227	33	30678	183667	0	3559
biciklist	0	1424	30	176	0	0	0	228	0	4195	5
znak	0	4734	351	2824	105	234	0	49	157	0	2028

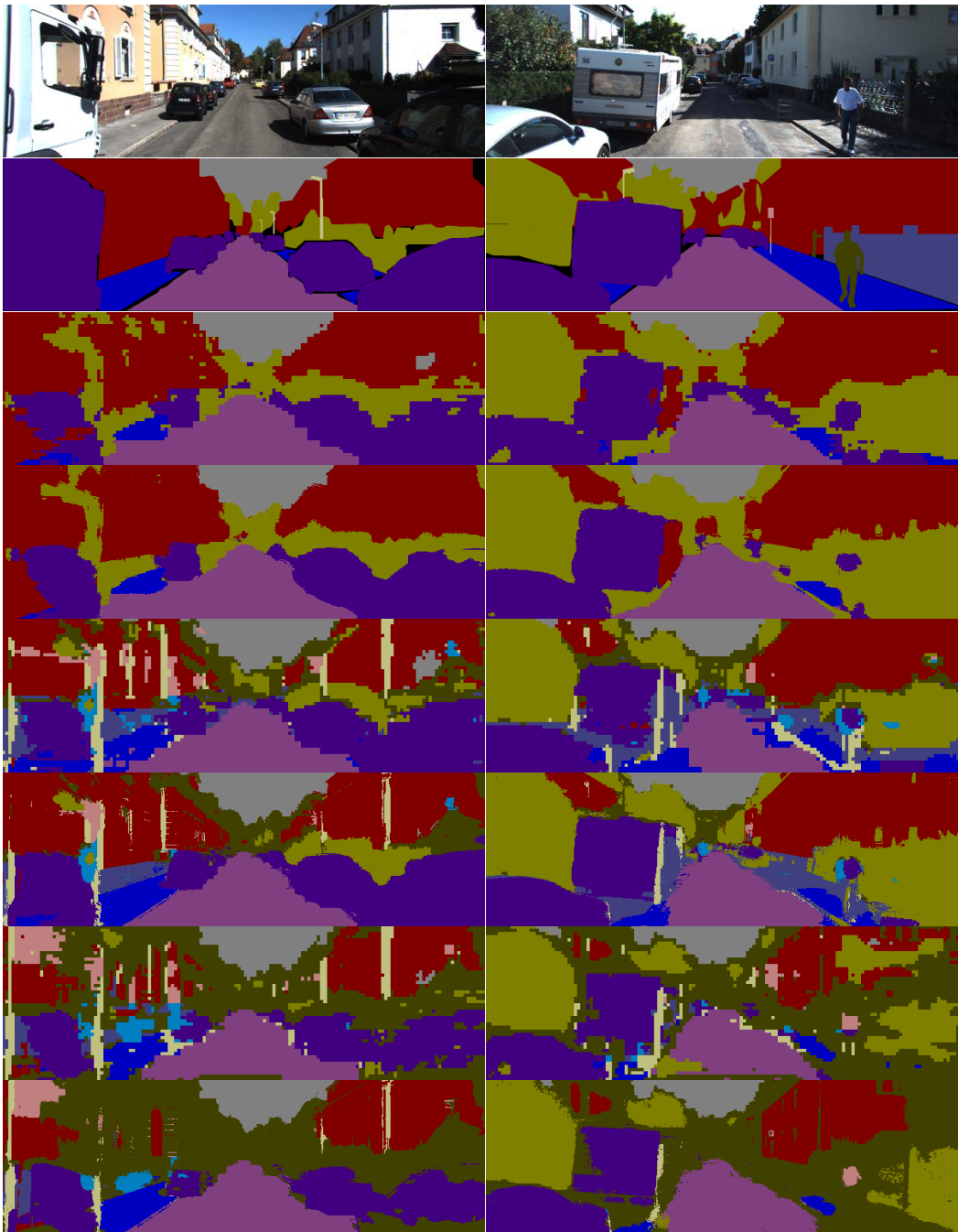
Tablica 15 Segmentacijski rezultati po razredima - Uravnoteženi gubitak na cijelom skupu
- optimalni parametri

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	0,92	44,18	25,32	6,21	0,15	9,01	6,67	0,85	6,50	0,04	0,15
zgrada	0,00	31,12	58,16	6,24	0,71	0,03	0,03	0,64	1,47	0,17	1,21	0,22
pješak	0,00	0,04	65,96	0,03	23,70	0,00	0,11	0,00	1,84	0,00	5,88	2,45
vegetacija	0,00	0,03	26,02	72,87	0,00	0,00	0,00	0,28	0,37	0,00	0,36	0,05
vozilo	0,00	1,26	23,68	1,53	70,55	0,53	0,50	0,00	0,63	0,54	0,39	0,40
ograda	0,00	0,33	54,86	35,10	2,64	1,49	0,01	0,00	0,92	3,97	0,07	0,62
cesta	0,00	0,01	4,44	0,13	3,20	0,03	87,18	0,00	0,66	4,28	0,00	0,07
nebo	0,00	1,74	9,79	0,99	0,00	0,00	0,00	87,40	0,07	0,00	0,03	0,00
stup	0,00	2,26	38,25	14,76	0,00	0,00	0,01	0,13	35,76	1,09	7,42	0,32
pločnik	0,00	0,12	21,87	1,31	5,62	0,01	16,95	0,00	1,15	52,76	0,00	0,21
biciklist	0,00	5,55	25,32	3,07	1,80	0,00	0,00	0,00	2,01	0,00	62,25	0,00
znak	0,00	0,00	71,04	0,01	28,11	0,00	0,43	0,00	0,00	0,00	0,00	0,41

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	920	44280	25381	6226	152	9035	6683	855	6516	38	148
zgrada	539356	1007979	108061	12239	552	543	11148	25458	2961	20997	3792
pješak	5	7579	3	2723	0	13	0	211	0	676	281
vegetacija	307	258069	722686	48	30	0	2753	3642	43	3602	507
vozilo	11001	207067	13356	617001	4679	4358	23	5522	4688	3370	3533
ograda	1049	176628	113006	8514	4791	37	0	2947	12785	229	1993
cesta	89	31975	928	23063	181	627589	6	4717	30788	0	501
nebo	3373	19020	1914	0	0	0	169805	132	0	50	0
stup	485	8223	3173	0	0	2	29	7687	235	1595	69
pločnik	444	84118	5036	21629	28	65191	2	4442	202933	0	789
biciklist	336	1534	186	109	0	0	0	122	0	3771	0
znak	0	7446	1	2947	0	45	0	0	0	0	43

Za potrebe analize utjecaja uvjetnog slučajnog polja na rezultate semantičke segmentacije, pogledajmo slučaj uravnoteženog gubitka (Tablica 12 i Tablica 13). Analogna analiza može se provesti i za ostale slučajeve. Promatrajući dijagonalne elemente u tablicama može se zaključiti kako CRF povećava točnost zastupljenijih klasa, dok smanjuje točnost manje zastupljenih klasa (pješak, ograda, stup, biciklist, znak). Slike dobivene sa CRF-om djeluju relativno lijepo i zaglađeno ljudskom oku, ali očito se to plaća sa određenim gubitkom točnosti piksela. Međutim dolazi do značajnog poboljšanja prosječne točnosti razreda (Tablica 8). Analogni zaključci vidljivi su i za slučaj negativne log izglednosti. Do smanjenja točnosti dolazi samo za slučaj pločnika kod CRF-a. Međutim, drugi problem predstavlja neoznačavanje ostalih slabo zastupljenih razreda, što nije bio slučaj za prethodno navedenu funkciju gubitka. Na temelju rezultata iz tablice (Tablica 13) vidljivo je kako se stvarni pješaci često označe kao vegetacija, vozila ili znakovi. CRF često razredom pješaka označi zgradu, odnosno stup. Analogna analiza može se provesti i za ostale razrede na temelju podataka u prethodnim tablicama, te nije detaljnije opisana.

