

Semantička segmentacija prirodnih scena dubokim neuronskim mrežama

Ivan Borko

Fakultet elektrotehnike i računarstva

ivan.borko@fer.hr

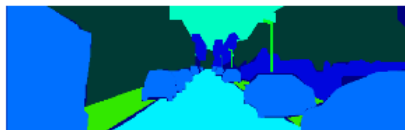
16. srpnja 2015.

- 1 Semantička segmentacija
- 2 Konvolucijske neuronske mreže
- 3 Implementacija
- 4 Testni skupovi
- 5 Rezultati





Semantička segmentacija

- pridjeljivanje labela svakom pikselu
- spaja klasične probleme detekcije objekata, segmentacije slika i višeklasne klasifikacije
- labela - oznake objekata predstavljenih slikom
- primjeri labela:
 - vegetacija
 - vozilo
 - građevina
 - cesta
 - osoba
 - životinja

Semantička segmentacija - primjer



-  vozilo
-  cesta
-  građevina
-  pločnik

-  nebo
-  vegetacija
-  prometni znak
-  pješak

Slika: Primjer semantičke segmentacije slike

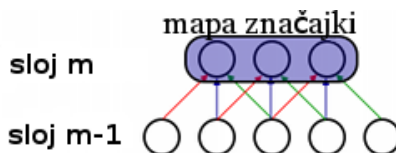
Konvolucijske neuronske mreže

Konvolucijske mreže - motivacija

- spadaju u duboke neuronske mreže - više od dva skrivena sloja
- računalni vid → veliki ulazni prostor
- veliki broj parametara
 - sporo učenje
 - prenaučenosť
- koncepti:
 - konvolucijski sloj
 - operator sažimanja

Konvolucijski sloj

- analiza vidnog korteksa mačaka i ljudi:
 - jednostavne stanice - reaguju na lokalne podržaje
 - kompleksne - veća područja i invarijantne na poziciju
- lokalna osjetljivost - detekcija objekata i oblika neovisno o poziciji
- konvolucijski operator - 1D konvolucija za zvuk, 2D za slike

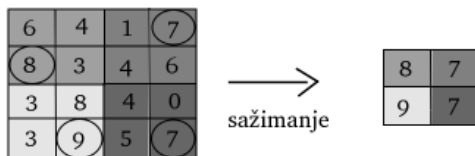


Slika: Zajedničke težine

Sloj sažimanja

(*engl.* pooling)

- vrsta nelinearnog poduzorkovanja
- podijeli ulaznu sliku na više nepreklapajućih pravokutnika
- više pristupa
 - sažimanje usprosjecivanjem (*engl. mean pooling*)
 - sažimanje maksimalnog odziva (*engl. max pooling*)
- važnost
 - smanjuje računsku složenost za gornje slojeve
 - povećava neosjetljivost na translacije u slici



Slika: Primjer sažimanja maksimalnog odziva

Implementacija

- alat korišten za programsku implementaciju
- služi za definiranje, optimiranje i prevođenje simboličkih izraza u C++ ili CUDA kod
- prednosti:
 - autodiferenciranje - automatsko računanje gradijenata
 - paralelno korištenje CPU i GPU (prototipiranje / evaluacija)
 - aritmetička pojednostavljenja, na primjer: $(x \cdot y)/x \rightarrow y$
 - ugrađene metode za poboljšanje numeričke stabilnosti određenih matematičkih izraza
- mane:
 - težak za učenje
 - otklanjanje grešaka

Isprobano više metoda:

- YUV kanali umjesto RGB
- Laplaceova piramida
- normalizacija vs
normalizacija blokova



Slika: Primjer Laplaceove piramide

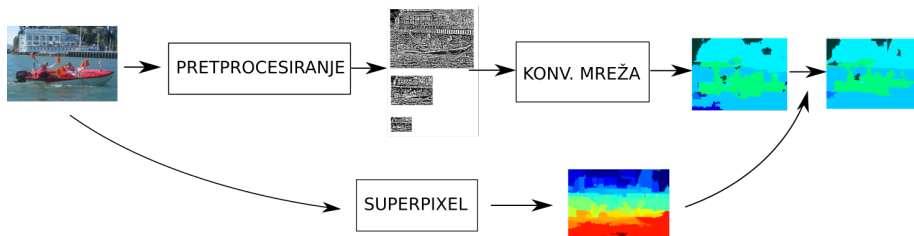
Postprocesiranje

- koristi se metoda superpiksela - grupiranje piksela po boji ulazne slike
- podešavanje triju parametara

