SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

POSLIJEDIPLOMSKI DOKTORSKI STUDIJ ELEKTROTEHNIKE I INFORMACIJSKIH TEHNOLOGIJA

KVALIFIKACIJSKI ISPIT

SEMANTIČKA SEGMENTACIJA SLIKA METODAMA DUBOKOG UČENJA

Dunja Božić-Štulić

Split, 8. prosinca 2017.g.

SADRŽAJ

1. Uvod

Semantička segmentacija slike zadnjih je godina postala predmetom interesa istraživača na području računalnog vida, te strojnog učenja. Razlog tome je što mnoge aplikacije današnjeg doba zahtjevaju precizne, te učinkovite mehanizme segmentiranja slike npr. autonomna vožnja, navigacija, pa čak i sustavi temeljeni na virtualnoj stvarnosti. Semantička segmentacija koristi se u razumijevanju 2D slika i videa, pa čak i 3D ili više-dimenzionalnih podataka, no unatoč širokoj upotrebi i dalje je jedna od glavnih tema u području računalnog vida. Kao takva spada u kompleksne zadatke računalnog vida, koji vode prema razumijevanju scena. Razumijevanje scene od iznimne je važnosti, te spada u jedne od osnovnih problema računalnog vida, zbog činjenice da se javlja velika potreba za aplikacijama koje zaključuju temeljem učenja sa slike. Neke od takvih aplikacija uključuju autonomnu vožnju [60, 61, 62], interakciju čovjeka s računalom [63], računalnu fotografiju [64], pretraživanje fotografija [65].

Standardi se pristup raspoznavanju objekata u ovakvim aplikacijama bazirao na tradicionalnim tehnikama računalnog vida, te strojnog učenja: ručni dizajn značajki, njihova agregacija te treniranje klasifikatora nad uzorcima. Usprkos velikoj popularnosti ovakvog načina rada dolaskom dubokog učenja problemi računalnog vida uključujući i samu semantičku segmentaciju, počeli su se rješavati korištenjem dubokih arhitektura. Najčešće korištene duboke arhitekture upravo su konvolucijske neuralne mreže (CNNs) [29, 28, 48] koje su svojim performancama nadmašile druge tehnike, posebno u smislu točnosti i efikasnosti. No usprkos izvrsnom uspjehu, područje dubokog učenja nije do kraja istraženo, te ga ne možemo opisati kao dobro istraženo područje.

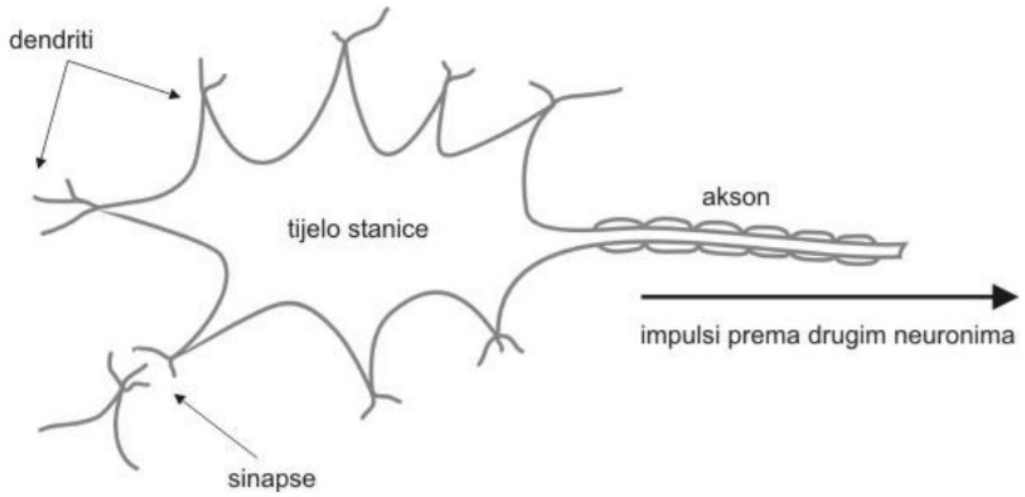
U ovom radu dat je pregled dubokih metoda učenja za semantičku segmentaciju slike koja je svoju primjenu pronašla u različitim područjima. Rad je organiziran na slijedeći način. U drugom poglavlju objašnjene su umjetne neuronske mreže, te detaljnije pojašnjena razlika između bioloških i umjetnih neurona. U trećem poglavlju opisane su konvolucijske neuronske mreže, njihova arhitektura i karakteristike. U četvrtom poglavju dat je pregled najvažniji arhitektura područja dubokog učenja, te je u petom poglavlju dat pregled arhitektura za semantičku segmentaciju slike.

2. Umjetne neuronske mreže

2.1 Biološki i umjetni neuron

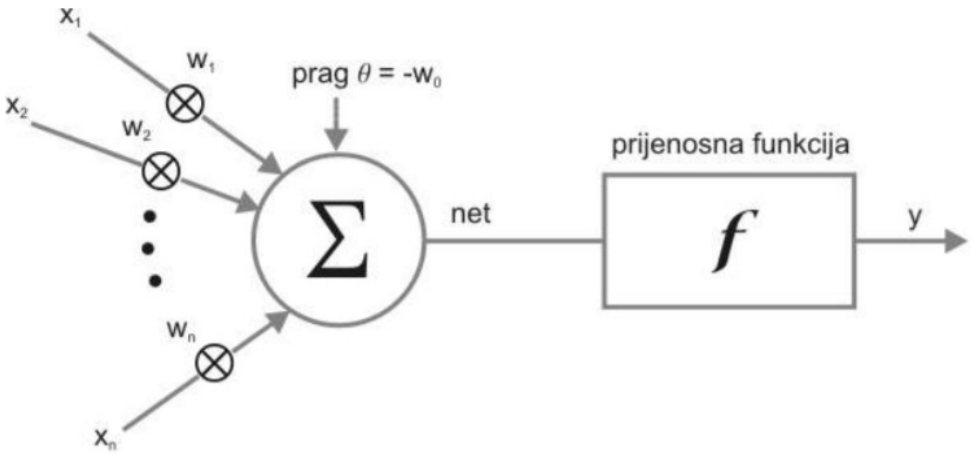
Za razumijevanje sposobnosti mozga nužno je upoznati građu njegova sastavnog dijela neurona. Ljudski mozak sastavljen je od oko 100 miljardi neurona kojih ima više od 100 vrsta i koji su shodno svojoj funkciji raspoređeni prema točno definiranom rasporedu. Svaki je neuron u prosjeku povezan s 104 drugih neurona. Četiri su osnovna dijela neurona: tijelo stanice (soma), skup dendrita (ogranaka), aksona (dugačke cijevčice koje prenose električke poruke) i niza završnih članaka. Slika 1. prikazuje građu neurona.

Tijelo stanice sadrži informaciju predstavljenu električkim potencijalom između unutrašnjeg i vanjskog dijela stanice (oko –70 mV u neutralnom stanju). Na sinapsama, spojnom sredstvu dvaju neurona kojim su pokriveni dendriti, primaju se informacije od drugih neurona u vidu post-sinaptičkog potencijala koji utječe na potencijal stanice povećavajući (hiperpolarizacija) ili smanjivajući ga (depolarizacija). U tijelu stanice sumiraju se post-sinaptički potencijali tisuća susjednih neurona, u ovisnosti o vremenu dolaska ulaznih informacija. Ako ukupni napon pređe određeni prag, neuron "pali" i generira tzv. akcijski potencijal u trajanju od 1 ms. Kada se informacija akcijskim potencijalom prenese do završnih članaka, onda oni, ovisno o veličini potenijala, proizvode i otpuštaju kemikalije, tzv. neurotransmitere. To zatim ponovno inicira niz opisanih događaja u daljnjim neuronima. Propagacija impulsa očigledno je jednosmjerna [1].



Slika 1. Građa neurona. Preuzeto iz [1].

Umjetne neuronske mreže (*engl. artificial neural networks*) privukle su pozornost istraživača 1943. godine, kada su Warren McCulloch i Walter Pitts predstavili prvi model umjetnih neurona. Umjetna neuronska mreža u širem je smislu riječi umjetna replika ljudskog mozga kojom se nastoji simulirati postupak učenja. To je paradigma kojom su implementirani pojednostavljeni modeli što sačinjavaju biološku neuronsku mrežu. Analogija s pravim biološkim uzorom zapravo je dosta klimava jer uz mnoga učinjena pojednostavljena postoje još mnogi fenomeni živčanog sustava koji nisu modelirani umjetnim neuronskim mrežama, kao što postoje i karakteristike umjetnih neronskih mreža koje se ne slažu s onima bioloških sustava.



Slika 2. Umjetni neuron. Pruzeto iz [1].

Neuronska mreža jest skup međusobno povezanih jednostavnih procesnih elemenata, jedinica ili čvorova, čija se funkcionalnost temelji na biološkom neuronu. Pri tome je obradbena moć mreže pohranjena u snazi veza između pojedinih neurona tj. težinama do kojih se dolazi postupkom prilagodbe odnosno učenjem iz skupa podataka za učenje. Neuronska mreža obrađuje podatke distribuiranim paralelnim radom svojih čvorova.

2.2 Arhitekture umjetnih neuralnih mreža

Općenito, arhitektura umjetnih neuronskih mreža može se podijeliti na tri dijela. Dijelovi se zovu slojevi, a dijelimo ih na:

- *Ulazni sloj*

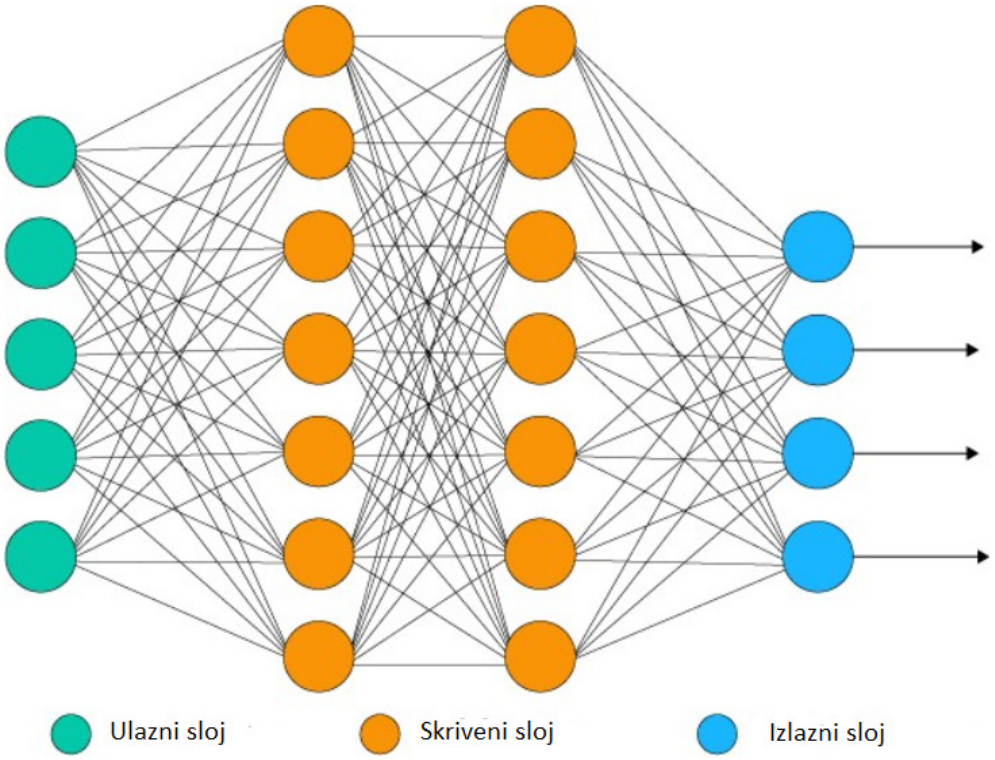
Ovaj sloj zadužen je za primanje informacija (podataka), signala, značajki ili mjera uzetih iz nekog vanjskog okruženja. Ovi ulazi uobičajeno su normalizirani unutar vrijednosti aktivacijkse funkcije. Normalizacja rezultira boljom preciznošću matematičkih funcija unutar mreže.

- *Skriveni sloj*

Ovi slojevi sastavljeni su od neurona, zaduženih za ekstrakciju uzoraka povezanih sa sistemom ili procesom koji se analizira.

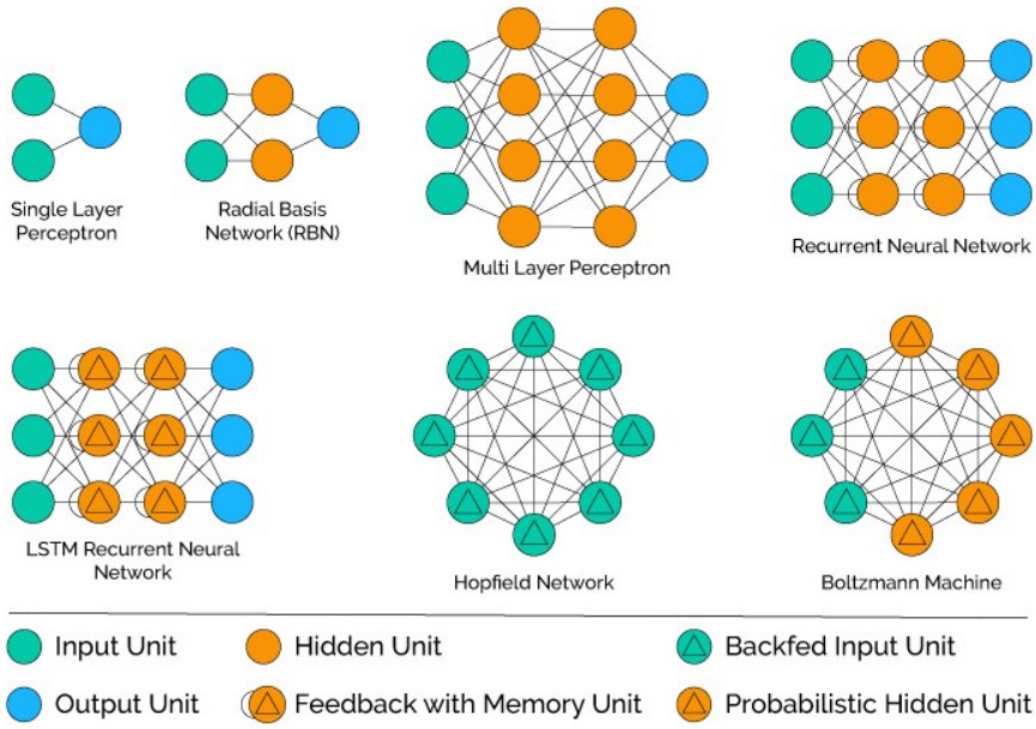
- *Izlazni sloj*

Izlazni sloj sastavljen je od neurona koji su zaduženi za prezentaciju izlaza mreže.



Slika 3. Grafički prikaz slojeva umjetne neuronske mreže. Prilagođeno i preuzeto iz [3].

