

UMJETNA INTELIGENCIJA I PRIMJENE U DIJAGNOSTICI



Davor Runje, dipl.ing., *Visoko učilište Algebra*

Neuronske mreže spadaju u obitelj algoritama strojnog učenja u kojima autor specificira arhitekturu, način učenja i podatke na kojima algoritam uči, a nakon toga algoritam sâm nalazi najbolji način kako riješiti zadani problem.

Uvod

U rujnu 2012. godine, algoritam *AlexNet* [2] pobijedio je na takmičenju prepoznavanja 1000 različitih vrsta objekata na testnom skupu od 1.2 milijuna fotografija i to s ogromnom razlikom u točnosti u odnosu na konkurentne metode. Osim ogromne razlike u točnosti, *AlexNet* je bio potpuno drukčija vrsta algoritma: *duboka neuronska mreža*. Tipični algoritmi koji su se do tada koristili bili su pažljivo napisani od strane ljudskih programera koji su znali svaki detalj implementacije i u potpunosti razumjeli kako oni rade. S druge strane, autor *AlexNeta*, Alex Krizhevsky, nije znao detalje kako njegov algoritam radi i kako točno razlikuje, na primjer, fotografije pasa od fotografija mačaka. Neuronske mreže spadaju u obitelj algoritama strojnog učenja u kojima autor specificira arhitekturu, način učenja i podatke na kojima algoritam uči,

Sažetak

Tehnika dubokog učenja neuronskih mreža pokrenula je pravu revoluciju u izradi algoritama koji imaju podjednaku ili čak bolju točnost od ljudskih stručnjaka na čitavom nizu kognitivnih audio, vizualnih i tekstualnih zadataka te ovaj nagli napredak nazivamo i novo proljeće umjetne inteligencije. Glavna karakteristika ove tehnike je potpuna nezavisnost od domene u kojoj se primjenjuje, tako da se isti algoritam primjenjuje kod prepoznavanja mačaka na slikama s interneta kao i prepoznavanja patologija na radiološkim snimkama. U ovom preglednom članku ćemo kratko opisati glavne karakteristike i procese dubokog učenja, te dati jedan primjer dijagnostike pneumonije na osnovi radiološke snimke pluća.

a nakon toga algoritam sâm nalazi najbolji način kako riješiti zadani problem. U konkretnom slučaju *AlexNeta*, radilo se o mreži s 650.000 umjetnih neurona. Uspjeh *AlexNeta* pokrenuo je val primjene neuronskih mreža u svim područjima, od računarskih znanosti do lingvistike te na kraju medicine. S vremenom, ovaj val inovacija baziran na dubokim neuronskim mrežama brendirao se pod zvučnim imenom umjetna inteligencija, mada u širem značenju tog pojma spada gotovo sva digitalna tehnologija od samih njenih početaka.

Zanimljivost je da ovaj val inovacija nije pokrenut nekim novim otkrićem. AlexNet se smatra jednim od najtjecajnijim članaka u računarstvu s više od 70.000 citata, ali u sebi zapravo ne sadrži neki novi teoretski rezultat ili uvid. Ovo je jednostavno bio prvi put da su tada poznate metode strojnog učenja upotrijebljene na velikoj neuronskoj mreži, a za to su trebala biti zadovoljena sljedeća dva uvjeta:

1. Velik skup podataka za učenje: u ovom slučaju, skup za učenje bila je baza od 14 milijuna fotografija.
2. Velika procesorska snaga: pokazalo se da standardni procesori koje pokreću naša računala nisu dobar izbor za učenje ovakvih sustava i da je nužno potrebno imati specijalizirane procesore koji su se povijesno koristili za obradu slika (tzv. grafičke kartice, mada se istim imenom nazivaju i procesori koji se ne mogu spojiti na ekrane i koriste se za obradu drugih vrsta signala).

Snažan val digitalizacije, potaknut razvojem interneta i mobilne telefonije, stvorio je ogromne količine podataka na kojima možemo trenirati ovakve modele. Mobilni telefoni opremljeni kamerama visoke razlučivosti i društvene mreže generirali su ogromne količine fotografija s opisima što se na njima nalazi (na primjer, Facebook trenira svoje modele koristeći fotografije s Instagrama koje su označene *hashtagovima* – ključnim riječima prefiksiranim znakom # [4]).