Primjeri rezultata segmentacije prikazani su na slici (Slika 14). Odozgo prema dole prikazane su: ulazna slika, anotirana slika, slika unarnih potencijala, te izlazi CRF-a redom za negativnu log izglednost, uravnoteženi gubitak, te uravnoteženi gubitak definiran nad cijelim skupom. Na temelju provedenog eksperimenta može se zaključiti kako prikladno definirana funkcija gubitka može drastično poboljšati performanse na slabije zastupljenim razredima.



Slika 14 Rezultati segmentacije - KITTI. Odozgo prema dole prikazane su: ulazna slika, anotirana slika, slika unarnih potencijala, te izlazi CRF-a redom za negativnu log izglednost, uravnoteženi gubitak, te uravnoteženi gubitak definiran nad cijelim skupom.

5.5.3. Uvjetna slučajna polja kao rekurzivne neuronske mreže

Rekurzivna neuronska mreža definirana prema članku [2], naučena je na KITTI skupu podataka. Za početnu inicijalizaciju težina korištena je javno dostupna mreža FCN-8s [3]. Mreža je trenirana na slikama veličine 192x608 uz fiksnu stopu učenja od 10^{-13} . Naučena mreža ostvaruje rezultate prikazane u tablici (Tablica 16).

Tablica 16 Točnost RNN mreže na KITTI skupu podataka

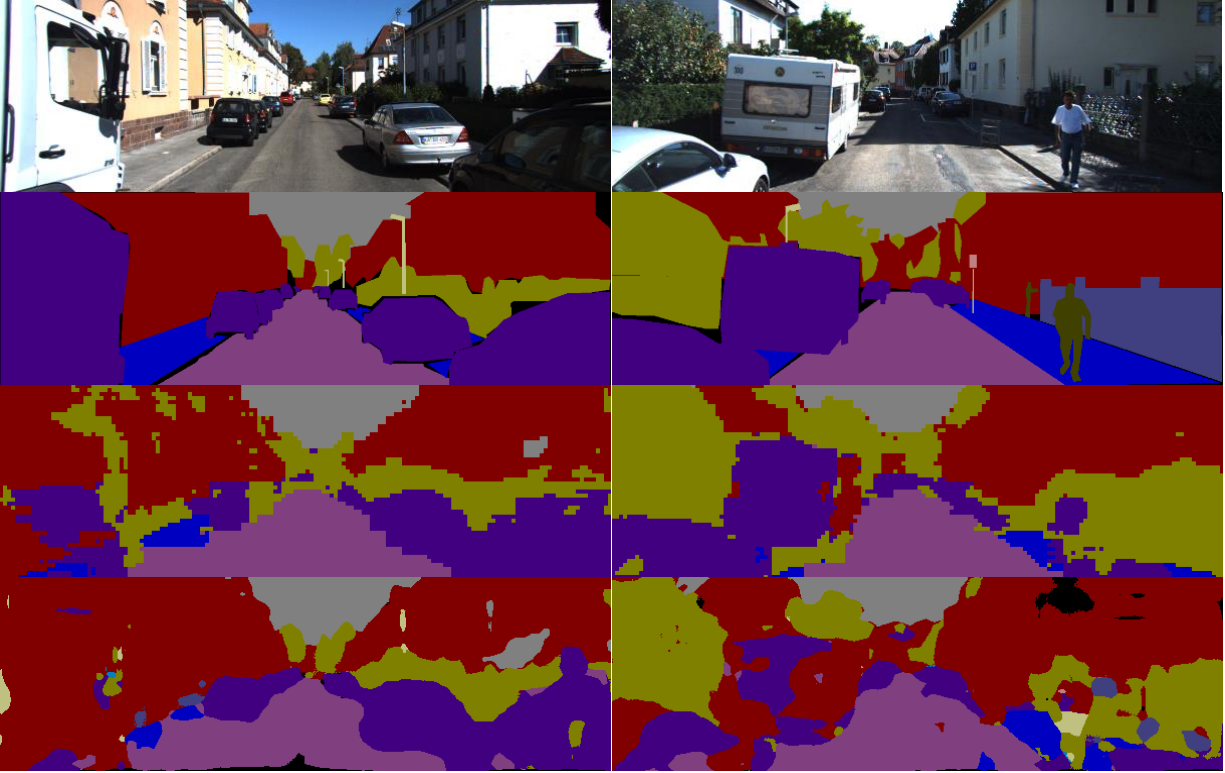
Točnost	76.440227 %
Točnost razreda	48.396190 %

Rezultati po razredima prikazani su u tablici (

Tablica 17). Slika 15 prikazuje primjere segmentiranih slika. Naučena mreža uspijeva dosta dobro označiti najzastupljenije razrede. Probleme joj stvaraju pješaci te prometni znakovi. Prilikom treniranja mreže trebalo bi razmisliti o definiranju drugačije funkcije gubitka, kako bi se uspjela podići točnost za slabo zastupljene razrede. Ako se promatra isključivo točnost piksela navedena mreža ostvaruje bolji rezultat nego mreža opisana u poglavlju 4.1.1. Glavna prednost ovakve mreže jest zajedničko učenje parametara uvjetnog slučajnog polja i konvolucijske neuronske mreže.

Tablica 17 Točnost po razredima za RNN

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	4,92	19,42	0,00	37,55	13,37	2,57	7,91	6,61	0,62	6,90	0,05	0,08
zgrada	2,37	88,25	0,00	5,17	1,38	0,81	0,21	1,18	0,42	0,18	0,01	0,03
pješak	0,03	40,20	0,26	13,44	33,46	0,27	0,00	0,00	11,73	0,02	0,40	0,20
vegetacija	0,05	7,66	0,00	90,35	0,30	0,10	0,10	1,20	0,23	0,00	0,00	0,00
vozilo	0,22	20,83	0,02	3,52	71,24	0,52	2,78	0,02	0,20	0,29	0,17	0,20
ograda	0,69	27,09	0,00	46,41	6,60	12,85	1,67	0,04	1,25	3,31	0,03	0,05
cesta	4,55	3,58	0,00	1,38	6,26	0,23	79,36	0,15	0,02	4,39	0,02	0,06
nebo	0,02	8,37	0,00	2,55	0,01	0,00	0,00	89,03	0,02	0,00	0,00	0,00
stup	0,42	36,23	0,00	32,87	0,97	4,44	0,03	0,80	23,68	0,11	0,44	0,00
pločnik	3,13	7,33	0,00	5,25	10,88	1,37	23,17	0,06	0,20	48,57	0,01	0,03
biciklist	0,00	55,91	0,00	12,64	1,62	0,00	0,00	0,13	7,63	0,00	22,07	0,00
znak	0,00	12,35	0,00	7,65	73,13	0,00	0,02	0,00	0,00	0,16	0,00	6,69



Slika 15 Uvjetno slučajno polje kao RNN. Odozgo prema dole: izvorna slika, anotirana slika, unarni potencijali, izlaz RNN mreže

5.5.4. Primjena naučenih značajki pri definiciji binarnih potencijala

U ovom poglavlju razmatra se alternativna definicija binarnih potencijala uvjetnog slučajnog polja. Prilikom definicije jezgri izgleda i zaglađivanja koriste se naučene značajke umjesto pozicija piksela u slici. Pritom je svaki piksel opisan sa 16 značajki koje su naučene konvolucijskom neuronskom mrežom prema radu [4] tako da maksimiziraju mjeru sličnosti između piksela. Učenje mreže i određivanje značajki za skup slika KITTI proveo je mag. ing. comp. Ivan Krešo.