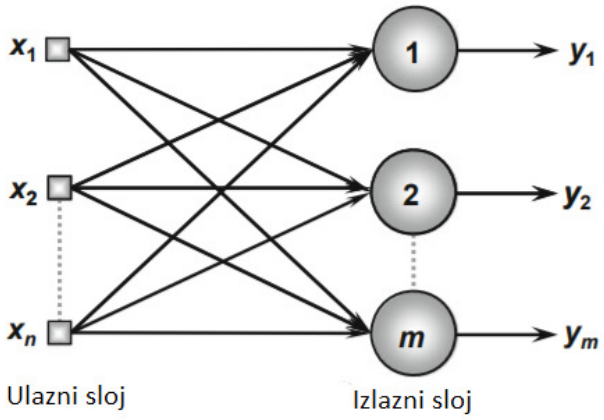
Na Slici 3. prikazan je grafički prikaz slojeva umjetne neuronske mreže. Osnovna arhitektura umjetnih neuronskih mreža, s obzirom na pozicije neurana, kao i njihovu povezanost dijeli se na: jednoslojna unaprijedna mreža, višeslojna unaprijedna mreža, povratna mreža. Na Slici 4. prikazne su različite umjetne neuronske mreže.



Slika 4. Prikaz različitih umjetnih neuronskih mreža. Preuzeto iz [3].

2.2.1 Jednoslojna unaprijedna mreža

Jednoslojna unaprijedna mreža sastoji se od jednog ulaznog, te jednog neuronskog sloja, koji je istodobno i izlazni sloj. Slika 5. prikazuje jednostavnu unaprijednu mrežu sastavljenju od n ulaza i m izlaza. Informacija uvijek ide u jedno smjeru, odnosno od ulaznog sloja do izlaznog. Iz Slike 5. vidljivo je da mreža ovakvog tipa arhitekture sadrži broj izlaza jednak broju ulaznih neurona. Ovakav tip arhitekture uobičajeno se koristi u problemima klasifikacije i linearnog filtriranja.



Slika 5. Jednoslojna unaprijedna mreža

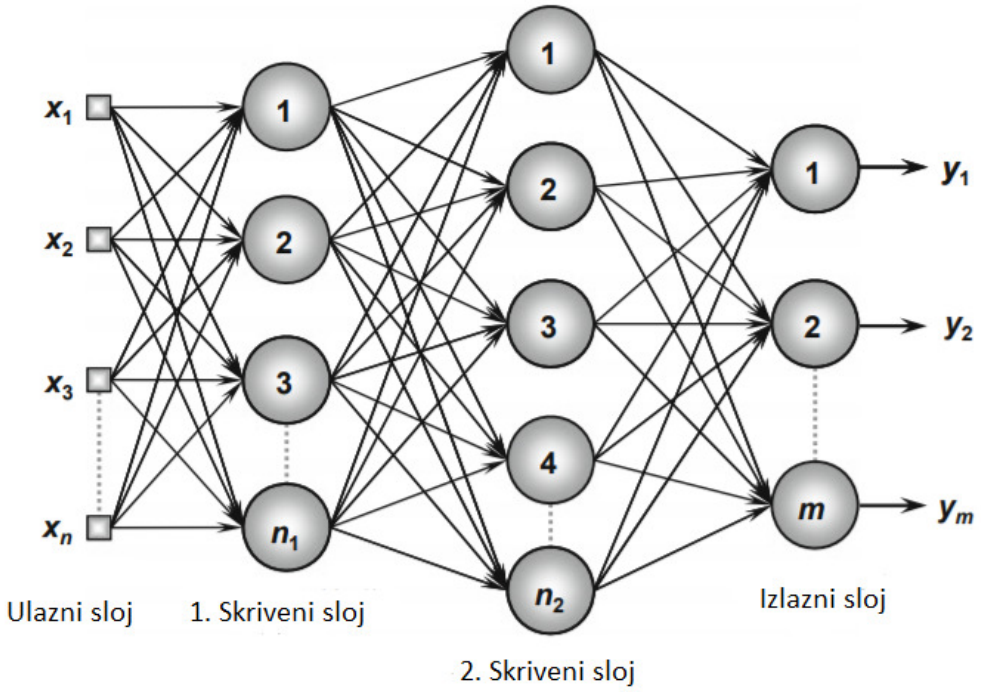
Među najpoznatijim arhitekturama ovog tipa su peceptron [4] i ADALINE [5].

2.2.2 Višeslojna unaprijedna mreža

Suprotno mrežama koje pripadaju predhodno opisanoj arhitekturi, višeslojne unaprijedne mreže sastavljene su od jednog ili više skrivenih slojeva (Slika 6.). Koriste se u rješavanju različitih problema, poput onih povezanih s aproksimacijom, klasifikacijom uzoraka, identifikacijom sustava, kontrolom procesa, optimizacijom, robotikom itd.

Slika 6. prikazuje višeslojnu unaprijednu mrežu sastavljenu od jednog ulaznog sloja, s n ulaznih uzoraka, dva skrivena sloja sastavljenih od n_1 i n_2 neurona, te izlaznog sloja sastavljenog od neurona u ovisnosti o broju izlaznih vrijednosti problema koji se analizira.

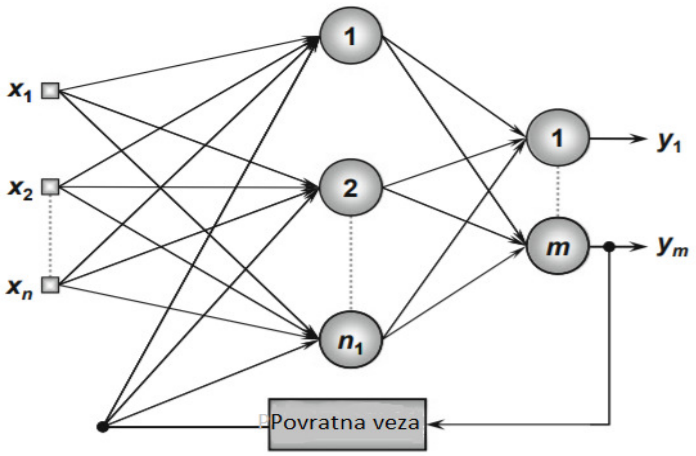
Među ovaj tip arhitekture spadaju višeslojni percepton [41] (engl. multilayer percepton), te RBF (engl. radial basis function). Iz Slike 4. vidljivo je da je broj neurona od kojih se sastoji prvi skriveni sloj, drugačiji od broja neurona koji čine ulazni sloj. U stvarnosti broj neurona skrivenog sloja ovisi o prirodi, te složenosti problema koji analiziramo u skrivenom dijelu mreže, kao i o kvaliteti i kvantiteti dostupnih podataka koje koristimo za analizu.



Slika 6. Višeslojna unaprijedna mreža

2.2.3 Povratna mreža

U arhitekturama povratnog tipa izlazi iz svakog neurona koriste se kao povratne informacije za druge neurone. Povratne informacije značajki kvalificiraju ovakav tip arhitekture za procesiranje dinamičkih informacija, što znači da ih možemo koristiti u vremenski promjenjivim sustavima, kao što su sustavi za optimizaciju i prepoznavanje, vođenja procesa itd.



Slika 7. Povratna umjetna neuronska mreža

Najpoznatije mreže ovog tipa su Hopfield [10] i percepton s povratnom informacijom neurona iz različitih slojeva [11]. Slika 7. prikazuje percepton mrežu s povratnom informacijom, gdje se jedan izlazni signal vraća natrag u srednji sloj, čime se postiže uzimanje u obzir i predhodnih izlaznih vrijednosti.

2.3 Procesi i svojstva učenja

Jedna od najrelevantnijih značajki umjetne neuronske mreže je njena sposobnost učenja iz prezentacije uzoraka. Jednom kada mreža nauči vezu između uzoraka i njenih izlaza, sposobna je generalizirati rješenje. Generalizacija rješenja podrazumijeva da je mreža sposobna dati izlaz koji je dovoljno blizu očekivanom (željenom) izlazu, bilo koje ulazne vrijednosti. Proces treniranja umjetne neuronske mreže sastoji se odF uobičajenih koraka „uglađivanja“ sinaptičkih težina i pragova neurona, kako bi se postigla generalizacija. Setovi uobičajenih koraka nazivaju se algoritmi učenja. Tokom njihova izvršenja, mreža izdvaja značajke sustava. Obično, kompletni skup uzoraka sadrži sve moguće iteracje ponašanja sustava podijeljene u dvije skupine, skupina za treniranje, te skupina za testiranje. Skup za treniranje sastoji se od 60-90% slučajnih uzorka iz kompletnog seta, te se koristi u procesu učenja. S druge strane, skup za testiranje sastoji se od 10-40% kompletnog set uzoraka, te se koristi za provjeru dali je mreža prihvatljivo generalizirala problem, te dali su rješenja unutar prihvatljivih razina. Ovakvim pristupom omogućava se validacija određene topologije. Prilikom dimenzioniranja ovih skupova važno je razmotriti i statističke značajke podataka. Tijekom procesa učenja umjetnih neuronskih mreža, svaka cijelovatia prezentacija uzoraka iz seta treniranja u svrhu prilagodbe praga i težina, naziva se epoha učenja.

2.3.1 Nadzirano učenje

Strategija nadziranog učenja sastoji se od skupa željenih izlaza za dati skup podataka. Drugim riječima, proces učenja sastavljen je od ulaznih signala, te pripadajućih izlaza. Nadzirano učenje zahtjeva tablicu ulazno –izlaznih podataka koja predstavlja sam proces, te njegovo ponašanje. Iz ovih informacija izraditi će se hipoteza o sustavu koji se uči. Sinaptičke težine i pragovi mreže kontinuirano se prilagođavaju primjenom usporednih akcija, tj. Algoritam učenja uspoređuje sličnost između proizvedenih rezultata , te dostupnih željenih izlaza. Mreža se smatra istrenirana kada je ta razlika u prihvatljivim intervalima vrijednosti. Nadzirano učenje je tipični oblik induktivnog zaključivanja, gdje se varijable mreže podešavaju poznavanjem a priori željenog izlaza za ispitivani sustav.

Donald Hebb predložio je prvu nadziranu strategiju učenja 1949. godine, inspiriranu neurofiziološkim promatranjima [53].

2.3.2 Nenadzirano učenje

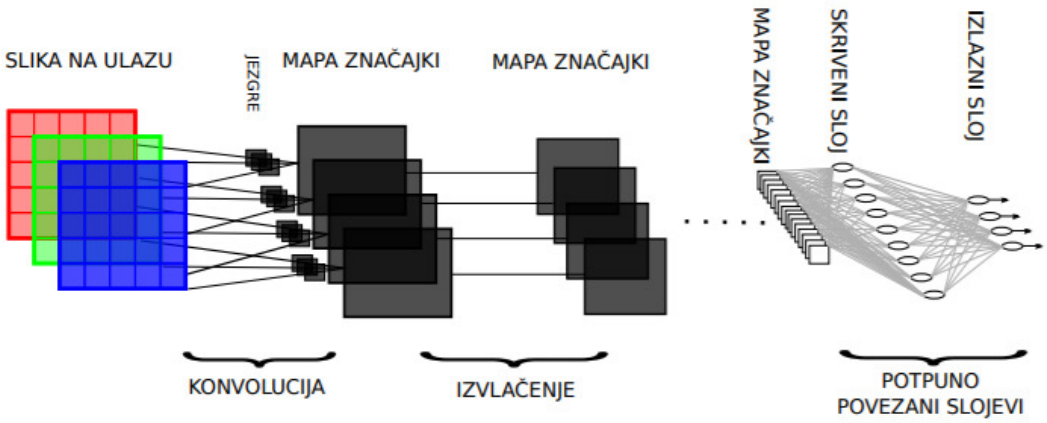
Za raziku od nadzornog učenja, ne nadzirano učenje ne zahtjeva znanje o željenom izlazu. Mreža mora sama zaključiti dali postoje sličnosti između elemenata koji čine set uzoraka, kreirajući klastere koji pokazuju sličnosti. Algoritam učenja podešava sinaptičke težine i pragove mreže kako bi se kreirali odgovarajući podskupovi. Dizajner mreže može prije procesa učenja definirati maksimalni broj klastera.

3. Konvolucijske neuronske mreže

Konvolucijske neuralne mreže (CNN) vrsta su neuronskih mreža s topologijom rešetke, specijalizirana za obradu podataka. Takve mreže mogu se prikazati kao proširenje klasičnih višeslojnih unaprijednih neuronskih mreža. Naime, unaprijedne neuronske mreže imaju nekoliko ograničenja, koja ih čine manje idealnim u riješavanju problema klasifikacije slike. ANN pretpostavljaju da su značajke nezavisne, što generalno nije održivo u većini stvarnih podataka. U kontekstu slike ovakva pretpostavka navodi da su pojedinačni pikseli međusobno nepovezani. Međutim, to nije slučaj sa velikom većinom slika, jer pikseli koji su bliski jedan drugome, vjerojatno pripadaju istom objektu ili vizualnoj strukturi, te bi samom tom činjenicom trebali biti tretirani na sličan način. Zbog navedenih problema zadatci klasifikacije slike riješavaju se konvolucijskim neuralnim mrežama (CNN).

3.1 Arhitektutra konvolucijskih neuralnih mreža

Na slici 1. prikazana je opća struktura konvolucijskih neuronskih mreža. Na ulazu može biti jedna monokromatska slika ili višekanalna slika u boji. Zatim slijede naizmjenice konvolucijski slojevi i slojevi sažimanja (engl. pooling). Na samom kraju se nalazi nekoliko potpuno povezanih slojeva (klasicni perceptron) koji su jednodimenzionalni, uključujući i izlazni sloj. Tipični primjeri konvolucijskih neuronskih mreža imaju oko desetak slojeva (cime jasno opravdavaju svoje mjesto u kategoriji dubokih neuronskih mreža). Konvolucijski slojevi i slojevi sažimanja imaju dvodimenzionalne "neurone" koji se nazivaju mapama značajki (engl. feature maps) koji u svakom sloju.

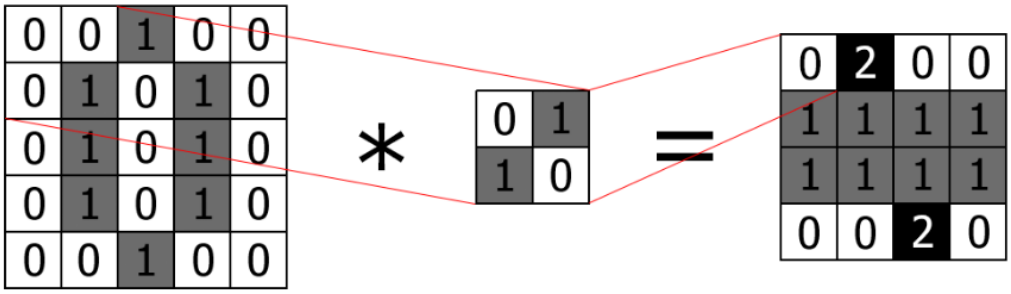


Slika 8. Opća struktura konvolucijskih neuronskih mreža

3.1.1 Konvolucijski sloj

U konvolucijskom sloju obavlja se operacija konvolucije nad matricom koju nazivamo kernel s ulaznom matricom. Na slici 9. prikazana je operacija konvolucije. Matrica na lijevoj strani

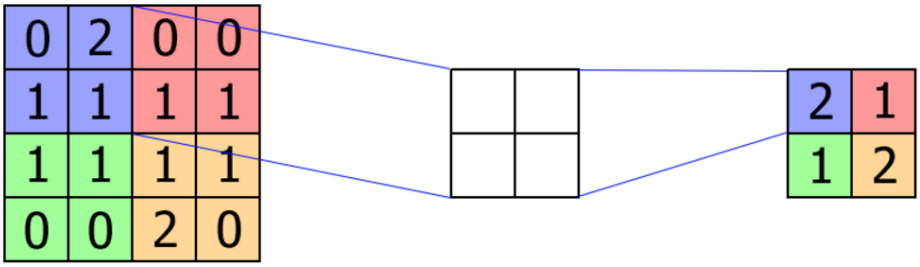
predstavlja ulazni podatak (obično se radi o slici) nad kojim se obavlja operacija konvolucije s srednjom matricom kernelom. Kernel je obično veličine $3x3$ ili $5x5$. Rezultat operacije konvolucije prikazan je u desnoj matrici.