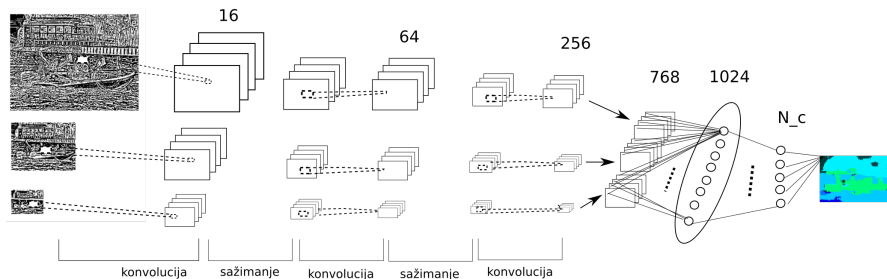


Slika: Segmentacija superpikselima

Arhitektura sustava



Arhitektura mreže



- *leakyReLU* aktivacijska fukcija
- 3 razine: svaka razina ima 3 sloja koja dijele filtere od 3 konvolucijska sloja
- na kraju je potpuno povezani sloj i sloj multinomijalne regresije koji vrši klasifikaciju određenog piksela

Negativna log izglednost

$$nll = -\frac{1}{N} \sum_{i=1}^N \ln P(Y_i = c_i)$$

Bayesova log izglednost - balansiranje razreda

$$bayesian_nll = -\frac{1}{N} \sum_{i=1}^N \frac{1}{P_{apr}(c_i)} \ln P(Y_i = c_i)$$

N je broj primjera u skupu za učenje, $P_{apr}(c_i)$ je apriorna vjerojatnost klase c_i , Y_i je izlaz mreže.

Testni skupovi

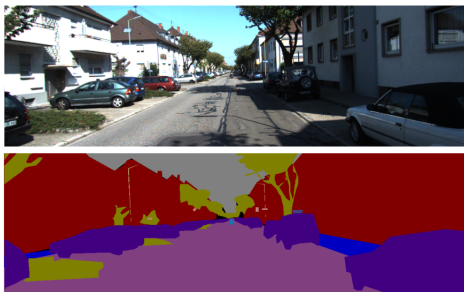
Stanford Background Dataset

- slike vanjskih scena, sakupljene iz raznih skupova podataka
- veličina 320 x 240 piksela
- 715 slika
- 8 semantičkih oznaka



Slika: Primjer slike i oznake iz skupa Stanford Background

- slike snimljene kamerom montiranom na vozilo, veličine 1241 x 376 piksela
- RGB, ali i dubinska komponenta
- 146 označenih: 100 za učenje, 46 za testiranje
- 12 semantičkih oznaka



Slika: Primjer slike i oznake iz skupa KITTI

Rezultati

Rezultati - Stanford Background

Metoda	Funkcija troška	Točnost(%)	Točnost razreda(%)	Brzina (sec)
Konv. mreža s 3 razine		75.7	59.3	0.05
	negativna log izglednost			
Konv. mreža s 3 razine + <i>superpixels</i>		76.1	59.7	0.11
	negativna log izglednost			
Konv. mreža s 3 razine		71.4	64.8	0.05
	Bayesova log izglednost			
Konv. mreža s 3 razine + <i>superpixels</i>		74.2	68.1	0.11
	Bayesova log izglednost			
Farabet et al. 2013		78.8	72.4	0.6
Farabet et al. 2013 + <i>superpixels</i>		80.4	74.6	0.7
Lempitzky et al. 2011		81.9	72.4	60
Munoz et al. 2010		76.9	66.2	12