Jezgre ovako definiranog uvjetnog slučajnog polja opisane su izrazima (46) i (47) pri čemu f_i predstavlja 16-dimenzionalni vektor koji opisuje i -ti piksel. I_i predstavlja svjetlosne intenzitete na poziciji i -tog piksela.

$$\text{jezgra zaglađivanja} = e^{\frac{-|f_i - f_j|^2}{2\theta_\gamma^2}} \quad (46)$$

$$\text{jezgra izgleda} = e^{-\frac{|f_i - f_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}} \quad (47)$$

Rezultati segmentacije prikazani su u tablici (Tablica 18). Za unarne potencijale korišteni su izlazi konvolucijske neuronske mreže trenirane s uravnoteženim gubitkom nad grupama slika. Optimalni parametri određeni su pretragom po rešetci (Tablica 19).

Tablica 18 Eksperimentalni rezultati za jezgre binarnih potencijala (46) i (47)

	Točnost	Točnost razreda
<i>Unarni potencijali</i>	68.493831 %	57.326605 %
<i>Optimalni parametri K</i>	73.394733 %	55.054725 %

Tablica 19 Optimalni parametri K za jezgre definirane prema (46) (47)

```
// jezgra zaglađivanja
s_gamma = 10
weight = 3

// jezgra izgleda
s_alpha = 70
s_beta = 3
weight = 2
```

Točnost po razredima prikazuju sljedeće tablice (Tablica 20, Tablica 12).

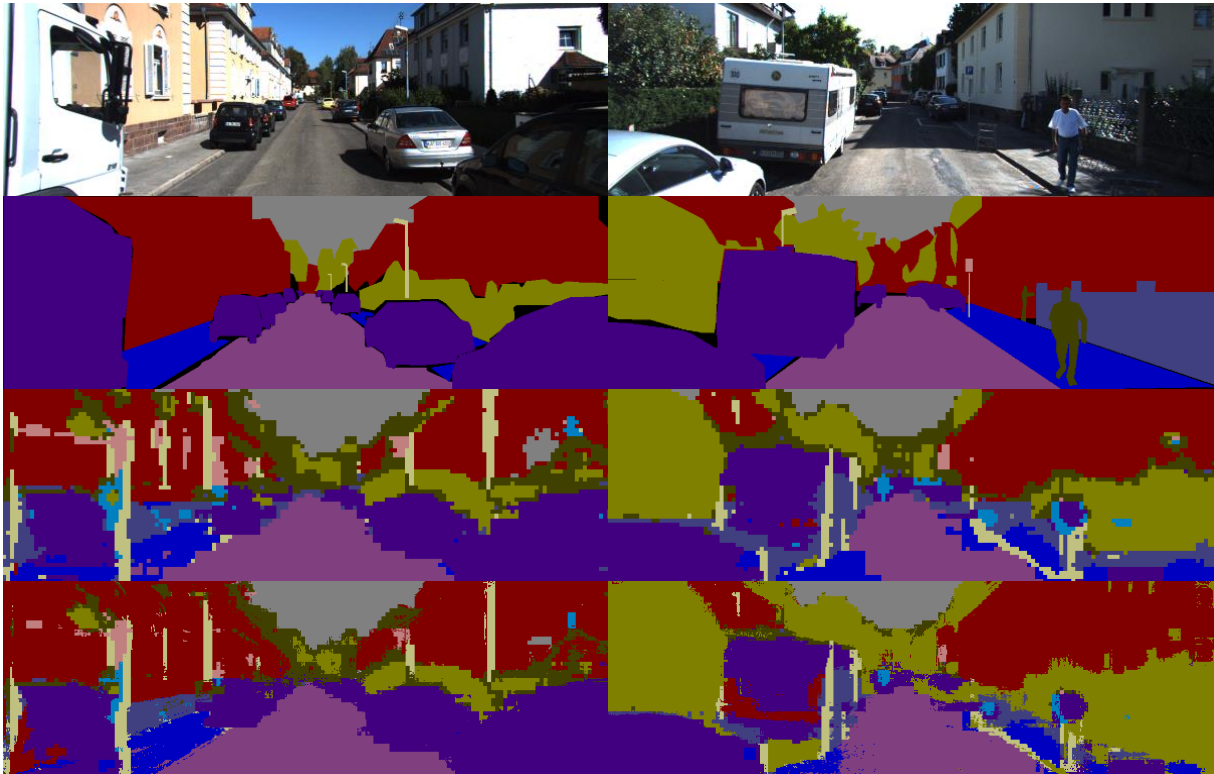
Analizom točnosti po razredima (Tablica 20) vidimo da i ovako definirani binarni potencijali još uvijek imaju problema s manje zastupljenim razredima čija se točnost u odnosu na unarne potencijale smanji.

Slika 16 prikazuje primjere segmentiranih slika za jezgre binarnih potencijala definirane prema (46) i (47). Odozgo prema dole prikazane su: ulazna slika, anotirana slika, slika unarnih potencijala, te izlazi CRF-a.

Tablica 20 Segmentacijski rezultati po razredima – jezgre binarnih potencijala (46) i (47) – optimalni parametri K

Ground truth / Inferred (%)	nepoznato	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	0,00	4,68	7,61	35,11	13,40	9,64	6,79	7,02	2,73	12,39	0,12	0,52
zgrada	0,00	70,45	6,98	11,36	2,25	2,77	0,10	1,34	2,74	0,79	0,79	0,44
pješak	0,00	9,84	4,52	14,36	49,96	5,08	0,00	0,00	4,85	0,23	0,85	10,31
vegetacija	0,00	2,25	2,87	91,31	0,20	0,42	0,00	1,00	1,27	0,30	0,14	0,23
vozilo	0,00	5,31	2,94	3,59	80,32	2,80	0,26	0,10	1,96	1,58	0,22	0,90
ograda	0,00	6,76	3,63	56,02	7,82	12,14	0,11	0,02	4,05	6,48	0,10	2,87
cesta	0,00	0,21	0,09	1,10	5,09	2,21	80,95	0,00	0,90	9,41	0,00	0,03
nebo	0,00	2,82	2,13	3,11	0,01	0,00	0,00	91,60	0,14	0,00	0,14	0,06
stup	0,00	8,33	6,70	24,78	0,61	3,84	0,00	0,69	47,76	1,20	5,41	0,68
pločnik	0,00	0,74	0,28	5,79	9,64	5,18	16,93	0,01	2,65	58,21	0,00	0,57
biciklist	0,00	14,31	2,43	7,59	3,98	0,00	0,00	0,00	3,00	0,00	65,00	3,68
znak	0,00	1,11	3,04	25,71	53,01	9,06	1,18	0,00	0,79	2,74	0,00	3,35

Ground truth / Inferred (broj piksela)	zgrada	pješak	vegetacija	vozilo	ograda	cesta	nebo	stup	pločnik	biciklist	znak
nepoznato	4688	7625	35189	13431	9662	6802	7038	2738	12423	119	519
zgrada	1220935	120930	196897	38961	47965	1728	23238	47488	13675	13658	7611
pješak	1131	519	1650	5741	584	0	0	557	26	98	1185
vegetacija	22332	28432	905495	1970	4209	6	9949	12599	2950	1423	2322
vozilo	46483	25738	31389	702500	24532	2263	889	17153	13806	1932	7913
ograda	21775	11674	180361	25191	39101	350	65	13053	20863	313	9233
cesta	1540	664	7942	36669	15890	582675	1	6467	67763	0	226
nebo	5470	4144	6043	18	2	0	177965	266	0	275	111
stup	1790	1441	5327	132	826	0	148	10267	258	1162	147
pločnik	2854	1069	22284	37074	19932	65108	26	10195	223879	0	2191
biciklist	867	147	460	241	0	0	0	182	0	3938	223
znak	116	319	2695	5557	950	124	0	83	287	0	351



Slika 16 Rezultati – jezgre binarnih potencijala definirane prema (46) (47). Odozgo prema dole prikazane su: ulazna slika, anotirana slika, slika unarnih potencijala, te izlazi CRF-a.