Slika 9. Vizualni prikaz operacije konvolucije

3.1.2 Sloj sažimanja

Slojevi sažimanja (*engl. pooling*) služe za smanjenje dimenzija mapi značajki, te za uklanjanje varijance. U slojevima sažimanja imamo i okvire s kojima prolazimo preko mape značajki. Mapa značajki sažima se na način da se okvir predstavi sa samo jednom vrijednošću. Tako npr., na slici 10. možemo vidjeti kako se okvir veličine $2x2$ reprezentira s jednom vrijednošću dobivenom iz 4 vrijednosti unutar okvira čime se mapa značajki smanjuje 4 puta. Pomicanje okvira u navedenom primjeru bio bi jednak 2 u horizontalnom, te 2 u vertikalnom smjeru. Sažimanje je moguće odraditi na dva načina: sažimanje usrednjavanjem, te sažimanje maksimalnom vrijednošću. Sažimanje usrednjavanjem (*engl. mean pooling*) uzima aritmetičku sredinu vrijednosti koje se nalze unutar okvira sažimanja. Sažimanje maksimalnom vrijenošću (*engl. max pooling*) uzima maksimalnu vrijednost unutar okvira sažimanja. Na slici 10. dan je primjer sažimanja maksimalnom vrijednošću.



Slika 10. Primjer sloja sažimanja

3.2 Svojstva konvolucijskih neuronskih mreža

Konvolucijske neuronske mreže imaju nekoliko svojstava koja im omogucavaju dobru generalizaciju prilikom višeklasne klasifikacije.

3.2.1 Dijeljenje težina

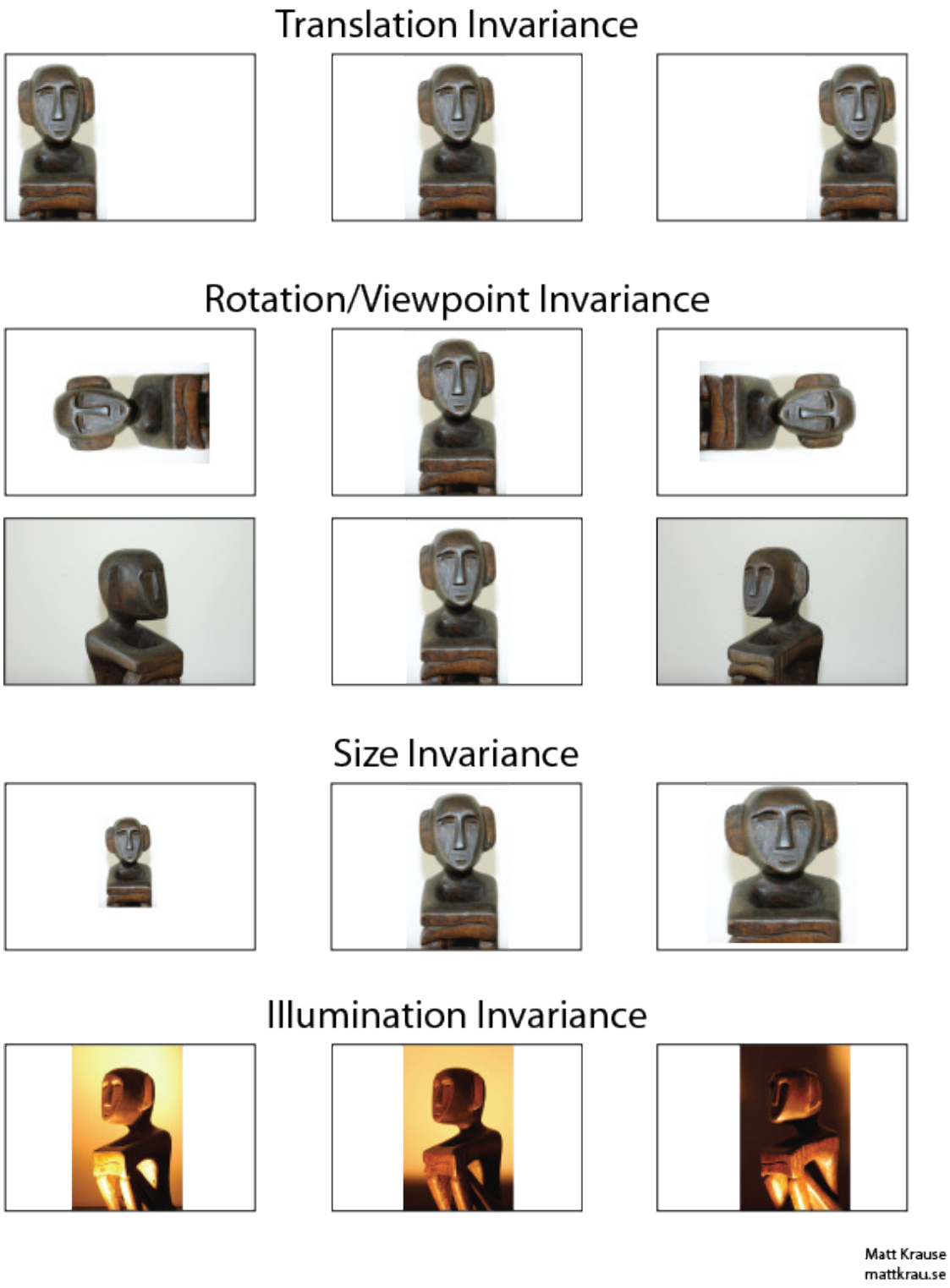
U konvolucijskim slojevima se za svaku konvoluciju jedne izvorne mape sa jednom izlaznom mapom koristi jedna jezgra (*engl. kernel*). Ukoliko se promatraju pojedini neuroni unutar mape, jasno je da svi ti neuroni dijele iste $K_W x K_h$ težine. Takvo dijeljenje težina omogućava da mreža nauči relevantne i diskriminativne značajke. Jezgre se specijaliziraju za određenu funkciju (primjerice - 13 detekcija horizontalnih i vertikalnih bridova, odziv na različite uzorke i sl.), te postaju slične npr. Haarovim i drugim značajkama. Bez dijeljenja težina dijelovi neuronske mreže mogli bi se pretrenirati na određeni detalj podataka. S dijeljenjem, na istu jezgru dolaze različiti podatci što povećava općenitost naučene značajke i poboljšava generalizacijske sposobnosti mreže.

3.2.2 Raspršena povezanost

Na slici 8. prikazana je raspršena povezanost. Korištenje raspršene povezanosti može uvelike pomoći u učenju različitih značajki ukoliko je skup za učenje takav da mreža ima težnju konvergiranju istom manjem broju značajki. Bez raspršene povezanosti (potpuna povezanost) sve mape primaju sve vrijednosti iz prethodnih mapa. U tom slucaju je moguće da dvije ili više mape konvergiraju ka istoj vrijednosti. Uvodenjem raspršene povezanosti mape dobivaju razli čite ulaze (samo neke mape prethodnog sloja) čime se osigurava konvergencija ka različitim vrijednostima.

3.2.3 Invarijantnost

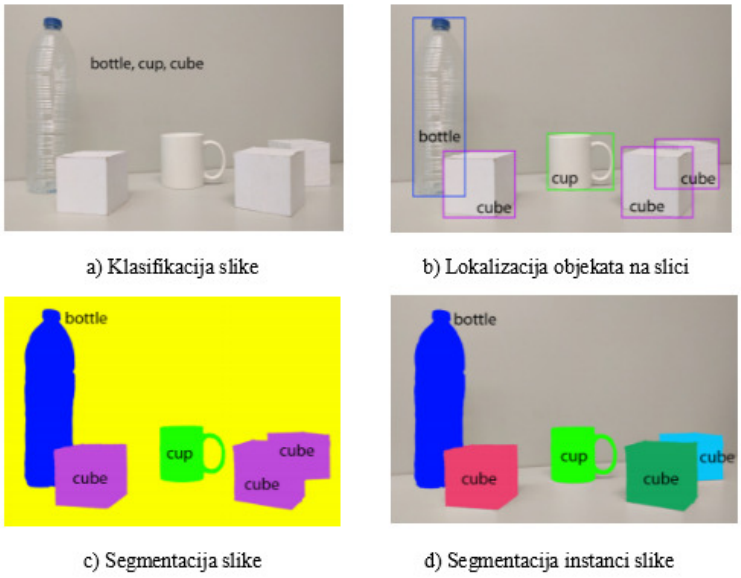
Invarijantnost omogućava konvolucijskoj neuronskoj mreži da bude otporna na male varijacije položaja značajki. Primarni mehanizam kojim se to postiže su slojevi sažimanja značajki koji smanjuju rezoluciju (odnosno dimenzionalnost) mapi značajki. S obzirom da se to postiže postepeno kroz više takvih slojeva, mreža i dalje uči međusobni položaj značajki (npr. očiju, nosa i ustiju kod detekcije lica), ali postaje otporna na manje varijacije u položaju. Slika 11. ilustrira neke tipove invarijantnosti (translacijsku, rotacijsku invarijantnost, invarijantnost veličine i invarijantnost osvjetljenja) [8].



Slika 11. Primjer različitih invarijantnosti mreže. Preuzeto iz [8].

4 Terminologija i osnovni koncepti dubokog učenja

Kako bi lakše razumijeli način na koji se semantička segmentacija riješava dubokim arhitekturama, važno je razumijeti da semantička segmentacija nije izolirano polje koje se istražuje, već prirodni korak ka boljem razumijevanju prirodnih scena. Prvi korak u boljem razumijevanju je **klasifikacija**, koja definira koji su objekti prisutni na sceni. **Lokalizacija** i **detekcija** slijedeći su korak, te nam osim klase koja je prisutna na sceni daju dodatne informacije o točnoj poziciji objekta. S obzirom na ove činjenice, prirodni korak dalje je **semantička segmentacija** čiji je cilj pridjeljivanja semantičkih oznaka dijelovima slike. Slika 12. prikazuje razvoj razumijevanja scene od grube klasifikacije objekata na slici do detaljnijeg opisa slike.



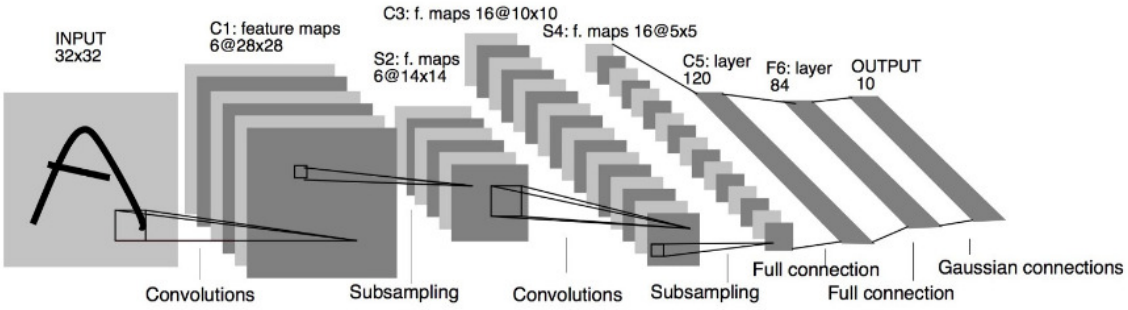
Slika 12. Razvoj prepoznavanja objekta ili razumijevanja scene od grube klasifikacije objekata na slici do detaljnijeg opisa slike (lokalizacija objekata, segmentacija istih, te segmentacija samih instanci slike). Preuzeto iz [52].

4.1 Standardne duboke arhitekture

Određene duboke arhitekture značajno su doprinjele području dubokog učenja, te su na neki način postale standard (LeNet5, AlexNet, VGG-16, GoogLeNet, ResNet). Ove arhitekture od velike su važnosti za problem segmentacije, jer se danas koriste kao jedan od blokova u semantičkim arhitekturama. Upravo zbog toga iduća poglavlja posvetit ćemo njima.

4.1.1 LeNet5 arhitektura

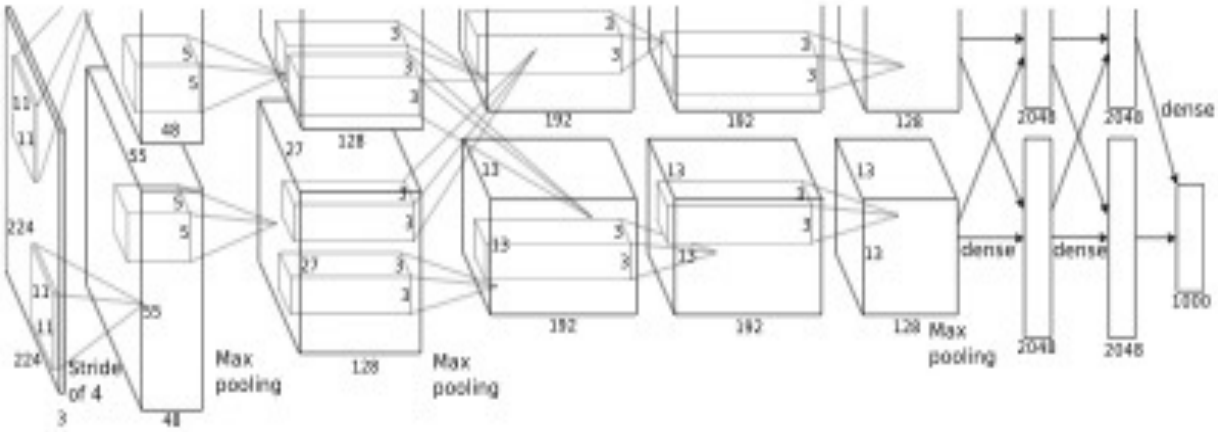
LeNet5 je prva konvolucijska neuralna mreža razvijena 1998. godine od strane LeCunna i Leona Bottou [29]. LeNet5 predviđena je za klasifikaciju rukom pisanih brojeva na poštanskim pošiljkama. Takva mreža imala je 3 konvolucijska sloja, bez slojeva sažimanja maksimumom, s potpuno povezanim slojem na kraju. Značajan je i LeCun-ov rad [17] u kojem se predstavlja arhitektura nazvana LeNet, čije inačice se koriste u mnogim verzijama konvolucijskih mrežama. U početnim se slojevima mreže izmjenjuju slojevi sažimanja maksimalnog odziva i mape značajki. Konkretno, prvi sloj se sastoji od 4 mape značajki, zatim slijedi sloj sažimanja maksimuma, pa sloj od 6 mapi značajki i opet sloj sažimanja maksimuma. Zadnji dio takve mreže je višeslojni perceptron na čije su ulaze spojeni izlazi zadnjeg sloja sažimanja maksimuma. Taj se višeslojni perceptron sastoji od 2 sloja. Prvi je skriveni sloj, a iza njega je sloj logističke regresije. Logistička regresija na kraju čini konačnu klasifikaciju. U konkretnom primjeru postoji 10 izlaza, jedan za svaku znamenku.