Procesorska snaga se po Mooreovom zakonu [6] udvostručavala svakih 18 mjeseci desetljećima i snaga procesora u današnjim mobilnim telefonima je 100 puta veća od snage superračunala iz 1980.-ih. Međutim, čak niti to nije bilo dovoljno za pokretanje ove revolucije u računarstvu, već su za to zaslužni specijalizirani procesori koji su se do tada pretežito koristili za video igrice te ih zbog toga i danas nazivamo grafičkim karticama (eng. *graphic processing unit*, odnosno *GPU* skraćeno). Neki od tih procesora se uopće ne mogu spojiti na ekran i obrađuju druge vrste podataka, kao što su to zvuk ili tekst, ali ime je i dalje isto. Klasični procesori koje nazivamo centralnim procesorima (eng. *central processing unit*, odnosno CPU skraćeno) sastoje se od dva dijela:

1. kontrolna jedinica koja upravlja time koje instrukcije treba izvršiti (korake algoritma), te

2. aritmetičko-logička jedinica koja izvodi aritmetičke i logičke operacije (na primjer, zbrajanje i množenje brojeva).

Razvojem tehnologije, tranzistori od kojih se grade procesori postajali su sve manji i mogli smo ih sve više staviti na istu površinu silicijskog supstrata. Taj dodatni prostor na supstratu počeo se koristiti na način da se više kopija istog procesora stavljalo na njega i tako su nastali višezegreni procesori koji se sastoje od jednokog broja nezavisnih parova kontrolnih i aritmetičko-logičkih jedinica. Današnji procesori imaju od nekoliko do čak 128 takvih jezgri unutar istog čipa i to im omogućava da izvode stotinu potpuno nezavisnih programa istovremeno. S druge strane, grafički procesori su povijesno bili specijalizirani za izvođenje programa na velikom broju točaka koje čine sliku (piksela). Iako se takvi programi mogu razbiti na potprograme koji se izvode istovremeno na nekom broju susjednih piksela, ti programi nisu nezavisni, već imaju potpuno iste korake koji se izvršavaju na različitim točkama. Da bi se optimizirao prostor na siliciju, takvi procesori imaju značajno manji broj kontrolnih jedinica u odnosu na aritmetičko-logičke jedinice i u stanju su izvršiti isti program na stotinama točaka istovremeno. Tako moderni grafički procesori danas imaju 10.000 aritmetičko-logičkih jedinica na 100 kontrolnih jedinica, za razliku od centralnih procesora koji imaju po 100 jedinica aritmetičko-logičkih i kontrolnih jedinica. Pokazalo se da ovakva arhitektura procesora daje dva reda veličine veću procesorsku snagu (100 puta veću) za obradu slika, ali i za treniranje dubokih mreža kod kojih je najzahtjevniji zadatak množenje velikih matrica realnih brojeva (točnije rečeno aproksimacija realnih brojeva zbog vrste digitalnog kodiranja koja koristi vrlo limitirani broj binarnih znamenaka, najčešće 32 ili 16 jedinica ili nula). Bez ovog ubrzanja, proces učenja i eksperimentiranja s velikim neuronskim mrežama bio bi jednostavno predug za praktičan rad.

Spektakularni rezultat na zadatku strojnog prepoznavanja objekata na fotografijama postignut AlexNetom je tako rezultat akumuliranog razvoja tehnologije i digitalizacije velikih količina podataka, a ne novim teoretskim rezultatom. Inspirirani tim uspjehom, istraživači su ubrzano krenuli upotrebljavati duboke neuronske mreže na teškim zadacima i postigli čitav niz poboljšanja u odnosu na postojeće *state-of-the-art* rezultate. Tako se danas problem percepcije slike i zvuka načelno smatra

riješenim (prepoznavanje objekata i općenito percepcija prostora na slikama, prepoznavanje govora), a postignut je i veliki napredak u obradi prirodnog jezika, mada se i dalje ne nazire njegovo rješenje [3].