Tablica: Rezultati na Stanford Background skupu podataka

Metoda	Funkcija troška	Točnost(%)	Točnost razreda(%)	Brzina (sec)
Konv. mreža s 3 razine (RGB)	negativna log izglednost	73.6	42.4	0.05
Konv. mreža s 3 razine (RGB) + <i>superpixels</i>	negativna log izglednost	75.1	43.1	0.11
Konv. mreža s 3 razine (RGB)	Bayesova log izglednost	69.8	55.3	0.05
Konv. mreža s 3 razine (RGBD)	negativna log izglednost	78.5	45.8	0.05
Konv. mreža s 3 razine (RGBD) + <i>superpixels</i>	negativna log izglednost	79.1	46.1	0.11
Ros et al. 2015		81.1	58.0	0.46

Tablica: Rezultati na KITTI skupu podataka

U okviru ovog rada implementiran je sustav koji postiže rezultate usporedive s najboljim implementacijama. Konvolucijske neuronske mreže su se pokazale kao dobar model za semantičku segmentaciju.

The End

Neuronske mreže - uvod

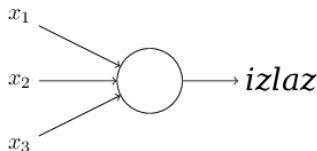
umjetni neuroni: perceptron

- linearni klasifikator
- aktivacijska funkcija: logistički sigmoid

Izlaz

$$P(Y = 0 | \mathbf{x}, \mathbf{w}, b) = \sigma\left(\sum_{j=0}^N w_j \cdot x_j + b\right) \quad (1)$$

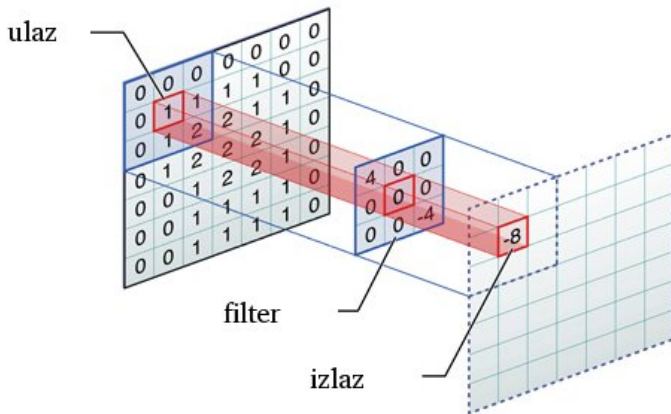
σ – logistička funkcija, $\sigma(x) = \frac{1}{1+e^{-x}}$



Slika: Perceptron

- povezivanje više neurona u slojeve
- modeliranje kompleksnih nelinearnih zavisnosti
- ulazni sloj, skriveni sloj, izlazni sloj
- učenje algoritmom unazadne propagacije (engl. *backpropagation*)
- negativne kritike
 - zapinjanje u lokalnim optimumima
 - nestajući gradijenti (engl. *vanishing gradient*)
- duboke mreže
- *state-of-the-art* u većini primjera računalnog vida

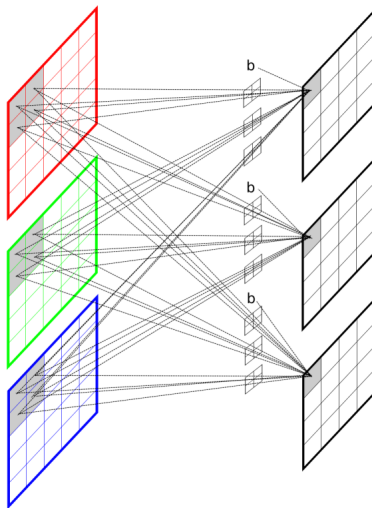
Konvolucijski operator



Slika: Konvolucija kroz više kanala¹

¹https://developer.apple.com/library/ios/documentation/Performance/Conceptual/vImage/Art/kernel_convolution.jpg

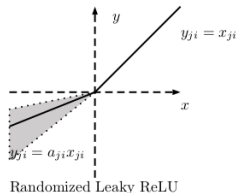
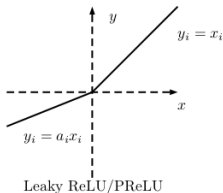
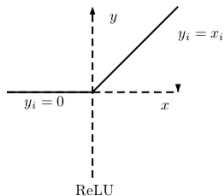
Prikaz konvolucije s više kanala



