

Duboke konvolucijske neuronske mreže za raspoznavanje znakova

Matija Ilijaš

2. srpnja 2015.

Sadržaj

Optičko prepoznavanje znakova

Umjetne neuronske mreže

Duboko učenje

Duboke konvolucijske mreže

Implementacija

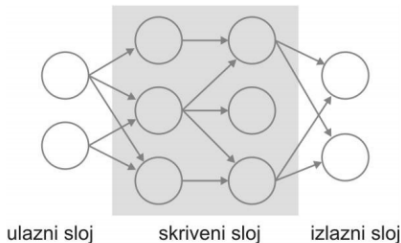
Rezultati

Optičko prepoznavanje znakova

- ▶ engl. Optical Character Recognition - OCR
- ▶ Područje računalnog vida, raspoznavanja uzoraka
- ▶ Dva glavna zadatka OCR sustava:
 - ▶ Segmentacija znakova iz slike
 - ▶ Klasifikacija znakova
- ▶ U ovom radu analiziraju se metode klasifikacije
- ▶ Standardna metoda provodi se u dva koraka:
 - ▶ Predprocesiranje slike znaka
 - ▶ Klasifikacija znaka

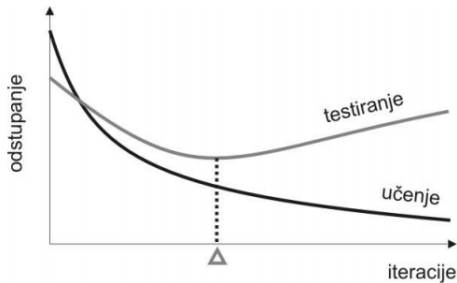
Umjetne neuronske mreže

- ▶ Algoritam strojnog učenja inspiriran strukturom i funkcionalnošću ljudskog mozga
- ▶ Paralelna obrada podataka, nelinearna veza ulaza i izlaza
- ▶ Neuron - težinska suma ulaza sa dodanom nelinearnošću
- ▶ Standardna arhitektura neuronske mreže
 - ▶ Potpuno povezana i aciklička
 - ▶ Sigmoidalna prijenosna funkcija



Učenje neuronskih mreža

- ▶ Neuronska mreža - nelinearna funkcija sa više varijabli
- ▶ Učenje mreže - traženje minimuma te funkcije s obzirom na grešku klasifikacije
- ▶ Koristi se Backpropagation algoritam koji implementira metodu gradijentnog spusta



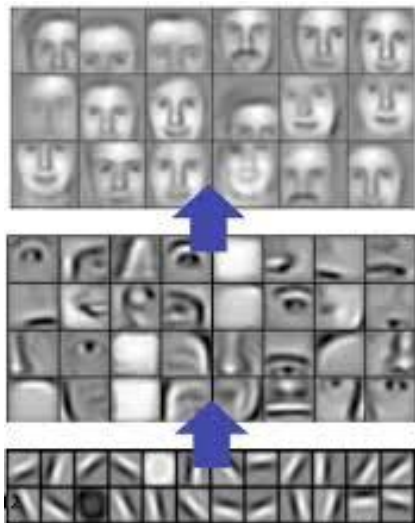
Duboko učenje

- ▶ Ideja prvi put prezentirana u radovima Kunihika Fukushima i Yanna Lecuna prije gotovo 30 godina
- ▶ Učenje korisnih značajki podataka povezanih u obliku hijerarhije - inspirirano vizualnim korteksom mozga sisavaca
- ▶ Uče se duboki kompleksni modeli na velikoj količini podataka - tek nedavno postalo moguće učenjem na grafičkim karticama
- ▶ Metode nenadziranog učenja
 - ▶ Učenje hijerarhije na neoznačenim podacima
 - ▶ Autoenkoderi, ograničeni Boltzmann strojevi
- ▶ Metode nadziranog učenja
 - ▶ Učenje Backpropagation algoritmom
 - ▶ Specijalne arhitekture (npr. konvolucijske)