Kako rade neuronske mreže

Osnovni element neuronskih mreža je umjetni neuron. Umjetni neuron prima na ulazu jednu ili više ulaznih numeričkih vrijednosti koje transformira u numerički izlaz. Ova transformacija sastoji se od linearne transformacije (svaka ulazna vrijednost pomnoži se s parametrom te se svi međurezultati zbroje), te od nelinearne funkcije koja definira izlaznu vrijednost. Na primjer, za ulazne vrijednosti 1 i 2 u neuron s parametrima 0,1, 0,2 i 0,3 i nelinearnom funkcijom x^2 , izlaz je 0,64:

$$(1 \times 0,1 + 2 \times 0,2 + 0,3)^2 = (0,8)^2 = 0,64$$

Jedan ovakav neuron je naizgled jednostavan, ali njegova prava snaga dolazi do izražaja kad se veliki broj ovakvih neurona organizira u mrežu. Svaki ulaz u mrežu je ulaz u svaki od n neurona iz prvog sloja koji generiraju n izlaza koristeći svoje parametre i nelinearnu funkciju na način koji je opisan gore. Ti izlazi su ulazi u sljedeći sloj koji generira nove izlaze i taj postupak se ponavlja dok se ne dođe do zadnjeg sloja čiji izlazi su izlazi mreže. Na taj način mreža transformira n ulaznih u m izlaznih vrijednosti, koristeći linearne transformacije određene parametrima mreže i fiksnim nelinearnim funkcijama neurona. Pokazalo se da ovakve mreže mogu aproksimirati bilo kakvu funkciju, pri čemu točnost te aproksimacije ovisi o broju neurona. U praksi se također pokazalo da duboke mreže (mreže s puno slojeva neurona) imaju bolju sposobnost učenja od plitkih, širokih mreža (minimalno dva sloja s velikim brojem neurona u njima), mada obje varijante mogu proizvoljno točno aproksimirati bilo kakvu funkciju.

Najveći praktični izazov je proces učenja ovakvih mreža. Cilj ovog procesa je nalaženje optimalnih parametara za koje mreža daje očekivani rezultat. Na primjer, u slučaju AlexNeta, neuronska mreža sastojala se od 650.000 umjetnih neurona karakteriziranih s 60 milijuna parametara, a funkcija koju je ta mreža aproksimirala je funkcija s 196.608 ulaznih varijabli (256x256 točaka za koju imamo tri vrijednosti intenziteta plave, zelene i crvene boje), a čiji izlaz je 1000 vrijednosti koje označavaju vjerojatnost da se na njima nalazi jedan od 1000 kategorija objekata

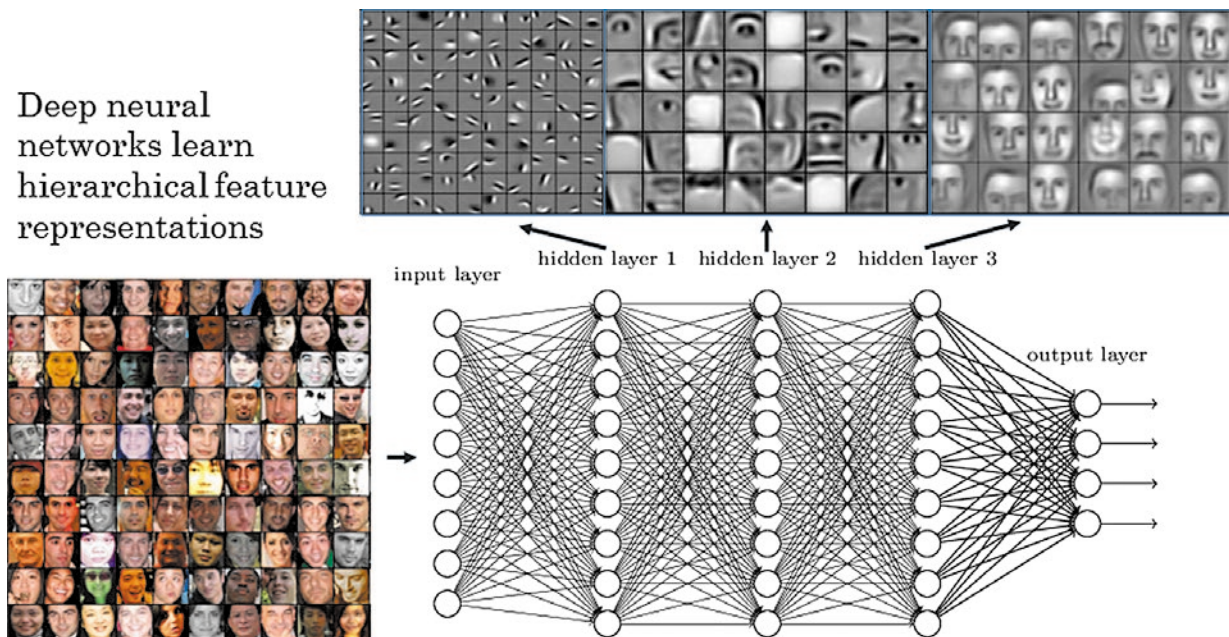
koje raspoznaje. Nasumično pogađanje 60 milijuna varijabli i testiranje koliko dobro sustav radi nije opcija; vjerojatnost da slučajno pogodite dobre parametre je jednostavno premala. Srećom, postoji bolja tehnika koja sukcesivno evaluira izlazne funkcije mreže na primjeru iz skupa podataka za treniranje te korigira postojeće parametre mreže da daju bolji rezultat na tom konkretnom primjeru. Na primjer, AlexNet je inicijaliziran slučajnim parametrima mreže tako da vjerojatnost svake od 1000 izlaznih klasa bude otprilike 0,1%. Recimo da se mreža krene trenirati slikom na kojoj se nalazi pas. Parametri mreže bit će promijenjeni tako da izlazna vjerojatnost mreže za psa sad bude 0,1001%, a za sve ostale klase 0,9999999%. Ovaj postupak se ponavlja za svaku od 15 milijuna slika dok izlazne vjerojatnosti klasa mreže ne budu vrlo blizu 100% za točnu klasu i 0% za netočnu. Naizgled jednostavan, ovaj postupak je izuzetno efikasan uz pretpostavku da imate veliki broj slika za treniranje i veliku procesorsku snagu koja je potrebna za veliki broj malih promjena parametra.