5.5.5. SegNet

S ciljem dobivanja unarnih potencijala učena je javno dostupna SegNet mreža [7]. Navedena mreža zanimljiva je zbog drugačijeg pristupa naglašavanju slabo zastupljenih razreda u odnosu na mrežu učenu u poglavlju 4.1.1. Naime, SegNet svaki razred u funkciji gubitka otežava s omjerom medijana frekvencija razreda i frekvencije promatranog razreda.

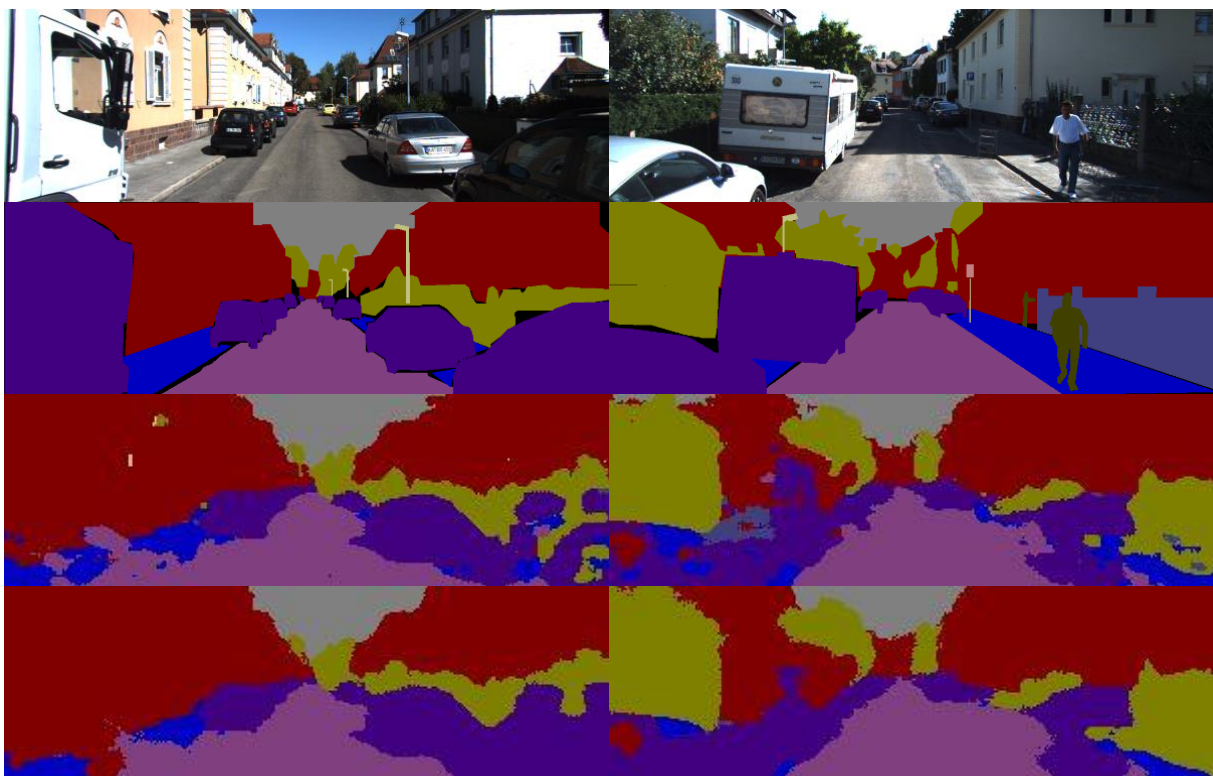
Mreža je trenirana na KITTI skupu slika. Međutim, zbog ograničenih računalnih resursa (GPU), slike su za potrebe treniranja smanjene usljed čega je u anotacijama nužno došlo do gubitka informacija.

Tablica 21 prikazuje ostvarene rezultate na KITTI skupu slika. Mreža sa uvjetnim slučajnim poljem ostvaruje porast točnosti piksela, međutim drastično poboljšanje točnosti razreda je izostalo. Razlog je što definirano otežavanje promatra cjelokupni skup podataka, pa se događa jednak efekt kao i kod uravnoteženog gubitka na cijelom skupu podataka iz poglavlja 4.1.1.

Tablica 21 Rezultati – Segnet unarni potencijali - KITTI

<i>Parametri</i>	<i>Segnet basic</i>	
	Točnost	Točnost razreda
CNN – uz ujednačavanje razreda	75.775556 %	44.434562 %
CNN – bez ujednačavanja razreda	74.540585 %	42.162338 %
CNN(sa ujednačavanjem) + CRF	77.748051 %	44.477265 %

Slika 17 prikazuje primjere segmentiranih slika za slučaj primjene ujednačavanja razreda.



Slika 17 Rezultati – SegNet – KITTI. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, izlaz CNN-a, izlaz CRF-a.

5.5.6. Konvolucijske značajke u binarnim potencijalima

U ovom poglavlju razmatra se definiranje binarnih potencijala uvjetnog slučajnog polja pomoću konvolucijskih značajki. Svi eksperimenti provode se nad skupom slika KITTI. Konvolucijske značajke izlučuju se iz mreže definirane u radu [1]. Izlaz posljednjeg konvolucijskog sloja daje 256 x 4 (rgbd) mape značajki, dakle 1024 značajke po pikselu. Ta brojka bila je prevelika za provedbu zaključivanja kod uvjetnog slučajnog polja pa se koristi PCA za preslikavanje vektora značajki u 40-dimenzijski prostor.

Pored standardnih jezgri zaglađivanja i izgleda, definiraju se dodatne jezgre (48)(49)(50).

$$S = e^{-\frac{|f_i - f_j|^2}{2\theta_\gamma^2}} \quad (48)$$

$$A = e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|f_i - f_j|^2}{2\theta_\beta^2}} \quad (49)$$

$$B = e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2} - \frac{|f_i - f_j|^2}{2\theta_\delta^2}} \quad (50)$$

f_i predstavlja vektor konvolucijskih značajki za i -ti piksel. Kod A jezgre radi se o trodimenzionalnom vektoru (veća dimenzija nije donijela nikakvo poboljšanje).

Tablica 22 i tablica 23 prikazuju rezultate dobivene uz primjenu jezgre S , zajedno s unarnim potencijalima dobivenim pomoću konvolucijske neuronske mreže [1] (učenom uz negativnu log izglednost) u ovisnosti o parametrima jezgre. Općenito, što se zada veća težina (w), veća važnost pridaje se binarnim konvolucijskim potencijalima, te se u konačnici dobiva slika označena sa najčešćom oznakom razreda.

Tablica 22 Točnost za promjenjivi parametar w

w	Točnost	Točnost razreda
1	74.244172 %	42.013998 %
3	72.770267 %	40.664243 %
5	68.704352 %	37.963380 %
Samo unarni potencijali	74.236634 %	42.176689 %

Tablica 23 Točnost za promjenjivi parametar theta

<i>theta</i>	<i>Točnost</i>	<i>Točnost razreda</i>
1	74.636435 %	42.456672 %
3	74.635292 %	42.447074 %
5	74.669254 %	42.435111 %
10	74.614252 %	42.216564 %
20	74.083214 %	41.567305 %
Samo unarni potencijali	74.236634 %	42.176689 %

Tablica 24 prikazuje rezultate za slučaj kada se uz jezgru *S* koriste i standardne jezgre zaglađivanja i izgleda. Težina *w* i parametar theta odabrani su jednostavnom pretragom po rešetci nad validacijskim skupom. Također su prikazani rezultati za slučaj kada se umjesto 40, koriste 3 konvolucijske značajke.