Učenje hijerarhije značajki

- ▶ Kvalitetna značajka - diskriminantna, robusna na promjene osvjetljenja i geometrijske transformacije
- ▶ Ručno određivanje značajki - promatranje svojstva ulaznih podataka te naglašavanje onih maksimalno diskriminantnih
- ▶ Negativna posljedica ručnog određivanja je gubitak dijela korisnih informacija
- ▶ Rješenje: određivanje diskriminantnih značajki izravno iz skupa podataka
- ▶ Jedna razina apstrakcije - LDA i PCA metoda
- ▶ Više razina apstrakcije - hijerarhija značajki naučena dubokim učenjem na skupu podataka

Učenje hijerarhije značajki



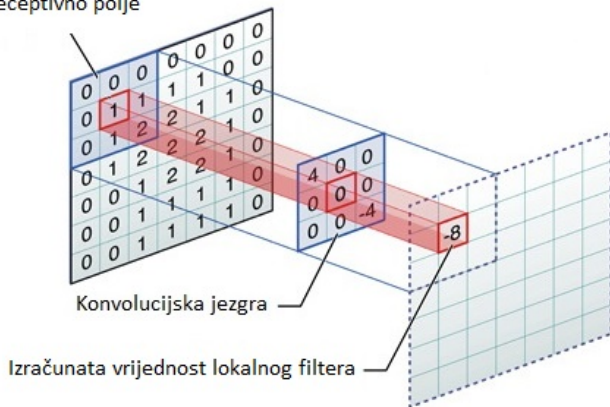
Duboke konvolucijske mreže

- ▶ Prvi model razvio Yann Lecun 1989. godine
- ▶ Iskorištavaju svojstvo lokalne prostorne korelacije prirodnih slika
- ▶ Lokalni prostorni filteri ostvareni konvolucijom oponašaju lokalne receptore vizualnog korteksa
- ▶ Cilj je naučiti konvolucijske jezgre specifične za dane podatke - učenje lokalnih značajki iz podataka
- ▶ Prvi dio arhitekture zadužen za formiranje hijerarhije značajki, drugi za klasifikaciju
- ▶ Nadzirano učenje Backpropagation algoritmom

Lokalni prostorni filteri

$$y(m, n) = x(m, n) * h(m, n) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x(i, j) h(m - i, n - j)$$

Lokalno receptivno polje

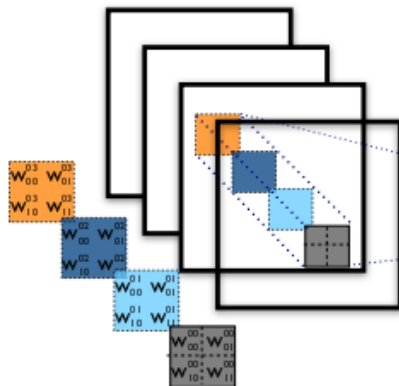


Dijeljenje težina

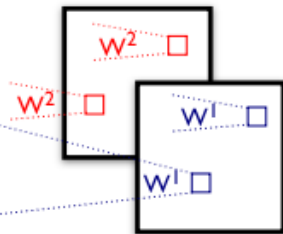
- ▶ Naučeni lokalni filter prepoznaje lokalnu značajku na samo jednoj poziciji u prostoru
- ▶ Mapa značajki - skup inačica lokalnih filtera koje pokrivaju sve moguće podprostore
- ▶ Element konvolucijskog sloja je matrica vrijednosti odziva pripadajućeg filtera za svaku poziciju u prostoru
- ▶ Dijeljenjem težina između parametara mreže smanjuje se ukupan broj slobodnih parametara, te se kao posljedica:
 - ▶ Smanjuje trajanje učenja
 - ▶ Poboljšava sposobnost generalizacije