Što uče neuronske mreže

Iako je gore opisani postupak matematički potpun, on ipak ne objašnjava kako točno funkcioniraju neuronske mreže i na osnovu čega postižu toliko visoku točnost.

Na slici 1 prikazana je neuronska mreža čiji je zadatak prepoznati jednu od četiri emocije na osnovu fotografije lica. Mreža se sastoji od četiri sloja neurona, pri čemu zadnji sloj ima četiri neurona čiji izlazi predstavljaju vjerojatnost jedne od četiri emocija. Mreža je naučena na skupu fotografija za koje je ručno označena jedna emocija. Na slici 1 nalaze se primjeri takvih ulaznih fotografija na lijevoj strani, a na gornjem dijelu se nalaze fiktivne ulazne slike za koji svaki od neurona unutar mreže ima maksimalnu vrijednost. Fotografija za koju neuron ima maksimalnu vrijednost zapravo nam govori za što točno je taj neuron specijaliziran. Tako vidimo da neuroni prvog sloja pozitivno reagiraju na razne točkice i linije pod određenim kutovima, da neuroni drugog sloja koriste informacije iz prvog sloja da bi reagirale na određene oblike očiju, nosa i usta, te da bi na osnovu tih informacija neuroni trećeg sloja reagirali na određene izraze lica. Završni sloj nam na kraju govori koji svi izrazi lica idu uz koju emociju. Ovo je vrlo razuman način da se ovaj kompleksni problem faktorizira u jednostavnije potprobleme, ali ono što je specifično za strojno učenje je da

Deep neural networks learn hierarchical feature representations



Slika 1. Što uče neuronske mreže

programer koji je napravio ovu mrežu nije dao nikakve instrukcije računalu kako da napravi ovu faktorizaciju. Kod klasičnog programiranja, programer daje vrlo precizne instrukcije računalu kako se algoritam izvršava i svaki detalj mora biti u potpunosti specificiran. Kod strojnog učenja neuronskih mreža, programer određuje arhitekturu mreže (koliko mreža ima slojeva, koliko neurona u pojedinom sloju, koju vrstu nelinearne funkcije na izlazu neurona, itd.) i na koji način će mreža učiti (skup podataka za treniranje, broj i veličina koraka), dok način na koji će problem biti riješen nije specificiran i algoritam učenja ima potpunu slobodu da particionira problem na potprobleme i riješi ih. Neki ovu metodu nazivaju i softverom koji sam piše softver. Posljedice ovakvog pristupa programiranja su dalekosežne i vjerojatno ćemo ih biti u potpunosti svjesni tek za dvadesetak godina.