Napomena: prilikom provedbe eksperimenata vezanih uz binarne potencijale korišten je drugi naučeni model mreže od onog korištenog za dobivanje prethodnih rezultata (Tablica 8). Iz tog razloga točnosti koje se tiču izvorne definicije uvjetnog slučajnog polja razlikuju se u ovom poglavlju neznatno od prethodno iznesenih.

Tablica 24 Rezultati uz jezgru *S*

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
40 konv. značajki	76.215502 %	42.823425 %
3 konv. značajke	75.645639 %	42.032121 %
Kitti – bez konv.	76.206687 %	42.760374 %

Poboljšanje postoji, ali nije pretjerano značajno (barem ne za ukupnu točnost).

Tablica 25 A jezgra i unarni potencijali

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
Unarni + A jezgra	74.883684 %	42.359186 %
Samo unarni	74.236634 %	42.176689 %

Tablica 26 prikazuje rezultate primjene *A* jezgre zajedno s unarnim potencijalima, te standardnom jezgrom zaglađivanja.

Tablica 26 A jezgra zajedno s unarnim potencijalima, te standardnom jezgrom zaglađivanja.

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
3 konv. značajke	75.525889 %	42.638588 %
40 konv. značajki	75.368883 %	42.757923 %
Kitti – bez konv.	76.206687 %	42.760374 %

Tablica 27 prikazuje rezultate primjene A jezgre zajedno s unarnim potencijalima, te standardnim jezgrama zaglađivanja i izgleda.

Tablica 27 A jezgra zajedno s unarnim potencijalima, te standardnim jezgrama zaglađivanja i izgleda.

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
3 konv. značajke	76.128881 %	42.558755 %
40 konv. značajki	76.310821 %	42.885995 %
Kitti – bez konv.	76.206687 %	42.760374 %

U nastavku se promatra utjecaj B jezgre. Tablica 28 prikazuje rezultate primjene B jezgre te unarnih potencijala.

Tablica 28 B jezgra i unarni potencijali

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
Unarni + konv.	74.917182 %	42.514437 %
Samo unarni	74.236634 %	42.176689 %

Tablica 29 prikazuje rezultate primjene B jezgre zajedno sa standardnom jezgrom zaglađivanja.

Tablica 29 B jezgra zajedno sa standardnom jezgrom zaglađivanja

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
3 konv. značajke	75.496615 %	42.799239 %
35 konv. značajki	75.390349 %	42.778151 %
Kitti – bez konv.	76.206687 %	42.760374 %

Tablica 30 prikazuje rezultate primjene B jezgre zajedno sa standardnim jezgrama zaglađivanja i izgleda.

Tablica 30 B jezgra zajedno sa standardnim jezgrama zaglađivanja i izgleda

<i>Parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
3 konv. značajke	76.142772 %	42.731477 %
35 konv. značajki	76.281780 %	42.891924 %
Kitti – bez konv.	76.206687 %	42.760374 %

Analizom rezultata prikazanih u prethodnim tablicama, može se zaključiti kako proširenje klasičnih jezgara izgleda i zaglađivanja s konvolucijskim značajkama, ne dovodi do značajnijeg poboljšanja performansi. Problem predstavlja i velika dimenzionalnost vektora konvolucijskih značajki u zadnjim slojevima detektora značajki, što iziskuje primjenu analize glavnih komponenti. Pokazuje se kako konvolucijske značajke ne podnose najbolje PCA, te se velika količina korisne informacije, koju klasifikacijski dio neuronske mreže uspješno obrađuje, izgubi prilikom smanjenja dimenzionalnosti vektora konvolucijskih značajki.

5.5.7. Dubinska informacija u binarnim potencijalima

U ovom poglavlju razmatra se primjena informacije o dubini prilikom definiranja jezgri binarnih potencijala. Svi rezultati dobiveni su na skupu KITTI. Definiraju se dodatne jezgre binarnih potencijala (51)(52)(53)(54).

$$DS = e^{-\frac{|d_i - d_j|^2}{2\theta_\gamma^2}} \quad (51)$$

$$DA_1 = e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|d_i - d_j|^2}{2\theta_\beta^2}} \quad (52)$$

$$DA_2 = e^{-\frac{|d_i - d_j|^2}{2\theta_\alpha^2} - \frac{|l_i - l_j|^2}{2\theta_\beta^2}} \quad (53)$$

$$DA_3 = e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|l_i - l_j|^2}{2\theta_\beta^2} - \frac{|d_i - d_j|^2}{2\theta_\delta^2}} \quad (54)$$

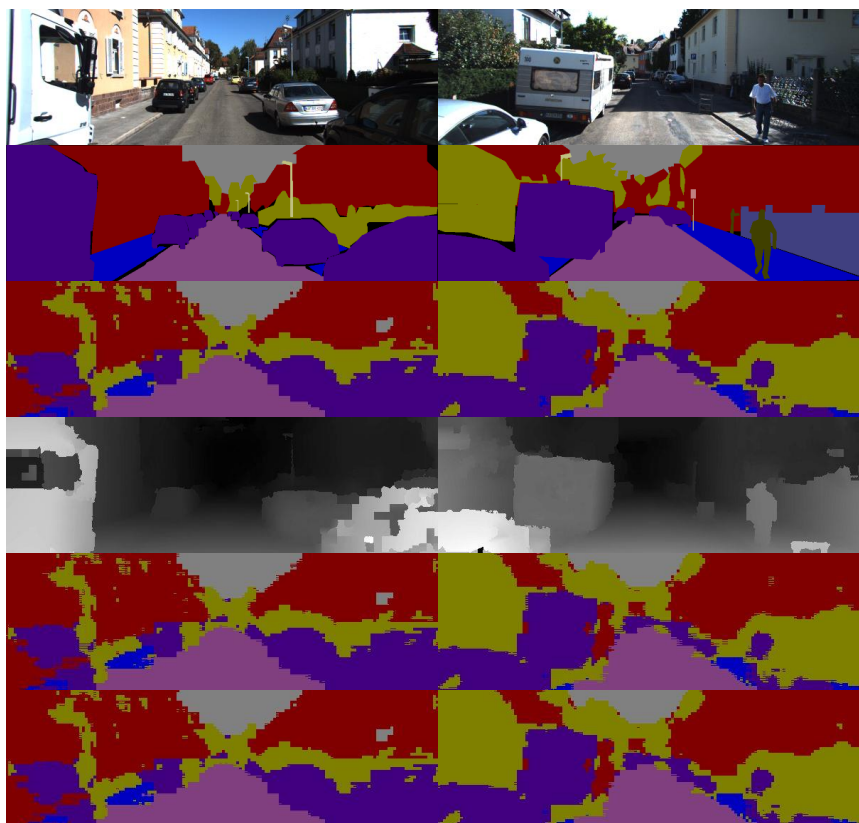
d_i predstavlja dubinu i -tog piksela.

Tablica 31 prikazuje utjecaj parametra theta (θ_γ) na točnost u slučaju primjene jezgre DS i unarnih potencijala.

Tablica 31 Jezgra DS i unarni potencijali - theta

<i>theta</i>	<i>Točnost</i>	<i>Točnost razreda</i>
1	74.568781 %	42.177644 %
2	74.565410 %	42.174357 %
3	74.547102 %	42.165857 %
5	74.517867 %	42.150678 %
10	74.451666 %	42.114772 %
20	74.378859 %	42.075338 %
Samo unarni potencijali	74.236634 %	42.176689 %

Slika 18 prikazuje izlaze CRF-a za različite vrijednosti parametra theta. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, unarni potencijali, dubinska mapa, izlazi za $\theta \in \{1,10\}$.