Slika 13. Arhitektura LeNet5. Preuzeto iz [29].

4.1.2 AlexNet arhitektura

AlexNet arhitektura pionir je dubokih konvolucijskih neuralnih mreža, predstavljena od strane Krizhevsky *et al.* [28]. AlexNet pobjednik je ILSVRC-2012 sa svojim TOP-5 testom točnosti od čak 84.6%, dok su ostali natjecatelji koristeći tradicionalne tehnike strojnog učenja postigli točnost od 73.8%, u istom izazovu. Arhitektura same mreže u stvarnosti je vrlo jednostavna. Sastoji se od pet konvolucijskih slojeva, slojeva sažimanja maksimalnom vrijednošću, te isto toliku ReLu jedinica, na kraju su tri potpuno povezana sloja. Na slici 5. prikazana je AlexNet arhitektura.

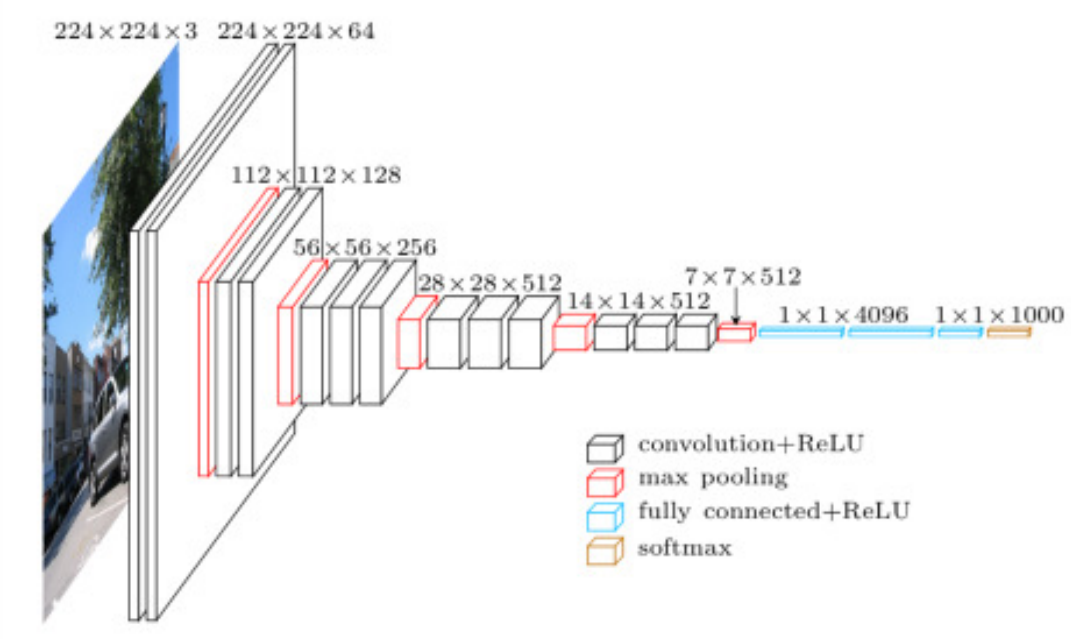


Slika 14. AlexNet arhitektura. Preuzeto iz [28].

4.1.3 VGG arhitektura

VGG (*engl. Visual Geometry Group*) je konvolucijska neuralna mreža predložena od strane K.Simonyan i A. Zissermana [47] sa Sveučilišta u Oxfordu. Predožili su različite modele i konfiguracije dubokih konvolucijskih neuralnih mreža, te jedan od svojih prijedloga prijavili na ILSVRC-2013. Taj model poznat je i pod nazivom VGG – 16, zbog činjenice da je sastavljen od 16 slojeva, postao popularan zbog postizanja TOP – 5 točnosti od čak 92.7%. Slika 15. prikazuje konfiguraciju VGG-16 modela.

Glavna razlika između VGG-16 modela, te njegovih predhodnika je u činjenici da je u prvim slojevima mreže korišteno puno konvolucijskih slojeva s malim receptivnim poljima, za razliku od dotadašnje prakse gdje su se koristila čak tri velika receptivna polja. Ovakav pristup doveo je do smanjena parametra, povećanja nelinearnosti, što se je dovelo do toga da je ovakav model lakše istrenirati.

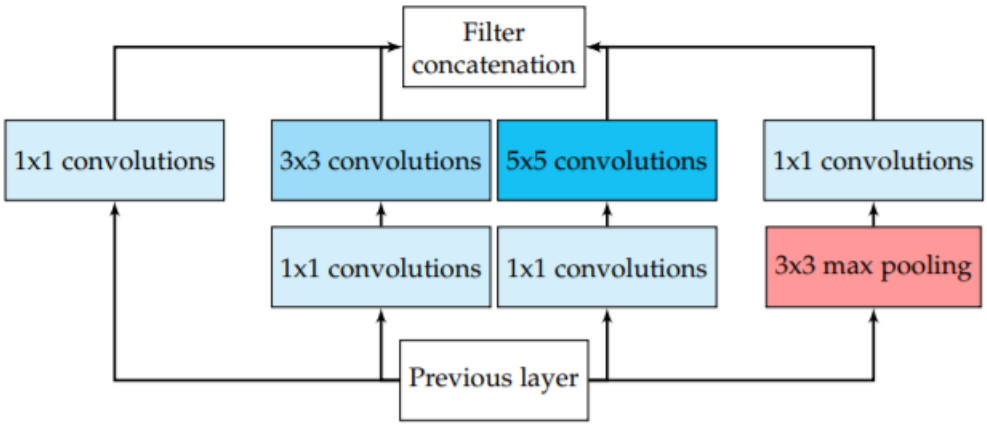


Slika 15. VGG arhitektura. Preuzeto iz [47].

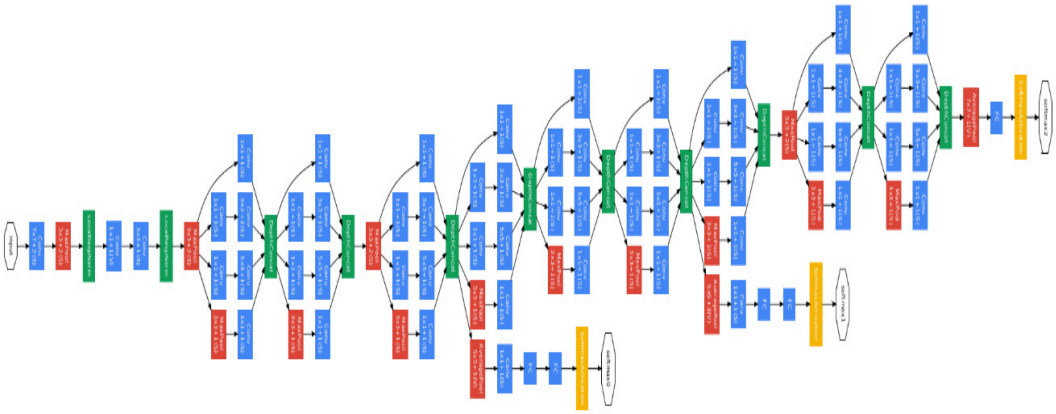
4.1.4 GoogLeNet arhitektura

GoogLeNet je mreža uvedena od strane Szegedy *et al*. [48], te osvojila izazov ILSVRC-2014 s TOP-5 testom točnosti od 93.3%. Ova duboka arhitektura karakteristična je po svojoj složenosti. Sastavljena je čak od 22 sloja, te novog bloka nazvanog početni modul (*engl. inception module*), prikazanog na slici 16. Uvođenjem novog modula Szegedy *et al*. [48] pokazali su da slojevi dubokih konvolucijskih neuralnih mreža mogu biti posloženi na više načina, a ne nužno u sekvecijalnom redu.

Početni modul sastoji se od Network in Network (NiN) sloja, sloja sažimanja, velikog konvolucijskog sloja, te manjeg konvolucijskog sloja. Svaki od ovih slojeva računaju se paralelno, te su popraćeni operacijom konvolucije s $1x1$ filterom da bi se smanjila dimenzija. Zahvaljujući ovakvim modulima ova mreža posebnu pažnju pridaje memoriji, te vremnu potrebom za izračun matričnih operacija, te na taj način smanjuje broj parametara i operacija. Kompletna arhitektura GoogLeNet mreže prikazana je na slici 17.



Slika 16. Početni modul (*engl. inception module*). Preuzeto iz [48].

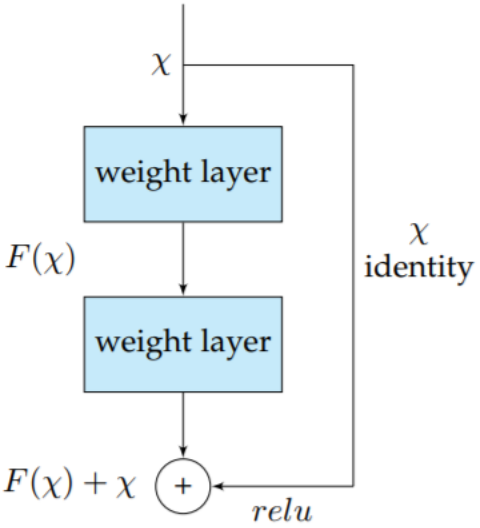


Slika 17. GoogLeNet arhitektura. Preuzeto iz [48].

4.1.5 ResNet arhitektura

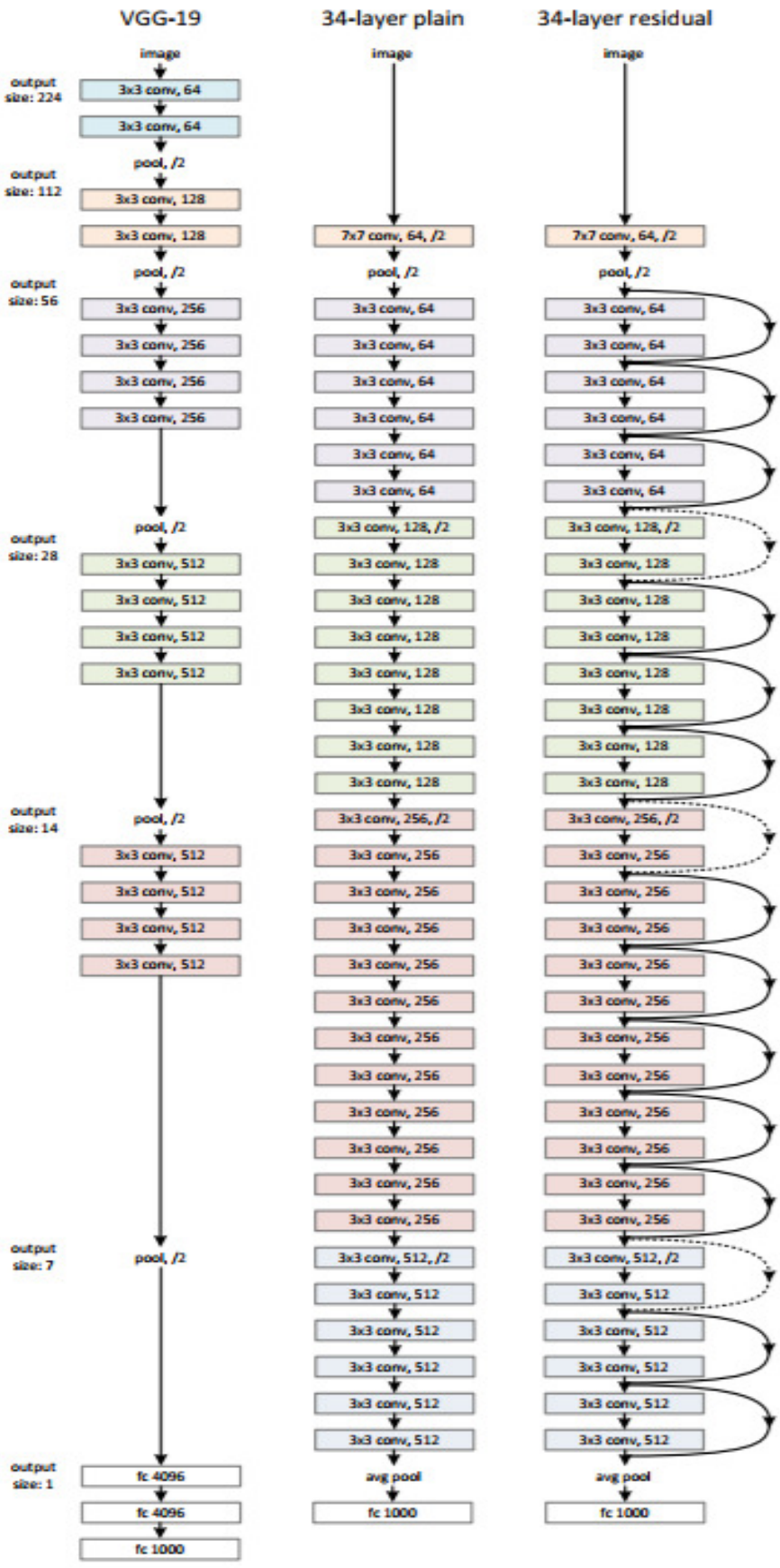
Microsoft-ova ResNet mreža prikazana na slici 19. [50] osvojila je 2016. godine izazov ILSVRC-2016 s 96.4% točnosti. Izuzev te činjenice, ova arhitektura dobro je poznata i po

svojoj dubini (sastoji se čak od 152 sloja), te po uvođenju preostalih blokova (*engl. residual blocks*) prikazanih na slici 18.



Slika 18. Preostali blokovi (engl. residual blocks). Preuzeto iz [50].

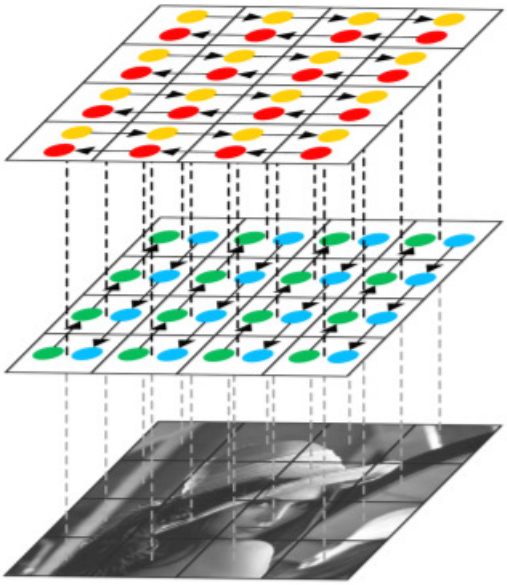
Preostali blokovi riješavaju problem treninga dubokih arhitektura uvodeći veze za preskakanje identiteta, tako da slojevi mogu kopirati ulaze na slijedeći sloj. Ideja iza ovakvog pristupa je da se osigura da slijedeći sloj uči nešto novo i drugačije od onoga što je već kodirano (budući da sloj ima i ulaz i izlaz od predhodnog sloja, te njegov ne promijenjeni ulaz). Ovakve veze među slojevima riješavaju i problem nestajećeg gradijenta.



Slika 19. ResNet arhitektura. Preuzeto iz [50].