Slika: Prikaz konvolucija

Aktivacijske funkcije

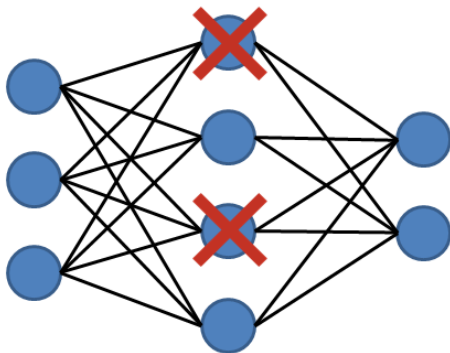
- logistička sigmoid funkcija
- hiperbolna tangens funkcija
- ReLU funkcije (engl. *Rectified Linear Unit*)
 - rješava problem zasićenja gradijenata
 - više vrsta: standardne, leaky, ...



Slika: ReLU funkcije

- sprječava prenaučenosť
- korištene tehnike
 - L_2 regularizacija, $L_2(\mathbf{w}) = \|\mathbf{w}\|_2 = \sum_{i=1}^n w_i^2$
 - umjetno povećanje skupa za učenje (slučajne transformacije)
 - **translacije** slike po x ili y osi za neki malen pomak, na primjer 5% slike
 - **rotacije** slike za kut od ± 7 deg
 - **skaliranje** (uvećavanje ili smanjivanje) slike za neki faktor koji je obično u raponu od $< 0.9, 1.1 >$
 - **smik** slike za kut od ± 5 deg
 - *dropout* - "isključivanje" neurona

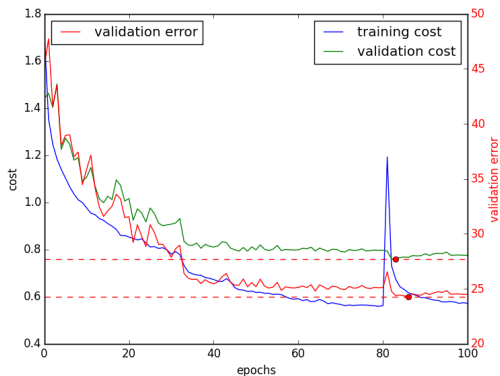
Dropout



Slika: Dropout

Učenje u dva koraka

- 1 učenje konvolucijskih slojeva - klasifikacija pojedinog piksela samo slojem logističke (multinomijalne) regresije
- 2 učenje klasifikatora - učenje potpuno povezanog sloja i sloja logističke (multinomijalne) regresije dodanih na konvolucije



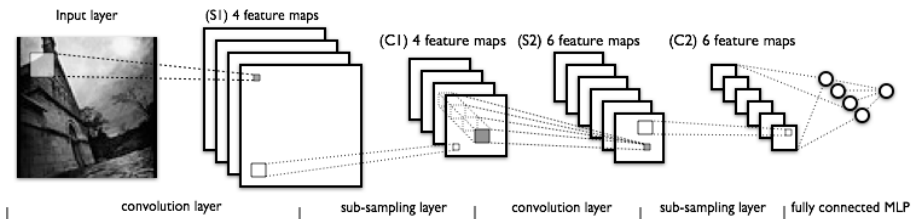
Konfiguracijska programske implementacije

- JSON format
- konfiguracija parametara mreže

```
{
  "evaluation": {
    "batch-size": 4
  },
  "network": {
    "layers": [16, 64, 256, 1000],
    "loss": "negative_log_likelihood",
    "builder-name": "build_multiscale",
    "seed": 23451
  },
  "training": {
    "optimization": "rms",
    "optimization-params": {
      "learning-rate": 0.0002,
      "momentum": 0.9
    },
    "epochs": -1,
    "learning-rate-decrease-params": {
      "no-improvement-epochs": 4,
      "min-learning-rate": 0.00001
    }
  },
}
```

Primjer mreže

- LeNet je primjer često citirane konvolucijske mreže
- neuron (m-1) sloja povezan samo sa prostorno bliskim neuronima m-tog sloja
- u početnim se slojevima izmjenjuju slojevi maksimalnog odziva i konvolucijski slojevi
- zadnji dio je potpuno povezani sloj na čije su ulaze spojeni izlazi zadnjeg sloja maksimalnog sažimanja



Slika: LeNet mreža