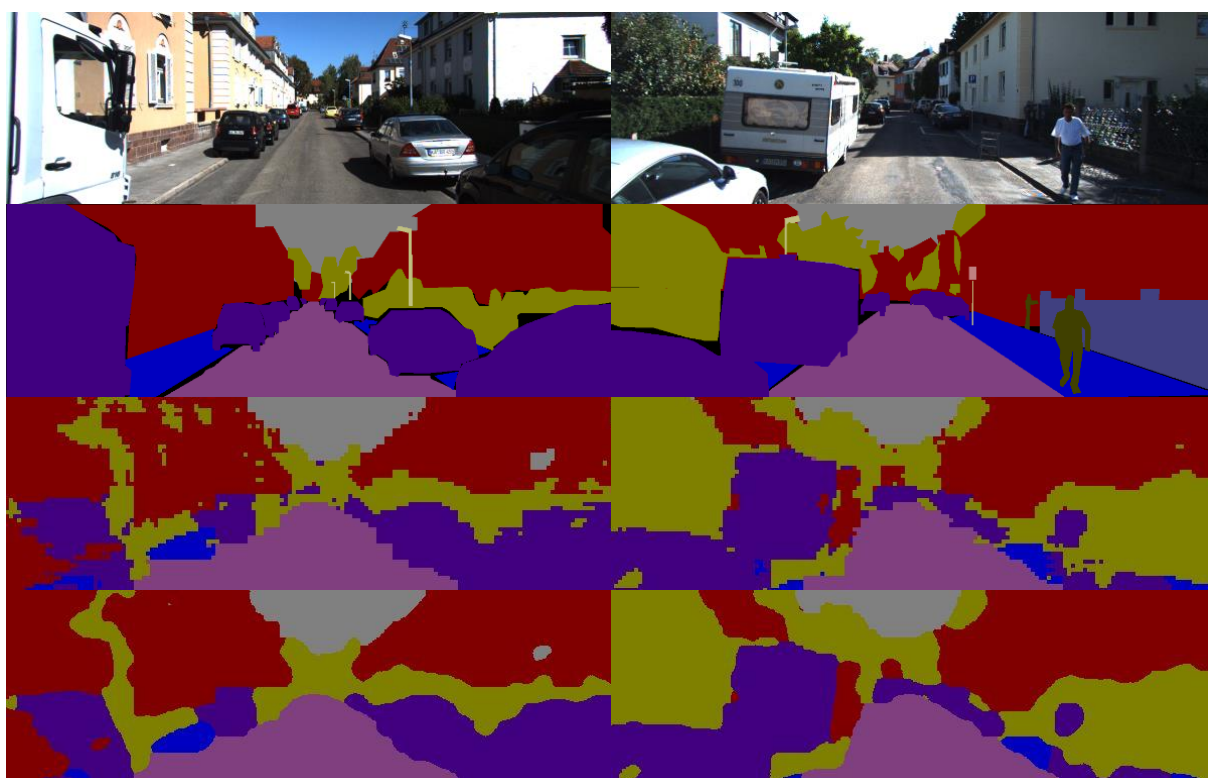
Slika 18 Primjena DS jezgri te unarnih potencijala – theta. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, unarni potencijali, dubinska mapa, izlazi za $\theta \in \{1,10\}$.

Tablica 32 prikazuje rezultate primjene DS jezgre zajedno sa standardnom jezgrom zaglađivanja, te unarnim potencijalima za parametre 1 i parametre 2.

Tablica 32 DS jezgra i jezgra zaglađivanja

<i>parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
DS + jezgra zaglađivanja	75.412591 %	42.544347 %
Samo unarni	74.236634 %	42.176689 %

Slika 19 prikazuje izlaze CRF-a. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, izlaz za unarne potencijale, izlaz za unarne potencijale + DS jezgru + jezgru zaglađivanja



Slika 19 DS jezgra i jezgra zaglađivanja. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, izlaz za unarne potencijale, izlaz za unarne potencijale + DS jezgru + jezgru zaglađivanja – najprije za parametre 1, a potom za parametre 2.

U naredne tri tablice (Tablica 33, Tablica 34, Tablica 35) parametri jezgri držani su fiksnima te se promatralo uz koje bi značajke imalo smisla kombinirati dubinu. Na temelju dobivenih rezultata, najviše smisla ima uzeti pozicije piksela i dubinu („bliske piksele na istoj dubini označi jednako“).

Tablica 33 DA₁ jezgra i unarni potencijali

<i>parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
DA ₁	75.389265 %	42.760510 %
Samo unarni	74.236634 %	42.176689 %

Tablica 34 DA₂ jezgra i unarni potencijali

<i>parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
DA ₂	73.951239 %	41.271207 %
Samo unarni	74.236634 %	42.176689 %

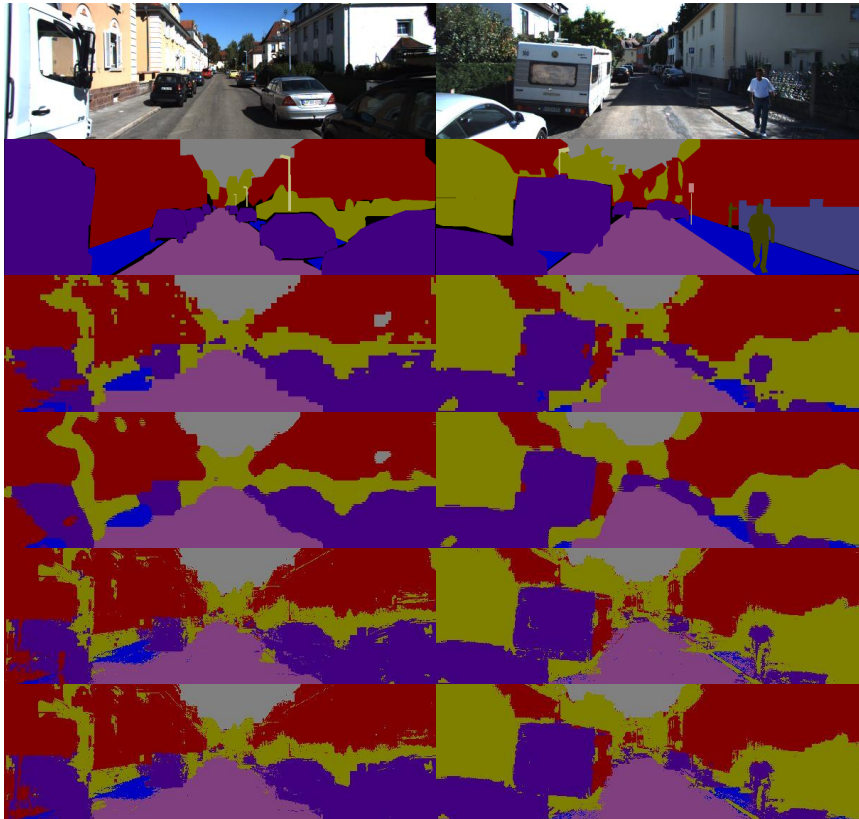
Tablica 35 DA₃ jezgra i unarni potencijali

<i>parametri</i>	<i>Točnost</i>	<i>Točnost razreda</i>
DA ₃	75.003260 %	42.071852 %
Samo unarni	74.236634 %	42.176689 %

Slika 20 prikazuje odgovarajuće izlaze CRF-a uz primjenu prethodno definiranih jezgri.