Dijeljenje težina

Mape značajki sloja m-1



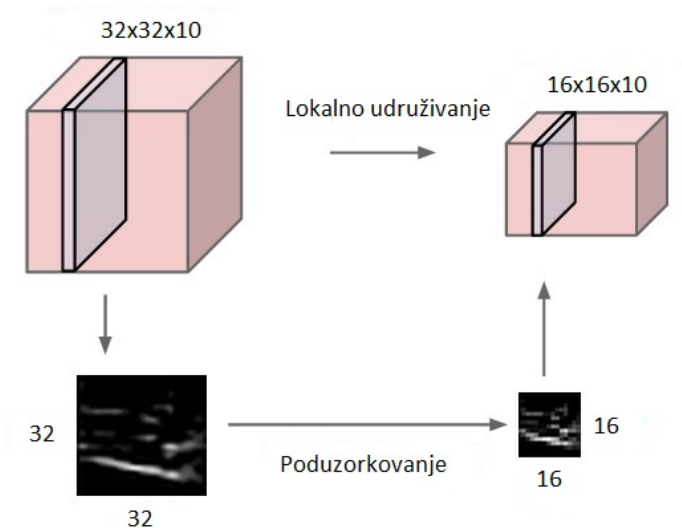
Mape značajki sloja m



Lokalno udruživanje

- ▶ engl. Pooling
- ▶ Oblik nelinearnog poduzorkovanja
- ▶ Slika se dijeli na nepreklapajuće podprozore fiksne veličine
- ▶ Maksimalno lokalno udruživanje - uzimanje maksimalne vrijednosti svakog podprozora
- ▶ Koriste se kao dodatni sloj nakon svakog sloja konvolucije, što
 - ▶ smanjuje količinu računanja u višim slojevima
 - ▶ ostvaruje jedan oblik invarijantnosti na translaciju

Lokalno udruživanje

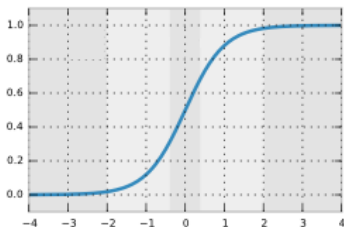


Ispravljena linearna jedinica

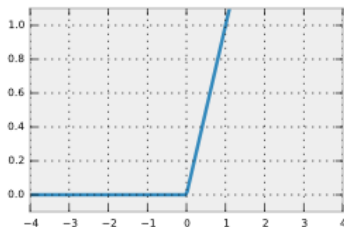
- ▶ engl. Rectified Linear Unit - ReLU
- ▶ Rješava problem nestajućeg gradijenta
- ▶ Ubrzava učenje svojim jednostavnim računanjem

$$f(x) = \max(x, 0)$$

Sigmoidalna funkcija

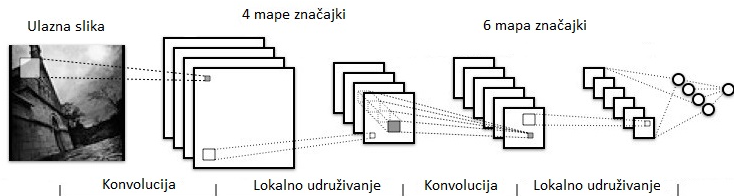


Ispravljena linearna jedinica



Konačni model

- ▶ Sastoji se od slojeva za formiranje hijerarhije značajki i slojeva za klasifikaciju
- ▶ Svaki sloj za formiranje hijerarhije sastoji se od:
 - ▶ Konvolucijskog sloja
 - ▶ Sloja ispravljenih linearnih jedinica
 - ▶ Sloja maksimalnog lokalnog udruživanja
- ▶ Slojevi za klasifikaciju su standardni i potpuno povezani



Implementacija

- ▶ Sustav za učenje i testiranje različitih arhitektura neuronskih mreža
- ▶ Prilikom razvijanja zadovoljena dva glavna uvjeta:
 - ▶ Brzo učenje kompleksnih modela na velikoj količini podataka
 - ▶ Detaljna analiza rada naučenih modela
- ▶ Sustav je podijeljen na dva podsustava razvijena sa zasebnim alatima:
 - ▶ Sustav za učenje
 - ▶ Sustav za testiranje
- ▶ Dodatno implementirana prilagodba OCR sustavu tvrtke Microblink