4.1.6 ReNet arhitektura

Kako bi se proširile povratne neuronske mreže (RNNs) do više dimenzionalnih zadataka, Graves *et al.* [51], predložili su višedimenzionalnu povratnu neuralnu mrežu (MDRNN). Ova arhitektura zamjenjuje svaku povratnu konekciju iz standardne RNN s d konekcija, gdje je d broj prostorno – vremenskih dimenzija podataka. Zasnovana na ovakvom pristupu Visin *et al*. [52] predložili su ReNet arhitekturu u kojoj se umjesto višedimenzionalnih RNN, koriste uobičajene RNN sekvence. Na ovaj način, broj RNN je linerano skaliran po svakom sloju na broj dimenzija d ulazne slike. U ovom pristupu svaki konvolucijski sloj (konvolucija + sažimanje) zamjenjuje s sa četiri RNN-ova koji se kližu po slici vertikalno i horizontalno u oba smjera kao što je prikazno na slici 20.



Slika 20. Jedan sloj ReNet arhitekture gdje je prikazano modeliranje vertikalnih i horizontalnih prostornih ovisnisti. Preuzeto iz [52].

4.2 Prijenosno učenje

Istrenirati duboku neuralnu mrežu od početka često i nije izvediv zadatak iz više razloga: potrebno je da skup podataka bude dovoljno velik (što nije čest slučaj), te dostizanje konvergencije može trajati predugo. Čak i u slučaju da je skup podataka dovoljno velik, te konvergencija ne potraje predugo, jednostavnije je krenuti od već pre- treniranih težina [80, 81]. Podešavanje težina nastavljajući proces treneriranja s pre-treniranom mrežom jedan je od najčešćih scenarija u prijenosnom učenju.

Yosinski *et al.* [82] dokazali su da je prijenosno učenje bolje od inicijaliziranja težina, te treniranja mreže od početka, čak i za slučajeve kada značajke nisu slične. No sama primjena prijenosnog učenja ponekad nije lagana, na primjer kod korištenja prijenosnog učenja postoje arhitekturalna ograničenja koja moraju biti zadovoljenja kako bi koristili pretreniranu mrežu.

4.3 Pretprocesiranje i povećanje podataka

Povećanje podataka uobičajena je tehnika, koja dokazano ima pozitivne učinke na treniranje dubokih modela za ubrzanje konvergencije ili u ulozi regulatora, što nam služi kako bi se izbjegla pretreniranost mreže i generalizacija [83].

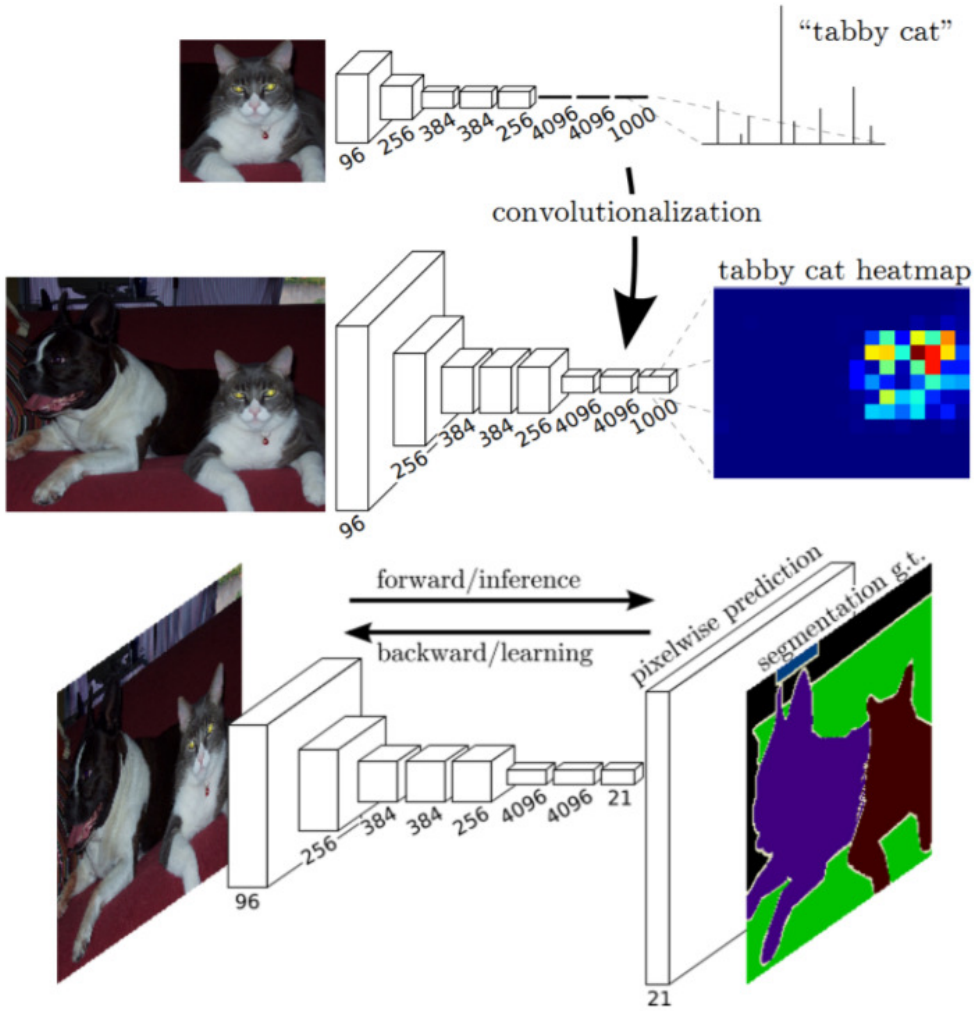
Proces povećanja podataka podrazumijeva primjenu seta transformacija na skup podataka ili na značajke. Najčećše se primjenjuju transformacije na skup podataka, s čime se generiraju novi podaci iz već postojećih. Transformacije koje se koriste u procesu povećanja podataka su translacija, rotacija, zamatanje, skaliranje, mijenjanje prostora boja, rezanje, ... Cilj ovih transformacija je generiranje novih primjera podataka, kako bi se kreirala što veća baza podataka, izbjeglo pretreniranje mreže (odnosno reguliralo model), postigao balans između klasa unutar skupa podataka, pa čak i sintetički stvorilo nove uzorke koji su reprezentativniji za za dati problem.

Povećanje podataka od posebnog je značaja za manje skupove podataka, te u već nekoliko scenarija dokazano efikasno. Tako je u [84], skup podataka od 1500 fotografija portreta sintetički povećan za četiri nove skale (0.6, 0.8, 1.2, 1.5), četiri rotacije (-45, -22, 22, 45), te četiri gamma varijacije (0.5, 0.8, 1.2, 1.5) kako bi se generirala baza podataka od 19000 slika. Ovakim pristupom postignuta je veća točnost njihovog modela za segmentaciju portreta sa 73.09 % na 94.2 %.

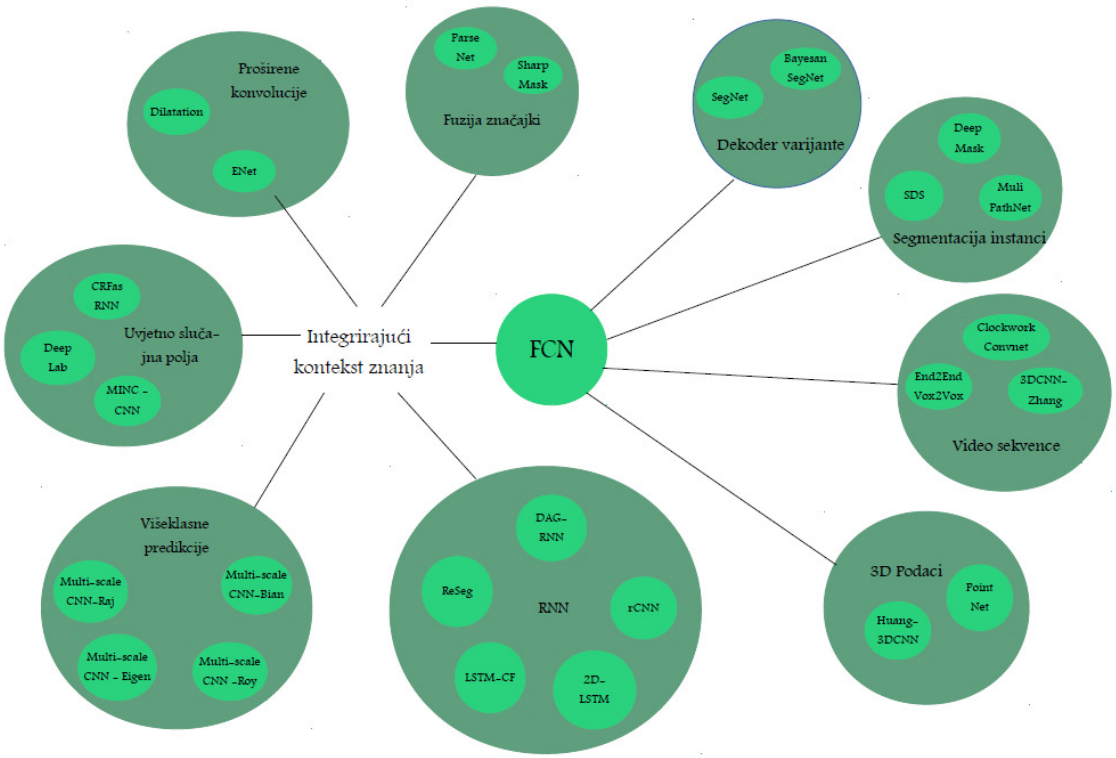
5 Metode semantičke segmentacije slike korištenjem dubokih konvolucijskih neuralnih mreža

Konstantni uspjeh dubokih arhitektura u raznim zahtjevnim zadatcima računalnog vida, posebice pristupa konvolucijskih neuralnih mreža (CNNs) za klasifikaciju slike ili detekciju objekata [67, 68, 69] motiviralo je istraživače da istraže sposobnosti ovakvih mreža za probleme laberiranja na razini piksela, poznate pod nazivom semantička segmentacija. Glavna prednost dubokih tehnika učenja nad tradicionalnim tehnikama prepoznavanja objekata je sposobnost učenja prikaza značajki „u hodu“, odnosno predstavljaju pristup u kojemu se sve (od znacajki do klasifikacije), u potpunosti, uči automatski na temelju skupa uzoraka. Pozitivne strane takvog pristupa uključuju prilagođenost naučenih značajki konkretnom problemu i njegovom skupu uzoraka, dijeljenje značajki između više klasa te učenje različitih značajki za različite modalnosti pojedinih klasa. Trenutno, jedna od najuspješnijih tehnika dubokog učenja za semantičku segmentaciju slike je „Fully Convolution Network“ (FCN) objavljena od strane *Long et al.* [70]. Njihov pristup temelji se na već postojećim temeljima klasične CNNs proširene na sposobnost učenja hierahije značajki. Uspješno su transformirali dobro znane modele za klasifikaciju – AlexNet [28], VGG (16 slojeva) [47], GoogLeNet [48] i ResNet [50] u jednu potpuno konvolucijsku mrežu tako da su zadnje potpuno povezane slojeve zamijenili s konvolcijskim slojevima u svrhu dobivanja prostorne mape, umjesto rezultat klasifikacije. Dobivene mape uzorkovali su koristeći frakcionirane konvolucije (poznate kao dekonvolucija [70, 71]) kako bi dobili izlaze označene po pikselima. Ovaj rad označava prekretnicu u dubokom učenju, jer dokazuje da CNNs mogu služiti i za ovakav tip problema, te učinkovito naučiti kako napraviti gusta predviđanja s ulazima proizvoljnih veličina. Također, postignut je i značajan napredak u točnosti segmentacije u odnosu na tradicionalne metode. Zbog svega navedenog, te mnogih drugih doprinosa FCN smatra se središtem dubokog učenja za semantičku segmentaciju slike. Slika 21. prikazuje proces FCN-a.

No usprkost svojoj snazi i fleksibilnosti FCN model i dalje ima poneke nedostatke koji ometaju njegove aplikacije u određenim situacijama.



Slika 21. Potpuna konvolucijska mreža FCN. Preuzeto iz *Long et al.* [70].



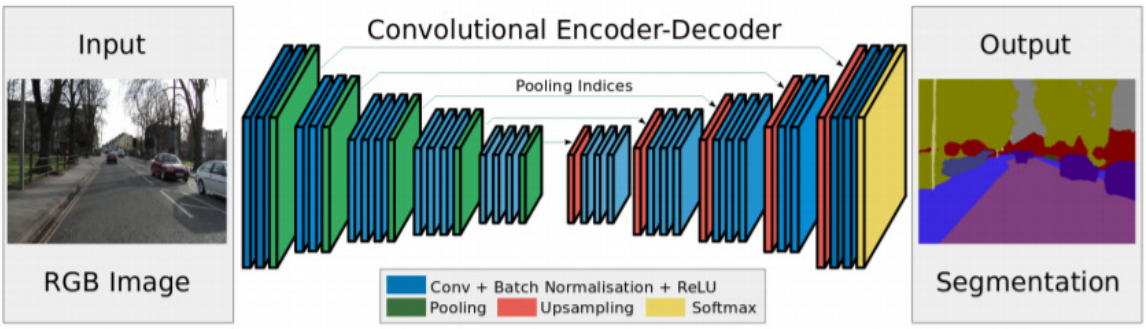
Slika 22. Vizualizacija istaženih metoda. Prilagođeno i preuzeto iz [].

5.1 Varijante dekodera

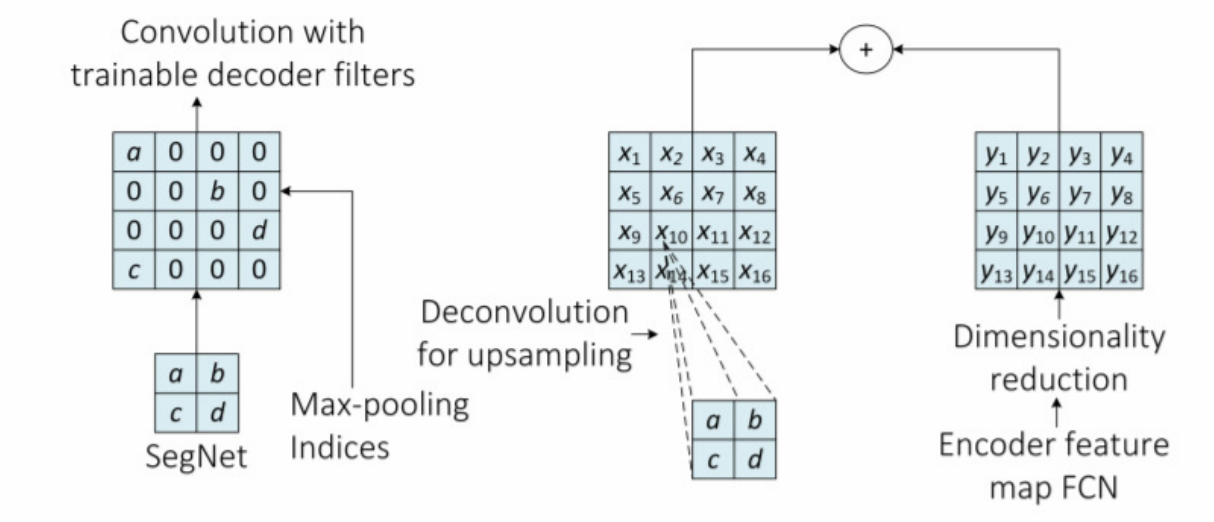
Osim FCN arhitekture, postoje i druge varijante dubokih arhitektura, razvijene s ciljem transformacije mreže čija je izvorna namjena bila klasifikacija slike, u mrežu pogodnu za segmentaciju. Nedvojbeno, FCN arhitektura i dalje je jedna od najpopularnijih, što ne znači da i su druge opcije loše za različite namjene. Općenito, svi kreću od iste ideje, a to je preuzeti mrežu za klasifikaciju, kao što je VGG -16, te ukloniti njene potpuno povezane slojeve. Ovaj dio novo nastale mreže naziva se encoder i proizvodi mape značajki ili reprezentaciju slike niske rezolucije. Glavni problem javlja se u dijelu kada je potrebno naučiti kako dekodirati ili mapirati slike niske rezolucije u predikcije po pikselima za segmentaciju. Ovaj dio procesa naziva se dekoder i obično predstavlja točku divergencije u dubokim arhitekturama ovog tipa.