Odozgo prema dole prikazane su:

- Izvorna slika
- Anotirana slika
- Izlaz za unarne potencijale
- unarni +
 - DA₁ jezgra
 - DA₂ jezgra
 - DA₃ jezgra



Slika 20 DA jezgra i unarni potencijali. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, unarni potencijali, unarni potencijali s DA_1 jezgrom, unarni potencijali s DA_2 jezgrom, unarni potencijali s DA_3

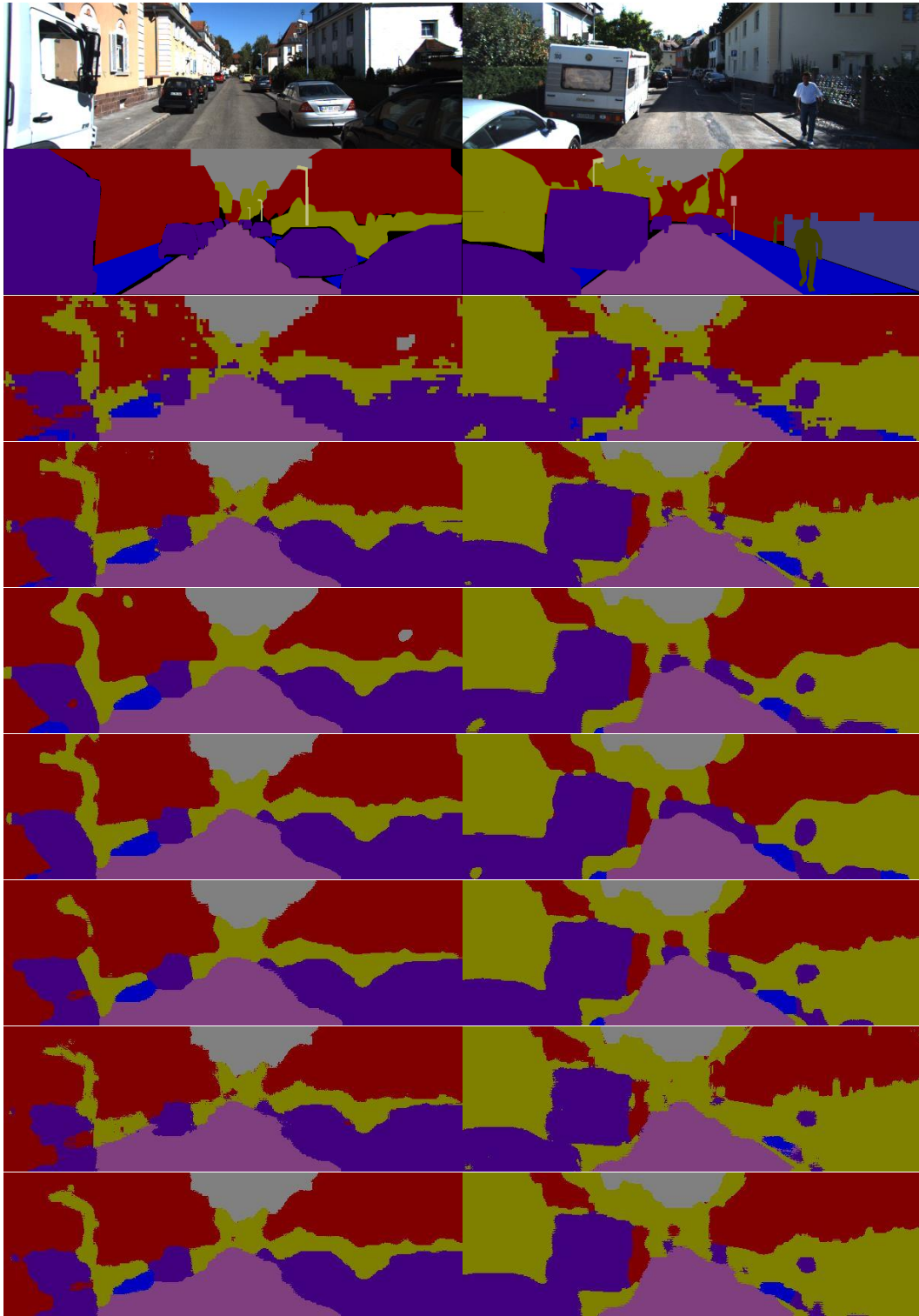
Tablica 36 prikazuje točnost segmentacije za različite kombinacije prethodno definiranih jezgara. Označimo sa U unarne potencijale, te sa $S(A)$ jezgru zaglađivanja (izgleda).

Tablica 36 Točnost za različite kombinacije korištenih jezgri

Parametri	Točnost	Točnost razreda
S + DA_1	75.459921 %	42.761453 %
S + DA_1 + A	76.481950 %	43.053443 %
DS + S + DA_1	74.768216 %	41.599739 %
DS + S + A	74.687989 %	41.042067 %
DS + S + DA_1 + A	75.736405 %	42.061556 %
Kitti – S + A	76.206687 %	42.760374 %

Slika 21 prikazuje rezultate segmentacije za različite kombinacije korištenih jezgri. Odozgo prema dole prikazane su:

- Izvorna slika
- Anotirana slika
- Izlaz za unarne potencijale
- unarni +
 - $S + A$
 - $S + DA_1$
 - $S + DA_1 + A$
 - $DS + S + DA_1$
 - $DS + S + A$
 - $DS + S + DA_1 + A$



Slika 21 Izlazi CRF-a za različite kombinacije korištenih jezgri. Odozgo prema dole prikazane su: izvorna slika, anotirana slika, unarni potencijali, unarni potencijali + S + A, unarni potencijali + S + DA₁, unarni potencijali + S + DA₁ + A, unarni potencijali + DS + S + DA₁, unarni potencijali + DS + S + A, unarni potencijali + DS + S + DA₁ + A

6. Zaključak

Semantička segmentacija predstavlja veoma zanimljivo i zahtjevno područje računalnog vida u kojem su grafički modeli, a prvenstveno uvjetna slučajna polja, našla široku primjenu.

Pokazano je kako klasična definicija uvjetnih slučajnih polja nema pretjeranu primjenu u rješavanju složenijih problema poput semantičke segmentacije zbog računski skupog zaključivanja. Međutim, definiranje binarnih potencijala u obliku Gaussovih funkcija, zajedno s metodom srednjeg polja omogućilo je brzu izvedbu zaključivanja kod gusto povezanih uvjetnih slučajnih polja.

Uvjetna slučajna polja, pored binarnih definirana su i pomoću unarnih potencijala. Eksperimentalno je istaknuta njihova važnost za semantičku segmentaciju. Prikladni unarni potencijali mogu se naučiti pomoću konvolucijskih neuronskih mreža. Za potrebe semantičke segmentacije problematični su slabo zastupljeni razredi u javno dostupnim skupovima slika budući da je neuronskoj mreži inicijalno teško naučiti takve razrede. Međutim, problem se može riješiti umjetnim povećavanjem gradijenata za slabo zastupljene klase prilikom učenja neuronske mreže što se efektivno može postići pogodnom definicijom funkcije gubitka. Također je demonstriran utjecaj dubine arhitekture mreže, te razmatranja višestrukih rezolucija ulazne slike na poboljšanje rezultata semantičke segmentacije.

U radu su razmatrane i potencijalne definicije novih Gaussovih jezgri. Pritom su u obzir uzete konvolucijske značajke i dubinska informacija piksela. Međutim, pokazalo se iznimno teško definirati binarne potencijale koji bi bili bolji od jednostavnih jezgri izgleda i glađenja nad RGB vrijednostima piksela slike. Definitivno se može zaključiti kako konvolucijske značajke nema pretjeranog smisla koristiti prilikom definicije binarnih potencijala kako zbog njihove velike dimenzionalnosti, tako i zbog slabog podnošenja analize glavnih komponenata. Dubinska informacija, za razliku od konvolucijskih značajki pokazala se uspješnom u povećanju performansi semantičke segmentacije.