Sustav za učenje

- ▶ Implementiran u razvojnom alatu Torch koji omogućuje paralelno izvođenje procesa učenja na grafičkoj kartici
- ▶ Napisane skripte u jeziku LuaJIT za lagano definiranje arhitektura mreža i određivanje parametara učenja
- ▶ Naučeni Torch modeli se spremaju u posebnom formatu i šalju sustavu za testiranje

```
-- Simple 2-layer neural network
model = nn.Sequential()
model:add(nn.Reshape(ninputs))
model:add(nn.Linear(ninputs,nhiddens))
model:add(nn.Tanh())
model:add(nn.Linear(nhiddens,noutputs))
```

Sustav za testiranje

- ▶ Implementiran u jeziku C++ korištenjem OpenCV biblioteke
- ▶ Za pokretanje Torch modela koristi se C++ biblioteka JTorh
- ▶ Analiza točnosti klasifikacije
 - ▶ Matrica greške klasifikacije
 - ▶ Prikaz krivo klasificiranih primjera
- ▶ Analiza kompleksnosti modela
 - ▶ Broj slobodnih parametara
 - ▶ Prosječno trajanje klasifikacije
- ▶ Simultano analiziranje više modela
 - ▶ Tablice točnosti po klasama
 - ▶ Tablice trajanja klasifikacije

Rezultati

- ▶ Modeli ućeni i testirani na internim skupovima podataka tvrtke Microblink
- ▶ Slike znakova sa osobnih iskaznica, tzv. strojno ćitljivih područja (MRZ)
 - ▶ 37 klasa znakova fonta ocrb
 - ▶ Obuhvaćaju znamenke, velika slova i znak manje
- ▶ Sintetizirani skup podataka
 - ▶ Dobiven slikanjem printanih imitacija osobnih iskaznica
 - ▶ Više razlićitih modela mobitela i uvjeta osvjetljenja
 - ▶ 300 tisuća slika za ućenje, 150 tisuća za testiranje
- ▶ Rućno oznaćeni skup podataka
 - ▶ Dobiven slikanjem osobnih iskaznica u razlićitim uvjetima
 - ▶ 12 tisuća rućno oznaćenih slika za testiranje

Rezultati

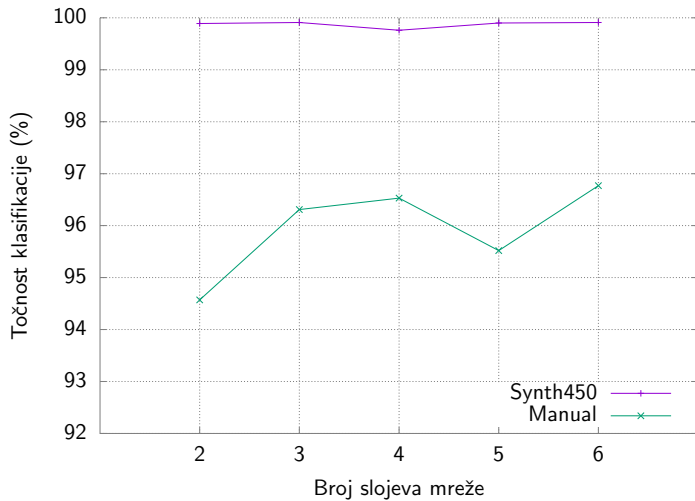
- ▶ Utjecaj predprocesiranja podataka
- ▶ Pronalaženje optimalne arhitekture
 - ▶ Standarnih neuronskih mreža
 - ▶ Konvolucijskih neuronskih mreža
- ▶ Usporedba optimalnih arhitektura
 - ▶ Točnost klasifikacije
 - ▶ Kompleksnost i trajanje klasifikacije
- ▶ Usporedba sa Microblink klasifikatorom