Transfer znanja

U praksi, rijetko kada imamo na raspolaganju toliko veliki skup podataka kao što je to ImageNet (15 milijuna slika), a koji je potreban da bi se trenirale velike neuronske mreže. U takvim slučajevima, gotovo uvijek se koristi tehnika zvana transfer znanja, a koja se koristi na sljedeći način:

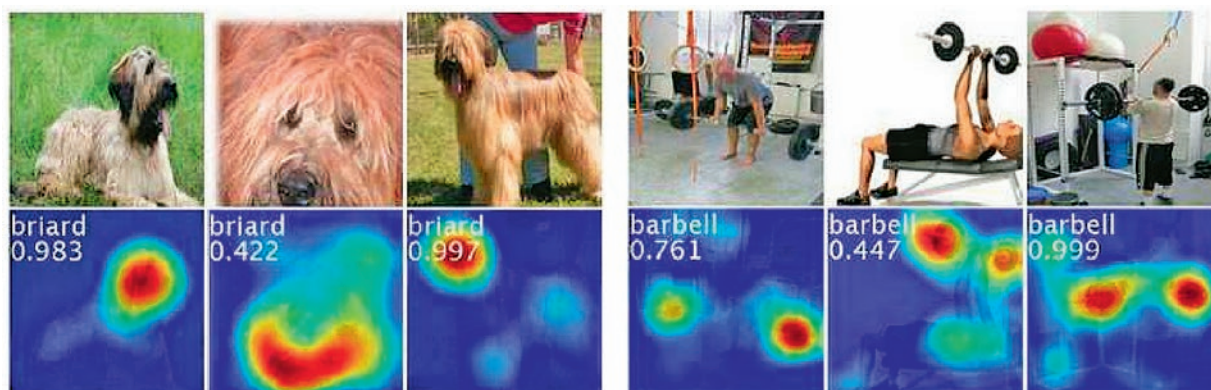
1. Prvotno se trenira mreža na nekom sličnom zadatku za koji postoji veliki skup podataka, na primjer, najčešće ImageNet za mreže koje obrađuju slike.

2. Nakon toga, istrenirana mreža prekonfigurira se za novi zadatak na način da se zadrže svi slojevi osim jednog ili nekoliko zadnjih koji se odbacuju. Tada se treniranje nastavlja na novom zadatku s manjim brojem podataka.

Uspješnost ove tehnike možemo intuitivno objasniti činjenicom da su početni slojevi mreža invarijantni na sâm zadatak. U gore navedenom primjeru prepoznavanja emocija iz fotografije lica, prvi sloj prepoznaje uzorke linija i krugova. Mreža koja bi, na primjer, prepoznavala vrstu voća u posudi, mogla bi koristiti isti sloj, a tek u daljnjim slojevima izgraditi potrebnu aparaturu za prepoznavanje distinktivnih karakteristika jabuka, krušaka i šljiva.

Interpretabilnost

Neuronske mreže često se nazivaju crnim kutijama (eng. *black boxes*) zbog njihove interne kompleksnosti i nemogućnosti jednostavne interpretacije dobivenih rezultata. Moderne mreže danas imaju više od 100 slojeva i milijune parametara i nije lako shvatiti način na koji postižu svoje rezultate. U brojnim primjenama ovo nije bitno, ali kako se ovakvi algoritmi sve češće primjenjuju u kritičnim sustavima, interpretabilnost rezultata počinje biti barem poželjno, a vrlo često i obavezno. Samovozeći automobili, nadgledanje i upravljanje nuklearnim elektranama i medicinska dijagnostika samo su