SegNet [73] je odličan primjer divergencije (Slika 23). Dekoder faza SegNet arhitekture sastavljena je od seta slojeva uzorkovanja, te konvolucijskih slojeva iza kojih slijedi softmax klasifikator koji predviđa oznaku piksela za izlaz koji ima istu rezoluciju kao ulazna slika. Svaki sloj uzorkovanja u dekoder fazi odgovara sloju sažimanja u enkoder fazi. Nad mapama koje su dobivene uzorkovanjem vrši se operacija konvolucije sa setom istreniranih filtera u svrhu dobivanja gustih značajki. U poslijednjem koraku mapa značajki vraća se u izvornu rezoluciju, te se prosljeđuje softmax klasifikatoru kako bi se dobila konačna semantička

segmentacija. Arhitekture bazirane na FCN modelu koriste naučene dekonvolucijske filtre kako bi povećali mape značajki. Nakon toga povećane mape značajki dodaju se jedna po jedna odgovarajućoj mapi značajki generiranoj u konvolucijskom sloju u encoder dijelu modela. Slika 24. prikazuje usporedbu oba pristupa.



Slika 23. Prikaz SegNet mreže. Preuzeto iz [73].



Slika 24. Usporedba decoder faze SegNet arhiteture, te FCN arhitekture. Preuzeto iz [53].

5.2 Integriranje znanja o kontekstu

Semantička segmentacija je problem koji zahtjeva integraciju informacija iz različitih prostornih mjerila. Također, podrazumijeva balans između lokalnih i globalnih informacija. Na jednu stranu, lokalne informacije ključne su za postizanje dobre točnosti na razini piksela, dok s druge strane važno je i integrirati informacije iz globalnog konteksta slike kako bi što bolje riješili lokalne nejasnoće. Vanilla CNN bori se s ovim balansom. Slojevi sažimanja, koji dopuštaju mreži da postigne stupanj prostorne invariancije i zadrži troškove komputacije u razumnim okvirima, pritom raspolažući s globalnim informacijama. Čak i originalna CNN

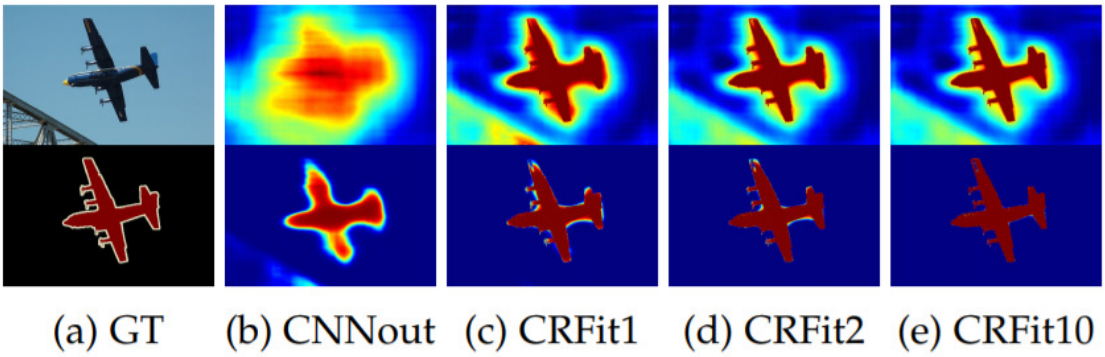
arhitektura, bez slojeva sažimanja, su ograničeni s obzirom da receptivno polje njihovih jedinica može rasti samo linearno s brojem slojeva.

Moguće je osvijestiti CNN o globalnim informacijama: precizranje kao korak naknadne obrade CFR-a, proširene konvolucije, višeskalarne agregacije ili čak odgoda modeliranja konteksta na drugu vrstu dubokih arhitektura kao što je RNN.

5.2.1 Uvjetna slučajna polja (CRF)

Kao što je spomenuto i u prijašnjem poglavlju, inherentna invarijanca na prostornu transformaciju kod CNN arhitektura ograničava prostornu točnost kod segmentacije. Jedan od uobičajenih pristupa kod preciziranja izlaza segmentacije, te poboljšanja sposobnosti samog sistema za detektiranje sitnih detalja je uvođenje koraka post- procesiranja pomoću uvijetnih slučajnih polja (*engl. Conditional Random Field CRF*). CRF omogućava kombiniranje informacija niske razine (npr. interakcija između piksela [74, 75]) s više klasnim sustavom koji određuje koji piksel spada u koju klasu. S ovakvom kombinacijom postignute su odlične performanse za ovisnosti velikog ranga, koje klasična CNN mreža ne uzima u obzir, kao i za sitne detalje sa slike.

DeepLab model [75, 76] koristi potpuno povezane parove CRF [78, 79], kao odvojeni korak procesiranja kako bi poboljšali rezultat segmentacije. Problem korištenja CRF kao koraka post-procesiranja je dugotrajno izvođenje algoritma.

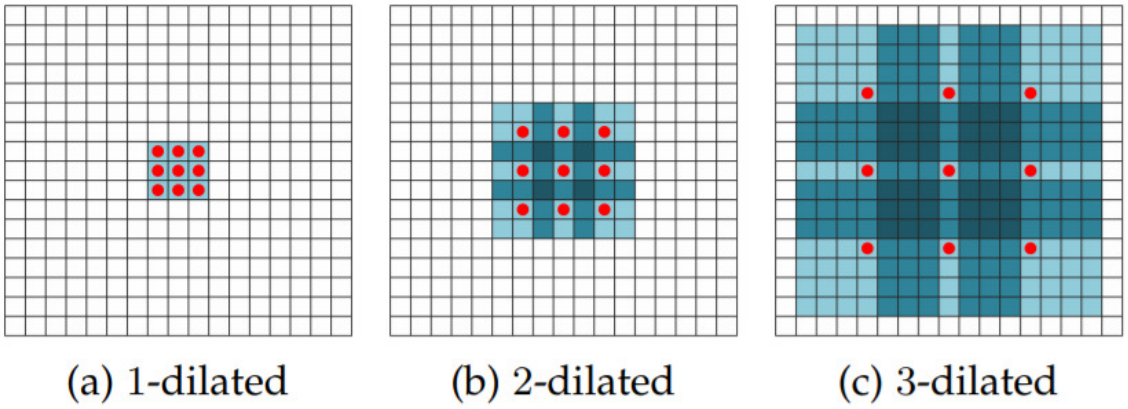


Slika 25. Uglađivanje izlazne segmentacijeske mape korištenjem CRF (prikaz po iteracijama). Prvi red prikazuje izlazne mape prije primjene softmax funkcije, dok drugi red prikazuje izlaz softmax funkcije. Preuzeto iz [75].

5.2.2 Proširene konvolucije

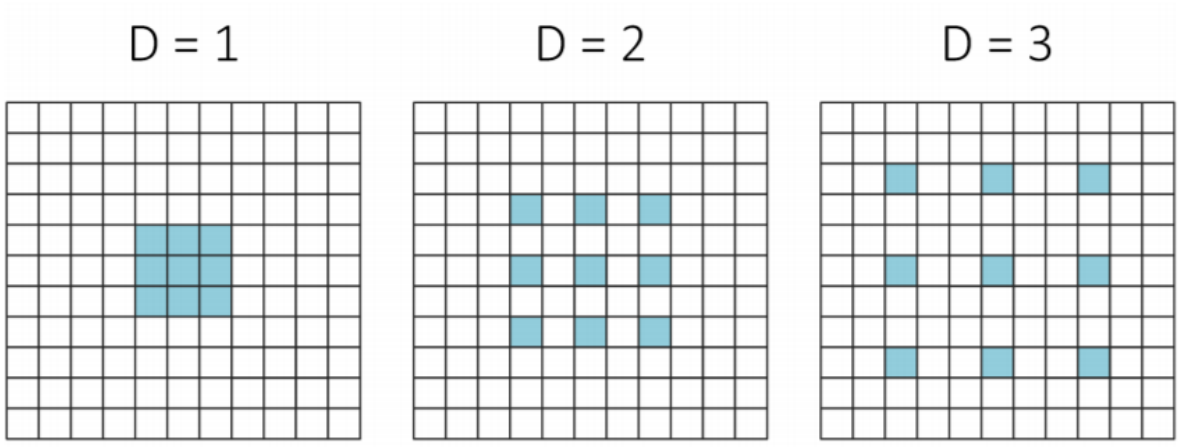
Proširene konvolucije, poznate pod imenom *à-trous* konvolucija, zapravo je generalizacija Keonecker-factored konvolucijskog filtra [85], koji podržava eksponencijalnu ekspanziju receptivnog polja bez gubitka rezolucije. Drugim riječima, proširene konvolucije, standardni

su tip konvolucija koje koriste filtere za uzorkovanje. Brzina dilatacije l kontrolira faktor uzorkovanja. Slika 26. prikazuje proširenu konvoluciju na 2D podacima. Crvene točkice predstavljaju ulaze filtera $3x3$, a područja obojanih rešetki predstavljaju receptivna polja. Kao što je vidljivo iz prikaza na slici receptivna polja rastu eksponencijalno, dok broj parametara po filteru zadržava linearni rast. Ovakve značajke proširene konvolucije omogućuju nam učinkovitu ekstrakciju gustih značajki u bilo kojoj rezoluciji.



Slika 26. a) 1- dilatacijska konvolucija, gdje svaka jedinica ima 3x3 receptivno polje, b) 2-dilatacijska konvolucija s $\mathbf{7x7}$ receptivnim poljem, c) 3-dilatacijska konvolucija s $\mathbf{15x15}$ receptivnim poljem. Preuzeto iz [86].

U praksi, isti efekt postići ćemo ako proširimo filter prije operacije konvolucije. Proširenje filtra ovisno o veličini proširenja ispuniti će prazne elemente s nulama. Slika 27. prikazuje primjer proširene konvolucije.



Slika 27. Elementi filtera (plavi) usklađeni s ulaznim elementima koristeći $\mathbf{3x3}$ proširenu konvoluciju s različitim stupnjevima. Od lijeva na desno 1,2 i 3. Preuzeto iz [53].

Najvažniji radovi u kojima je korišten princip proširene konvolucije su od *Yu el al.* [86], već spomenuti DeepLab (ovog puta njihova poboljšana verzija) [76], te mreža u realnom vremenu

Enet [87]. Svi oni koriste kombinaciju proširenih konvolucija s povećanjem dilatacije kako bi dobili čišće receptivno polje, bez dodatnih troškova i bez gubljenja podataka u mapi značajki. Ovi radovi pokazali su da je proširena konvolucija usko povezana s višeskalarnim predikcijama, što je detaljnije objašnjeno u idućem poglavlju.

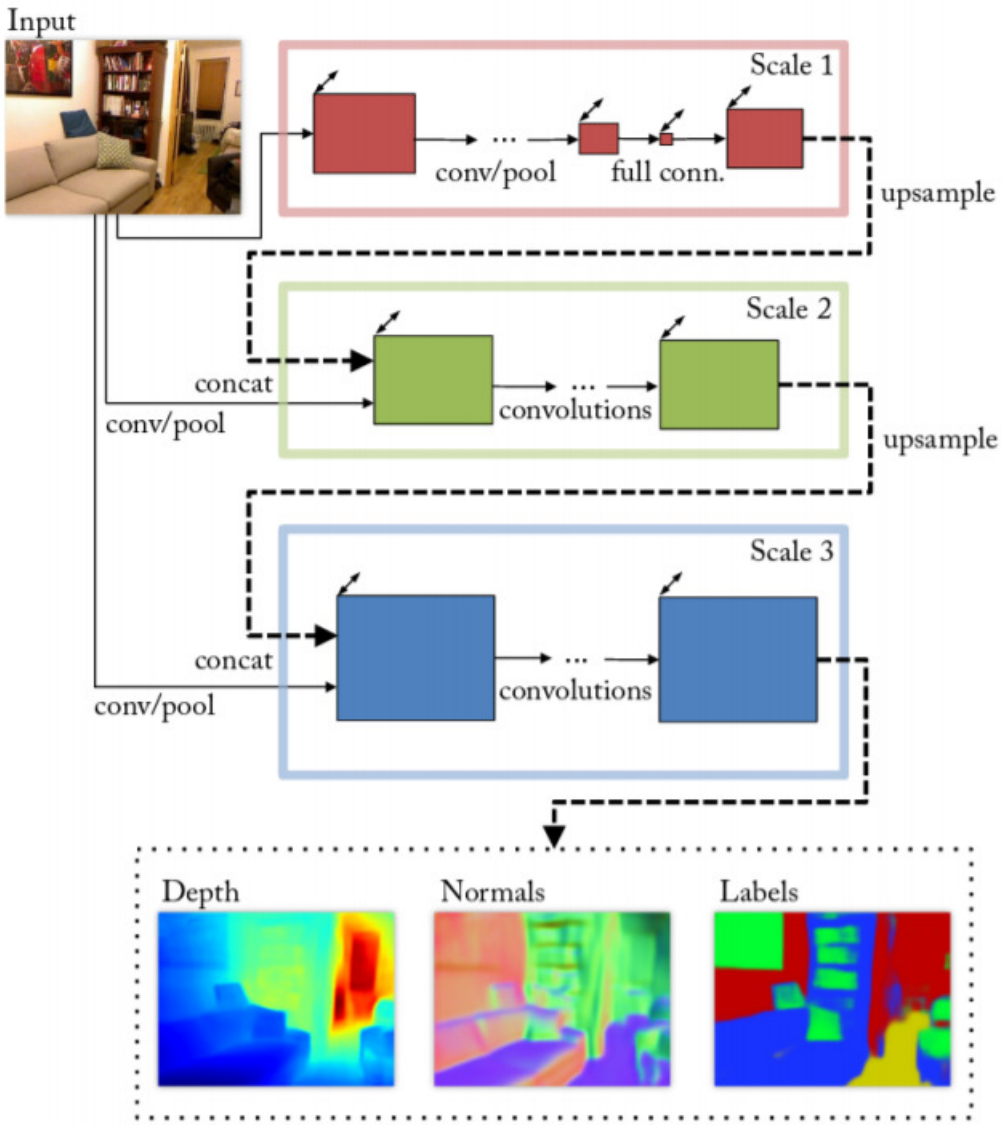
5.2.3 Višeskalarne predikcije

Višeskalarne predikcije jedan su od načina riješavanja problema integriranja znanja o kontekstu. Gotovo svaki parametar CNN mreže utječe na skalu generirane mape značajki. Drugim riječima, ista arhitektura će imati utjecaj na broj piksela ulazne slike, koja odgovara pikselu iz mape značajki. Ovo bi značilo da filter implicitno uči detektirati značajke u određenim mjerilima (s određenim stupnjom invarijancije).