Utjecaj predprocesiranja podataka

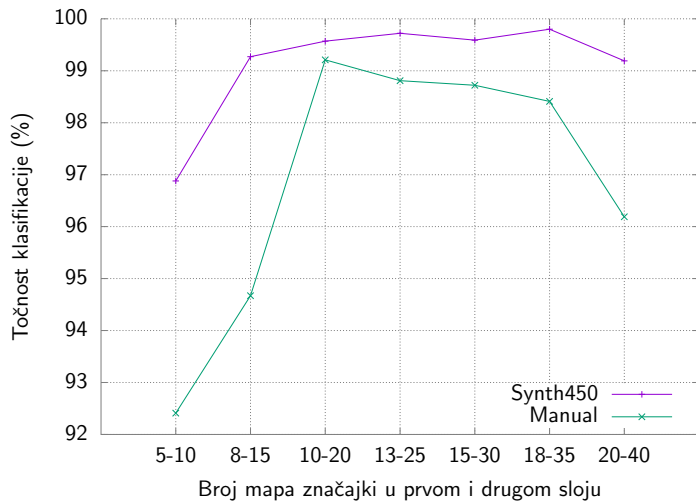


Skup podataka	Adaptivni prag	Microblink interni	Bez predprocesiranja
Sintetizirani	98.29%	97.87%	99.57%
Ručno označeni	96.35%	97.66%	99.21%

Pronalaženje optimalne standardne arhitekture



Pronalaženje optimalne konvolucijske arhitekture



Usporedba točnosti optimalnih arhitektura

Skup podataka	Standarna arhitektura	Konvolucijska arhitektura
Sintetizirani	99.91%	99.57%
Ručno označeni	96.31%	99.21%

- Konvolucijska arhitektura bolje generalizira na ručno označenom skupu

Arhitektura	Znamenka '0'	Slovo 'O'	Znamenka '8'	Slovo 'B'
Standarna	74.15%	91.97%	98.69%	38.57%
Konvolucijska	97.55%	95.35%	99.81%	99.29%

- Najveća razlika u točnosti kod teških slučajeva sličnih klasa

Usporedba kompleksnosti optimalnih arhitektura

Svojstvo	Standardna arhitektura	Konvolucijska arhitektura
Računske operacije	170550	97780
Slobodni parametri	170550	6930
Trajanje klasifikacije	0.202 ms	0.223 ms

- ▶ Po broju računskih operacija standardni model gotovo dvostruko kompleksniji, no prosječno trajanje klasifikacije je približno jednako
- ▶ Konvolucijska arhitektura ima za red veličine manji broj slobodnih parametara od standardne arhitekture

Usporedba sa Microblink klasifikatorom

Klasifikator	Sintetizirani skup	Ručno označeni skup
Microblink	86.93%	91.69%
Standarna mreža	99.91%	96.31%
Konvolucijska mreža	99.57%	99.21%

- ▶ Neuronske mreže bolje klasificiraju na oba testna skupa

Klasifikator	Prosječno trajanje klasifikacije
Microblink	0.025 ms
Standarna mreža	0.202 ms
Konvolucijska mreža	0.223 ms

- ▶ Microblink klasifikator za red veličine brži od neuronskih mreža

Zaključak

- ▶ Pokazana je prednost prosljeđivanja neobrađenih slika dubokom klasifikatoru u svrhu očuvanja informacija slike
- ▶ Dodavanjem dubine bilo kojoj arhitekturi neuronske mreže pospješuje se njena sposobnost generalizacije
- ▶ Pokazana je prednost korištenja konvolucijske arhitekture zbog dodatnog ograničenja na njenu kompleksnost što pospješuje generalizaciju
- ▶ Rezultati klasifikatora tvrtke Microblink lošiji su od rezultata standardnih i konvolucijskih neuronskih mreža, što se može objasniti:
 - ▶ Negativnim utjecajem predprocesiranja podataka
 - ▶ Korištenjem značajki samo jedne razine apstrakcije
- ▶ Naučene mreže su za red veličine sporije od Microblink klasifikatora, no još uvijek dovoljno brze za primjenu u stvarnom vremenu na mobilnom uređaju

Hvala na pažnji!