Potencijalni budući rad uključivao bi eksperimentiranje s kombiniranjem odluka više različitih mreža (ili pak kombiniranje odluka neuronskih mreža i modela izgrađenih drugim postupcima) kako bi se dobili kvalitetniji unarni potencijali. Pored toga moguće je proširiti definiciju uvjetnog slučajnog polja s potencijalima višeg reda, kao i učiti

binarne potencijale neuronskim mrežama čime se dodatno mogu poboljšati performanse uvjetnog slučajnog polja.

Zahvale

Zahvaljujem se mentoru, izv. prof. dr. sc. Siniši Šegviću na kontinuiranoj podršci i pomoći pri izvedbi ovog rada. Također se zahvaljujem i mag. ing. Ivanu Kreši na izvedbi konvolucijske mreže za definiranje unarnih potencijala, dr. sc. Josipu Krapcu na pruženim savjetima i idejama tijekom izrade rada, kao i mag. ing Ivanu Borku na pomoći prilikom generiranja unarnih potencijala konvolucijskom mrežom Farabeta i LeCuna [9].

Literatura

- [1] Borko, I. Semantička segmentacija prirodnih scena dubokim neuronskim mrežama. Diplomski rad. Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2015.
- [2] Shuai, Z., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P. H. S. Conditional Random Fields as Recurrent Neural Networks. IEEE International Conference on Computer Vision, 2015.
- [3] Long, J., Shelhamer, E. Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), stranice 3431-3440, 2015.
- [4] Žbontar, J., LeCun, Y. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. arXiv preprint arXiv:1510.05970, 2015.
- [5] Krähenbühl, P., Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. Neural Information Processing Systems, 2011.
- [6] Badrinarayanan, V., Kendall, A., Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv preprint arXiv:1511.00561, 2015.
- [7] Kendall, A. Caffe-SegNet, <https://github.com/alexgkendall/caffe-segnet>, 21.1.2016.
- [8] Adams, A., Baek, J., Davis, M. A. Fast High-Dimensional Filtering Using the Permutohedral Lattice. Computer Graphics Forum, 2010.
- [9] Farabet, C., Couprie, C., Najman, L., LeCun, Y. Learning Hierarchical Features for Scene Labeling. IEEE Trans. Pattern Anal. Mach. Intell., 35(8):1915-1929, 2013.
- [10] McCallum, A. Efficiently inducing features of conditional random fields. Conference on Uncertainty in AI (UAI), 2003.
- [11] Wick, M., Rohanimanesh, K., McCallum, A., Doan, A. A discriminative approach to ontology alignment. International Workshop on New Trends in Information Integration (NTII), 2008.
- [12] Roth, D., Yih, W. Integer linear programming inference for conditional random fields. International Conference on Machine Learning (ICML), stranice 737-744, 2005.
- [13] Kudo, T., Yamamoto, K., Matsumoto, Y. Applying conditional random fields to

- Japanese morphological analysis. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004.
- [14] Quattoni, A., Collins, M., Darrell, T. Conditional random fields for object recognition. Advances in Neural Information Processing Systems, stranice 1097-1104, 2004.
- [15] Krizhevsky, A., Sutskever, I., Hinton, G.E. Imagenet classification with deep convolutional neural networks. Annual Conference on Neural Information Processing Systems, stranice 1106-1114, 2012.
- [16] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. International Conference on Learning Representations (ICLR), 2014.
- [17] Lin, G., Shen, C., van den Hengel, A., Reid, I. Efficient piecewise training of deep structured models for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [18] Arnab, A., Jayasumana, S., Zheng, S., Torr, P.H.S. Higher order potentials in end-to-end trainable conditional random fields. CoRR abs/1511.08119, 2015.
- [19] Lafferty, J., McCallum, A., Pereira, F.C.N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data, 2001.
- [20] Szeliski, R. Computer Vision: Algorithms and Applications. Springer, 2010.
- [21] Rother, C. MRFs and CRFs for Vision: Models & Optimization, http://www.di.ens.fr/willow/events/cvml2010/materials/INRIA_summer_school_2010_Carsten.pdf, 21.4.2016.
- [22] Krešo, I., Čaušević, D., Krapac, J., Šegvić, S. Convolutional scale invariance for semantic segmentation. U postupku recenzije (GCPR), 2016.
- [23] Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), stranice 1-16, 2014.
- [24] Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B. The cityscapes dataset. CVPR Workshop on The Future of Datasets in Vision, 2015.
- [25] Geiger, A., Lenz, P., Stiller, C., Urtasun, R. Vision meets robotics: The kitti dataset. International Journal of Robotics Research (IJRR), 2013.
- [26] Ros, G., Ramos, S., Granados, M., Bakhtiary, A., Vazquez, D., Lopez, A.M. Vision-based offline-online perception paradigm for autonomous driving, IEEE

Winter Conference on Applications of Computer Vision (WACV), stranice 231-238, 2015.

- [27] Gould, S., Fulton, R., Koller, D. Decomposing a Scene into Geometric and Semantically Consistent Regions. Proceedings of International Conference on Computer Vision (ICCV), 2009.
- [28] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A. The PASCAL Visual Object Classes (VOC) challenge, IJCV, 2010.
- [29] German, S., German, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions, stranice 721-741, 1984.
- [30] Eigen, D., Fergus, R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. IEEE International Conference on Computer Vision (ICCV), stranice 2650-2658, 2015.

Sažetak

Primjena uvjetnih slučajnih polja s Gausovim potencijalima za semantičku segmentaciju

Denis Čaušević

Uvjetna slučajna polja su diskriminativni probabilistički grafički model strojnog učenja s brojnim primjenama u klasifikaciji strukturiranih podataka. Problem zaključivanja kod uvjetnih slučajnih polja svodi se na maksimizaciju aposteriorne vjerojatnosti, što je ekvivalentno minimizaciji energije definirane potencijalima prvog, drugog ili višeg reda. U ovom radu razmatramo binarne potencijale modelirane Gausovim funkcijama zbog mogućnosti modeliranja interakcije među prostorno udaljenim čvorovima grafa i efikasnog zaključivanja metodom srednjeg polja. Ovako definirana slučajna polja primijenili smo za poboljšanje rezultata semantičke segmentacije dobivenih dubokim konvolucijskim neuronskim mrežama. Eksperimenti su provedeni na standardnim skupovima podataka u području razumijevanja urbanih prometnih scena (KITTI, Cityscapes). Postignuti rezultati jasno potvrđuju potencijal ovog pristupa za poboljšanje segmentacijske performanse modernih dubokih modela. Ostvarena performansa vrlo je blizu najboljih svjetskih rezultata u području semantičke segmentacije urbanih scena.

Ključne riječi: uvjetna slučajna polja, metoda srednjeg polja, semantička segmentacija, konvolucijske neuronske mreže, probabilistički grafički modeli

Summary

Application of conditional random fields with Gaussian potentials for semantic segmentation

Denis Čaušević

Conditional random fields are probabilistic discriminative graphical machine learning models with a wide variety of applications for structured prediction. The inference problem in conditional random fields can be formulated as posterior probability maximization, which is equivalent to energy minimization defined on first, second and higher order potentials. In this paper we consider conditional random fields with Gaussian functions as binary potentials. These potentials enable modelling of interactions between spatially distant graph nodes and allow efficient inference through mean field approximation. This paper elaborates the application of such an approach for semantic segmentation, while combining the before mentioned binary potentials with convolutional unary potentials. Experiments done on publicly available datasets for urban traffic scene understanding (Cityscapes, KITTI) prove that this approach has a high potential to further improve segmentation performance of modern deep models. The performance achieved in this paper is very close to state-of-the-art results for semantic segmentation of urban scenes.

Keywords: conditional random fields, mean field approximation, semantic segmentation, convolutional neural networks, probabilistic graphical models