Raj *et al.* [89] predložili su višeskalarnu verziju potpuno povezane konvolucijske mreže VGG-16. Njihova mreža sadrži dva dijela. Prvi dio procesira ulaznu sliku u originalnoj rezoluciji, dok drugi dio poduplava rezoluciju ulazne slike. U prvom dijelu mreža je plitka , dok je u drugom dijelu potpuno povezana VGG -16 s dodanim konvolucijskim slojevima. Rezultat drugog dijela modela uzorkovan je i kombiniran s rezultatom prvog dijela mreže. Kombinirani izlaz ova dva dijela prolazi kroz set konvolucijskih slojeva, kako bi se generirao završni izlaz mreže. Rezultat ovakvog pristupa je robusna mreža otporna na varijacije u skali.

Roy *et al.* [90] drugačije su pristupili ovom problemu, koristeći mrežu sastavljenu od 4 višeskalarne CNN mreže. Ove četiri mreže imaju arhitekturu identičnu onoj predlozženoj od strane Eigen *et al.* [88]. Jedna od četiri mreže zadužena je za pronalazak semantičkih oznaka scene. Ova mreža izvlači značajke iz sekvence različitih skala (Slika 28.).

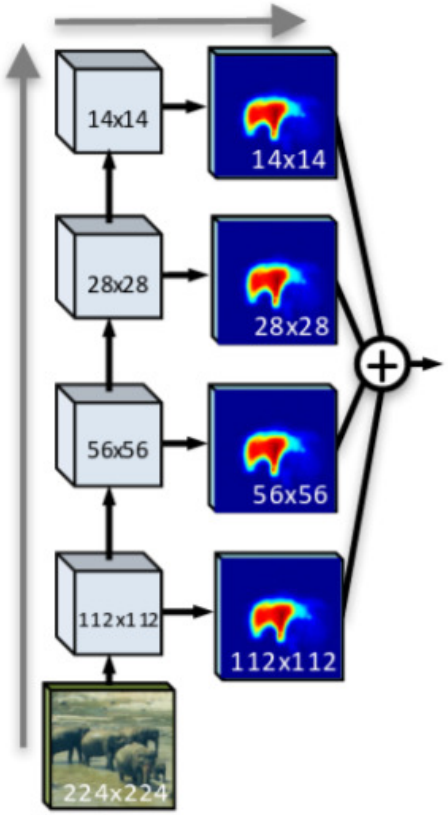
Još jedan zapaženi rad je mreža predložena od strane Bien *et al.* [91]. Njihova predložena mreža sastavljena je od n FCN mreža koje rade s različitim skalama. Značajke koje su izvučene iz svake od mreža stapaju se, te se provlače kroz dodatni konvolucijski sloj, kako bi se dobila završna segmentacija. Glavni doprinos njihove predložene arhitekture je proces učenja u dvije faze koji uključuje treniranje svake mreže posebno, a nakon toga kombiniranje mreža, te korištenje prijenosnog učenja na zadnjem sloju. Ovaj višeskalarni model dopušta dodavanje proizvoljnog broja novo istreniranih mreža.



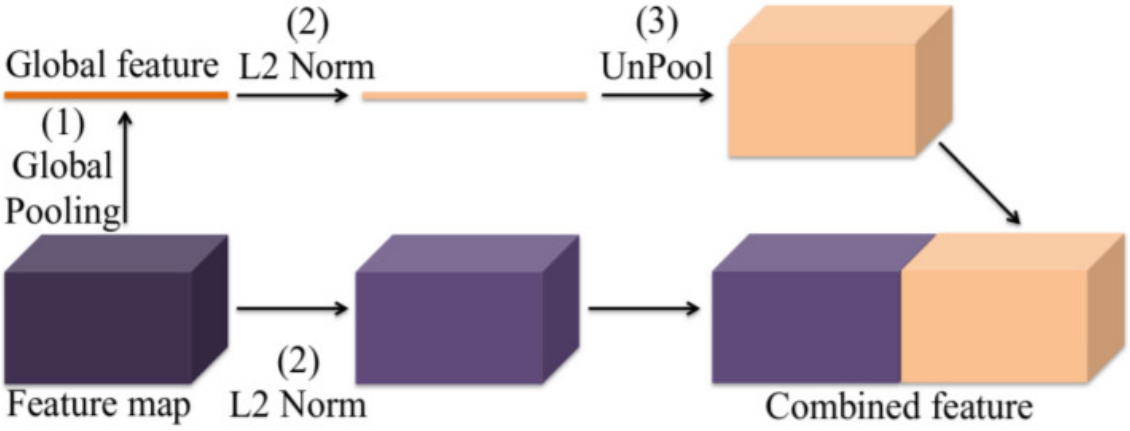
Slika 28. Višeskalarna CNN arhitektura (Eigen *et al.* [88]). Mreža progresivno pročišćava izlaz pomoću sekvence skala kako bi procjenila dubinu, normalu, te obavlja semantičku segmentaciju preko RGB ulaza. Preuzeto iz [88].

5.2.4 Fuzija značajki

Fuzija značajki je tehnika koja podrazumijeva spajanje globalnih značajki (izvučenih iz predhodnog sloja mreže) s lokalnim značajkama izvučenim iz slojeva koji slijede, te predstavlja jedan od načina integracije konteksta o informaciji kod arhitektura namjenjenih za segmentaciju. Uobičajene arhitekture kao što je FCN koriste prekinute konekcije kako bi izveli fuziju , tako da kombiniraju mape značajki iz različitih slojeva (Slika 29.). Još jedan od mogućih pristupa je ranija fuzija. Pristup ranije fuzije korišten je u ParseNet [92] mreži u konteksnom modulu. Tamo su globalne značajke sažete na istu prostornu veličinu lokalnih značajki, te povezane kako bi se generirale kombinirane značajke koje se koriste u idućem sloju ili kako bi se istrenirao klasifikator. Slika 30. prikazuje ovaj proces.



Slika 29. Prikaz arhitekture koja izvodi kasniju fuziju mape značajki, na način da prvo daje predikcije neovisno, a nakon toga izlazi se spajaju, te se donosi odluka o završnoj segmentaciji. Preuzeto iz [92].



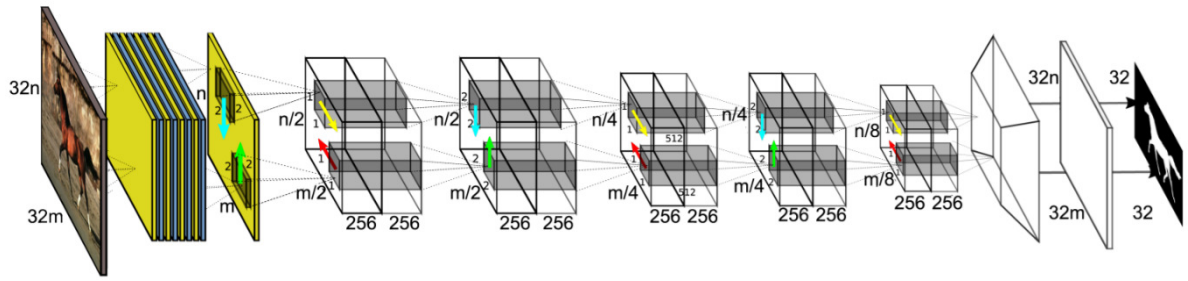
Slika 30. ParseNet-ov kontekst modul u kojem se globalne značajke iz predhodnog sloja kombiniraju s značajkama idućeg. Preuzeto iz [92].

Ideja fuzije značajki nastavlja se radom od strane Pinheiro *et al.* u mreži nazvanoj SharpMask [93], koja koristi modul za uvođenje značajki iz predhodnog sloja u slijedeći. Detaljniji opis dat je u poglavlju o segmentaciji instanci slike.

5.2.5 Povratne Neuralne Mreže

CNN uspješno su primjenjene na višedimenzionalne podatke, kao što su slike. No usprkos tome, ovakav tip mreža oslanja se na specifične kernele, čime je arhitektura limitirana za riješavanje problema segmentacije.

Po uzoru na ReNet model za klasifikaciju slike Visin *et al.* [94] predložili su arhitekturu za semantičku segmentaciju nazvanu ReSeg [95] prikazanu na Slici 31. U ovakvom pristupu, ulazna slika obrađuje se u prvom sloju VGG-16 mreže, te se dobivena mapa značajki prosljeđuje u jedan ili više ReNet slojeva za prijenosno učenje. Na kraju, dobivena mapa značajki se smanjuje pomoću sloja sažimanja temeljem transponirane konvolucije. U njihovom pristupu korištene su GRU jedinice budući da postižu odlične performace u balansiranju korištenja memorije i računalne moći. Vanilla RNN imaju problem modeliranja dugoročnih ovisnosti, većinom zbog problema nestajućeg gradijenta. Nekoliko ivedenih modela kao što su LSTM [96] mreža, te GRU [97] uspješno izbjegavaju ovaj problem.



Slika 31. Prikaz ReSeg mreže. VGG-16 konvolucijski slojevi prikazani su žutom i plavom bojom.Ostatak arhitekture bazira se na ReNet arhitekturi. Preuzeto iz [95].

Inspirirani istom ReNet arhitekturom Li *et al.* [98], predložili su novi LSTM-CF model za označavanje scena. Njihov pristup koristi dva različita izvora podataka: RGB i dubinu. RGB izvor oslanja se na varijantu DeepLab arhitekture [76] povezujući značajke na tri različite skale kako bi obogatili značajke (inspirirani radom [99]).

Modelirane globalnog konteksta slike povezano je s 2D pristupom tako što se mreža razvija vertikalno i horizontalno preko ulazne slike. Na temelju ove činjenice, Byeon *et al.* [100]

predložili su jednostavnu 2D LSTM arhitekturu u kojoj se ulazna slika dijeli na ne preklapajuće prozore, koji se dalje šalju u četiri odvojena LSTMs memorijska bloka. U ovom radu daje se naglasak na računalnu jednostavnost koristeći jedno jezgreni procesor, te na jednostavnosti samog modela.

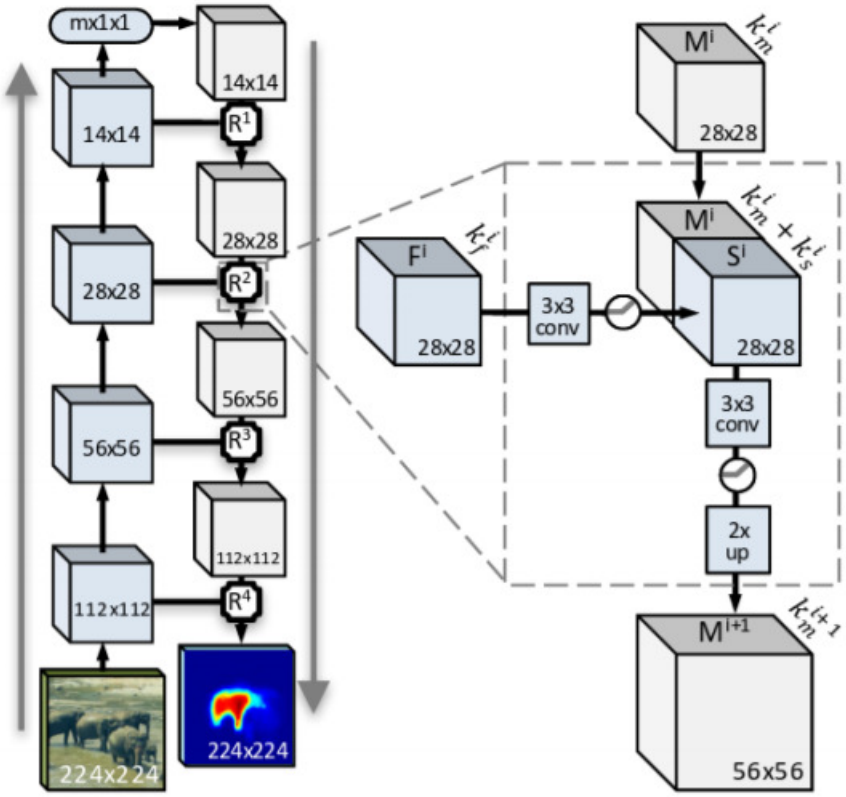
Još jedan pristup za hvatanje globalnog konteksta slike oslanja se na korištenje većeg prozora, kako bi se modelirao veći kontekst. Ovakvim pristupom smanjuje se rezolucija slike, te se javljaju problemi vezani za preklapanje prozora. Unatoč ovom problemu, Pinheiro *et al.* [101] predstavili su rCNN koja vrši treniranje s različitim veličinama prozora, uzimajući u obzir predhodne predikcije dobivene različitim veličinama prozora. Ovakvim pristupom, labele se automatski uglađuju čime se povećavaju performanse.

Neusmjereni ciklički grafovi (*eng. Undirected cyclic graphs* (UCGs)) prihvaćeni za modeliranje konteksta slike za semantičku segmentaciju [102]. RNN nisu direktno povezani s UCG i samo riješenje je razdijeljeno u više usmjerenih grafova (DAGs), u ovakvom pristupu slike se obrađuju u tri različita sloja: mapa značajki slike dobiva se preko CNN mreže, modeliranje konteksta slike dobiva se preko DAG-RNNs i dekonvolucijski sloj koristi se za sažimanje mape značajki. Ovim radom prikazano je kako se RNN mreža može koristiti zajedno s grafom kako bi se uspjepno modelirao kontekst slike.

5.2 Segmentacija instanci slike

Segmentacija instanci prirodni je slijed semantičke segmentacije, te ujedno predstavlja jedan od najvećih izazova u usporedbi s ostalim segmentacijskim tehnikima. Cilj segmentacije instanci je dobiti prikaz objekata iste klase podijeljene u različite instance. Automatizacija ovog procesa nije jednostavna, jer broj instanci nije unaprijed poznat, te evaluacija dobivenih instanci nije bazirana na pikselima kao što je bio slučaj kod semantičke segmentacije. Segmentacije instanci slike ne istraženo je područje, no interes je motiviran mogućnošću primjene u praksi. Označavanje instanci pruža nam dodatne informacije za zaključivanje nepoznatih situacija, za brojanje elemenata koji pripadaju istoj klasi, te za detekciju određenih objekata koje treba dohvatiti u robotskim zadacima.

Iz gore navedenih razloga Hariharan *et al.* [103] predložili su SDS (*engl. Simultaneous Detection and Segmentation*) metodu kako bi popravili performanse već postojećih modela. Njihov pristup bazira se na segmentaciji slike, a zatim detekcija kandidata koji pripadaju istoj klasi. Oko detektiranih kandidata izdvajaju regije, koje se dalje prosljeđuju u adaptiranu verziju R-CNN [93].



Slika 32. SharpMask's arhitektura. Preuzeto iz [104].

6 Zaključak

U nastojanju pronalaska idealnog rješenja za semantičku segmentaciju slika metodama dubokog učenja u ovom radu dat je pregled postojećih metoda. Objašnjenji su osnovni koncepti dubokog učenja, te navedene osnovne duboke arhitekture za klasifikaciju slike, kao i izvedene arhitekture za semantičku segmentaciju prirodnih scena. No, unatoč velikom broju radova u području dubokog učenja semantička segmentacija slike, nije do kraja istražena, te još ne postoji arhitektura koja je u ovom zadatku nadmašila ostale. U ovom poglavlju data je kratka usporedba postojećih arhitektura, njihovih prednosti i nedostataka.

Standardni pristupi računalnog vida za razumijevanje slike na razini piksela većinom su se svodili na TextonForest i Random Forest klasifikatore. Nakon što su CNN arhitekture postigle zavidan uspjeh u klasifikacija slike, malom modifikacijom osnovne CNN arhitekture počele su se koristiti i za semantičku segmentaciju. Jedan od prvih pristupa dubokog učenja za problem semantičke segmentacije je klasifikacija dijelovima. Svaki piksel klasificirao se u klasu koristeći dijelove slike oko njega. Razlog korištenja dijelova slike su upravo potpuno povezani slojevi mreže, koji zahtjevaju fiksnu veličinu ulazne slike. FCN arhitektura upravo je iz ovog razloga u svojoj mreži uklonila potpuno povezane slojeve, čime je problem fiksne veličine ulazne slike uklonjen. No, izuzev potpuno povezanih slojeva jedan od glavnih problema korištenja CNN arhitekture za semantičku segmentaciju upravo su slojevi sažimanja. Slojevi sažimanja odbacuju informacije o slici, ali i gube informacije o točnoj poziciji piksela. Točna pozicija piksela od presudne je važnosti za problem semantičke segmentacije. Dva su pristupa rješenju ovog problema: encoder- dekoder arhitekture, te arhitekture koje koriste proširene konvolucije. Enkoder-dekoder arhitekture u enkoder fazi smanjuju prostornu dimenziju slike koristeći slojeve sažimanja, dok u dekoder fazi obnavljaju detalje pronađenih objekata, te vraćaju sliku na prvobitnu dimenziju. Ovakvi tipovi koriste konekcije s enkoderom kako bi što uspješnije obnovili detlje slike u dekoder fazi. Problemi ovih arhitektura su male preciznosti segmentacije. Drugi tip arhtekture za semantičku segmentaciju slike koriste proširene konvolucije, te ne koriste slojeve sažimanja. Problem u ovakovom pristupu su izlazne segmentirane mape, čija je veličina 1/8 od stvarne slike. Osim korištenja jedne od ove dvije arhitekture, u literturi su korištena i uvjetna slučajna polja kao korak postprocesiranja. Primjenom uvjetni slučajnih polja završna segmentacijska mapa uglađuje se po pretpostavci da pikseli sličnog intenziteta spadaju u istu klasu. Korak

postprocesiranja usporava cijeli proces segmentacije, uz malo povećanje preciznosti od 1-2%. U Tablici 1. sumirane su najvažnije arhitekture za semantičku segmentaciju slike.

Tablica 1. Usporedba najvažnijih arhitektura za semantičku segmentaciju.

Naziv	Arhitektura	Preciznost	Efikasnost	Kontribucije	Nedostatci
Fully Connected Network	VGG-16 (FCN)	**62.2%**	*	Prvi rad u području.	Izlazne segmentirane mape su „grube“.
SegNet	VGG-16 + Decoder	**59.9 %**	**	Encoder-decoder	Preciznost ove arhitekture nije zadovoljavajuća.
Dilatation	VGG-16	**73.5%**	*	Proširene konvolucije	Segmentacijska mapa je 1/8 od ulazne slike.
DeepLab v1&v2	VGG-16/ResNet-101	**79.7%**	*	Proširene konvolucije + CRF	Segmentacijska mapa je 1/8 od ulazne slike.
DeepLab v3	VGG-16/ResNet-101	**85.7%**	**	Proširene konvolucije + CRF + ASAP	Segmentacijska mapa je 1/8 od ulazne slike.

Na problemima semantičke segmentacije slike metodama dubokog učenja se i dalje intenzivno radi, kako bi se povečala preciznost segmentacije uz zadovoljavajuću veličinu segmentacijske mape ali i rješili probleme gubljenja informacija koje uvode slojevi sažimanja. U tom smjeru se vidi i budući rad kandidatkinje, s jedne strane istraživanja razvoja što preciznijih arhitektura koje bi posebno bile prilagođene semantičkoj segmentacija slika prirodnog krajolika. Ovo je područje do sada uglavnom bilo usmjereno prema primjeni kod autonomne vožnje vozila, dok su nama istraživački ciljevi semantička segmentacija slika snimljenih letećim, lebdećim i stacionarnim sustavima motrenja i nadzora nepristupačnih terena prirodnog krajolika, s obzirom da je budući usmjeren prema zadacima kognitivnog vida inteligentnih observera prirodnog krajolika.

LITERATURA

[1] Baldi, P. and Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. Neural Networks, 2:53–58.

[2] Baldi, P. and Hornik, K. (1994). Learning in linear networks: a survey. IEEE Transactions on Neural Networks, 6(4):837–858. 1995.

[3] Barlow, H. B. (1989). Unsupervised learning. Neural Computation, 1(3):295–311.

[4] Becker, S. (1991). Unsupervised learning procedures for neural networks. International Journal of Neural Systems, 2(1 & 2):17–33.

[5] https://www.xenonstack.com/blog/overview-of-artificial-neural-networks-and-its-applications.

[6] Bishop, C. M. (1993). Curvature-driven smoothing: A learning algorithm for feed-forward networks. IEEE Transactions on Neural Networks, 4(5):882–884.

[7] https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/.

[8] https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/.

[9] https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html.

[10] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities,Proc. Natl. Acad. Sci. USA 79:2554 (1982).

[11] S. Gupta, R. Girshick, P. Arbelaez, and J. Malik, "Learning rich ´ features from rgb-d images for object detection and segmentation," in European Conference on Computer Vision. Springer, 2014, pp. 345–360.

[12] H. Zhu, F. Meng, J. Cai, and S. Lu, "Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," Journal of Visual Communication and Image Representation, vol. 34, pp. 12 – 27, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320315002035

[13] M. Thoma, "A survey of semantic segmentation," CoRR, vol. abs/1602.06541, 2016. [Online]. Available: http://arxiv.org/abs/ 1602.06541

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classifi- cation with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[18] A. Graves, S. Fernandez, and J. Schmidhuber, "Multi- ´ dimensional recurrent neural networks," CoRR, vol. abs/0705.2011, 2007. [Online]. Available: http://arxiv.org/ abs/0705.2011

[19] Hinton, G. E., Dayan, P., Frey, B. J., Neal, R. M. The" wake-sleep" algorithm for unsupervised neural networks. Science 268, 5214 (1995), 1158– 1161.

[20] Hinton, G. E., Osinero, S., Teh, Y.-W. A fast learning algorithm for deep belief nets. Neural computation 18, 7 (2006), 1527–1554.

[21] Hinton, G. E., Srivastava, N., Krizehevsky, A., Sutskever, I., AND Salakhtudinov, R. R. Improving neural networks by preventing coadaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012).

[22]] C. Liu, J. Yuen, and A. Torralba, "Nonparametric scene parsing: Label transfer via dense scene alignment," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 1972–1979.

[23] Hochreiter, Schmidhuber, J. Long short-term memory. Neural computation 9, 8 (1997), 1735–1780.

[24] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences 79, 8 (1982), 2554–2558.

[25] Hornik, K., Stinchcombe, M., White, H. Multilayer feedforward networks are universal approximators. Neural networks 2, 5 (1989), 359–366.

[26] Kirpatrick, S., Gelatt, C. D., Vecchi, M. P., ET AL. Optimization by simulated annealing. science 220, 4598 (1983), 671–680.

[27] Kohonen, T. The self-organizing map. Neurocomputing 21, 1 (1998), 1–6.

[28] Krizehevsky, A., Sutskever, I., Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (2012), pp. 1097–1105.

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

[30] Lee, H., Grosse, R., Rangata, R., NG, A. Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th Annual International Conference on Machine Learning (2009), ACM, pp. 609–616. [31] Liou, C.-Y., Cheng, W.-C., Liou, J.-W., Liou, D.-R. Autoencoder for words. Neurocomputing 139 (2014), 84–96.

[32] Manabe, S., Wetiierland, R. T. Thermal equilibrium of the atmosphere with a given distribution of relative humidity.

[33] Meunier, D., Lambotite, R., Forinto, A., Erdche, K. D., AND Bullmore, E. T. Hierarchical modularity in human brain functional networks. Frontiers in neuroinformatics 3 (2009).

[34] Mitchell, T. M. Machine learning. wcb, 1997.

[35] Nguyen, A., Yossink, J., Clune, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. arXiv preprint arXiv:1412.1897 (2014).

[36] Oja, E. Simplified neuron model as a principal component analyzer. Journal of mathematical biology 15, 3 (1982), 267–273.

[37] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, “Learning a deep convolutional network for light-field image superresolution,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 24–32.

[38] Orponen, P. Computational complexity of neural networks: a survey. Nordic Journal of Computing 1, 1 (1994), 94–110.

[39] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, “Deep learning for content-based image retrieval: A comprehensive study,” in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 157–166.

[40] Person, K. On lines and planes of closest fit to system of points in space. philiosophical magazine, 2, 559-572, 1901.

[41] Rossenblat, F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review 65, 6 (1958), 386.

[42] Smolensky, P. Information processing in dynamical systems: Foundations of harmony theory.

[43] Szegadzy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Andfergus, R. Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199 (2013).

[47] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprintarXiv:1409.1556, 2014.

[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[50] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[51] A. Graves, S. Fernandez, and J. Schmidhuber, “Multi- ´ dimensional recurrent neural networks,” CoRR, vol. abs/0705.2011, 2007. [Online]. Available: http://arxiv.org/ abs/0705.2011.

[52] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. In Parallel Distributed Processing. Vol 1: Foundations. MIT Press, Cambridge, MA, 1986.

[53] D.O. Hebb. The organization of behavior: A neurophychological study. WileyInterscience, New York, 1949.

[54] E. Oja. Simplified neuron model as a principal component analyzer. Journal of mathematical biology, 15(3):267–273, 1982.

[55] Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. In L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, editors, Large-Scale Kernel Machines. MIT Press, 2007.

[56] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? Journal of Machine Learning Research, 11:625–660, 2010.

[57] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504, 2006.

[58] G.E. Hinton, S. Osindero, and Y.W. Teh. A fast learning algorithm for deep belief nets. Neural Computation, 18(7):1527–1554, 2006.

[59] http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/

[60] A. Ess, T. Muller, H. Grabner, and L. J. Van Gool, "Segmentation- ¨ based urban traffic scene understanding." in BMVC, vol. 1, 2009, p. 2.

[61] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 2012, pp. 3354–3361.

[62] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3213–3223.

[63] M. Oberweger, P. Wohlhart, and V. Lepetit, "Hands deep in deep learning for hand pose estimation," arXiv preprint arXiv:1502.06807, 2015.

[64] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, "Learning a deep convolutional network for light-field image superresolution," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 24–32.

[65] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 157–166.

[66] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, "Toward automatic phenotyping of developing embryos from videos," IEEE Transactions on Image Processing, vol. 14, no. 9, pp. 1360–1371, 2005.

[67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classifi- cation with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[68] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[69] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[70] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[71] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision. Springer, 2014, pp. 818–833.

[72] M. D. Zeiler, G. W. Taylor, and R. Fergus, “Adaptive deconvolutional networks for mid and high level feature learning,” in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 2018–2025.

[73] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” arXiv preprint arXiv:1511.00561, 2015.

[74] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” in ACM transactions on graphics (TOG), vol.23, no.3. ACM, 2004, pp. 309–314.

[75] J. Shotton, J. Winn, C. Rother, and A. Criminisi, “Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context,” International Journal of Computer Vision, vol. 81, no. 1, pp. 2–23, 2009.

[76] A. Kendall, V. Badrinarayanan, and R. Cipolla, “Bayesian segnet: Model uncertainty in deep convolutional encoderdecoder architectures for scene understanding,” arXiv preprint arXiv:1511.02680, 2015.

[77] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” preprint arXiv:1412.7062, 2014.

[78] V. Koltun, “Efficient inference in fully connected crfs with gaussian edge potentials,” Adv. Neural Inf. Process. Syst, vol. 2, no. 3, p. 4, 2011.

[79] P. Krahenbuhl and V. Koltun, “Parameter learning and conver- ¨ gent inference for dense random fields.” in ICML (3), 2013, pp. 513–521.

[80] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1717–1724.

[81] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, “Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks,” in European Conference on Computer Vision. Springer, 2008, pp. 69–82.

[82] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in Advances in neural information processing systems, 2014, pp. 3320–3328.

[83] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: when to warp?” CoRR, vol. abs/1609.08764, 2016.

[84] X. Shen, A. Hertzmann, J. Jia, S. Paris, B. Price, E. Shechtman, and I. Sachs, “Automatic portrait segmentation for image stylization,” in Computer Graphics Forum, vol. 35, no. 2. Wiley Online Library, 2016, pp. 93–102.

[85] S. Zhou, J.-N. Wu, Y. Wu, and X. Zhou, “Exploiting local structures with the kronecker layer in convolutional networks,” arXiv preprint arXiv:1512.09194, 2015.

[86] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” arXiv preprint arXiv:1511.07122, 2015.

[87] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “Enet: A deep neural network architecture for real-time semantic segmentation,” arXiv preprint arXiv:1606.02147, 2016.

[88] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2650–2658.

[89] A. Raj, D. Maturana, and S. Scherer, “Multi-scale convolutional architecture for semantic segmentation,” 2015.

[90] A. Roy and S. Todorovic, “A multi-scale cnn for affordance segmentation in rgb images,” in European Conference on Computer Vision. Springer, 2016, pp. 186–201.

[91] X. Bian, S. N. Lim, and N. Zhou, “Multiscale fully convolutional network with application to industrial inspection,” in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016, pp. 1–8.

[92] W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” arXiv preprint arXiv:1506.04579, 2015.

[93] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollar, “Learning ´ to refine object segments,” in European Conference on Computer Vision. Springer, 2016, pp. 75–91.

[94] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. C. Courville, and Y. Bengio, “Renet: A recurrent neural network based alternative to convolutional networks,” CoRR, vol. abs/1505.00393, 2015.

[95] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville, “Reseg: A recurrent neural network-based model for semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2016.

[96] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[97] K. Cho, B. Van Merrienboer, D. Bahdanau, and Y. Bengio, “On ¨ the properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.

[98] Z. Li, Y. Gan, X. Liang, Y. Yu, H. Cheng, and L. Lin, “RGB-D scene labeling with long short-term memorized fusion model,” CoRR, vol. abs/1604.05000, 2016.

[99] G. Li and Y. Yu, “Deep contrast learning for salient object detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 478–487.

[100] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, “Scene labeling with lstm recurrent neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3547–3555.

[101] H. Pinheiro and R. Collobert, “Recurrent convolutional neural networks for scene labeling.” in ICML, 2014, pp. 82–90.

[102] B. Shuai, Z. Zuo, G. Wang, and B. Wang, “Dag-recurrent neural networks for scene labeling,” CoRR, vol. abs/1509.00552, 2015.

[103] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, “Simulta- ´ neous detection and segmentation,” in European Conference on Computer Vision. Springer, 2014, pp. 297–312.

[104] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

POPIS OZNAKA I KRATICA

ANN	artifical neural networks
CNN	convolutinal neural networks
CRF	conditional random fields
DAG	directed acyclic graph
GRU	gated recurrent unit
LSTM	long short-term memory
MLP	multi-layer perception
RNN	recurrent neural network
UCG	undirected cyclic graphs
VGG	visual geometry group