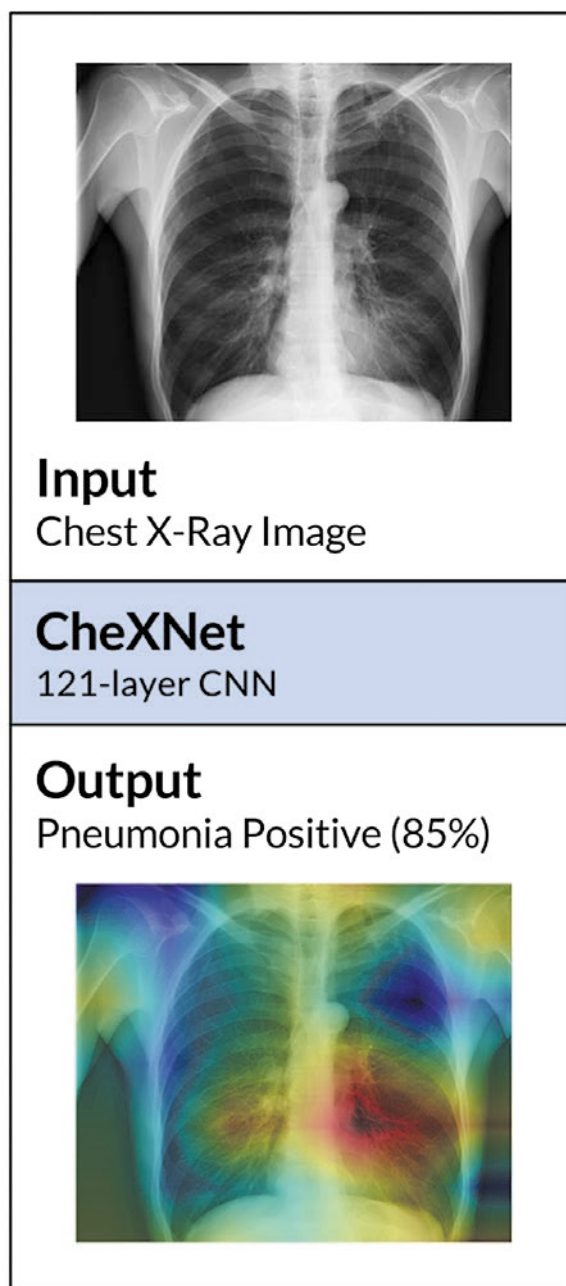
Slika 2. Aktivacijska mapa lokalizira prediktivno relevantni dio slike

neka od područja u kojima je apsolutno nužno razumjeti i interpretirati rezultate dobivene neuronskim mrežama. Zadnjih godina se područje interpretabilnosti sve više razvija, a za neke primjene imamo i vrlo zadovoljavajuće tehnike. Za primjer ćemo navesti tehniku mapiranja aktivacijskih klasa (eng. *Class Activation Mapping*, skraćeno CAM) koja računa intenzitet kojim svaki piksel slike utječe na izlazni rezultat [7].

Duboko učenje u dijagnostici

U medicini postoje veliki skupovi digitaliziranih podataka i duboko učenje ima sve veći broj primjena. Osim samih patologija koje prepoznaju na osnovu slike, velika primjena neuronskih mreža je medicinska segmentacija slike. Kao primjer, detaljnije ćemo prikazati mrežu CheXnet za detekciju pneumonije iz radiološke slike [5]. Ulaz u mrežu je crno-bijela slika rezolucije 224x224, a izlaz je vjerojatnost pneumonije od 0 do 100%. Arhitektura CheXneta ima čak 121 sloj i pripada obitelji mreža zvanih *DenseNet*, a trenirana je tehnikom prijenosa učenja s početne mreže trenirane na ImageNetu na 112.120 frontalnih radioloških slika pluća od 30.805 jedinstvenih pacijenata koji imaju jednu od 14 različitih patologija. To znači da smo prvo uzeli 15 milijuna fotografija pasa, mačaka, automobila i slično da bismo istrenirali mrežu koja ih prepoznaje, nakon toga odbacili zadnji sloj mreže i zamijenili ga novim, te istrenirali mrežu do kraja na stotinjak tisuća radioloških slika. Da je netko prije deset godina napisao ovakvu rečenicu, mislili bismo da se šali, ali tu smo gdje jesmo i milijuni slika mačaka s interneta pomažu nam da bolje prepoznamo patologije na radiološkim slikama pluća.

Najzahtjevniji dio treniranja ovakve mreže je prikupljanje i anotacija podataka. Za testiranje točnosti algoritma i



Slika 3. Radiološka slika i aktivacijska mapa

	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

Slika 4. Usporedba točnosti radiologa i algoritma

usporedbu s ljudskom točnošću, slučajnim odabirom odabrano je 420 radioloških slika koje su nezavisno pregledali i dijagnosticirali četiri radiologa i algoritam. Važno je napomenuti da dotičnih 420 radioloških slika nije bilo korišteno za treniranje mreže te da su prvi puta bile upotrijebljene na testu. Od četiri radiologa, samo jedan je imao neznatno bolju F1 mjeru od algoritma (harmonijska sredina između preciznosti i odziva), dok su trojica bila značajnije lošija.

Za kraj, da bismo pomogli radiolozima da potvrde danu dijagnozu, koristi se CAM tehnika za lokalizaciju dijela slike zbog koje je algoritam prepoznao patologiju. Na primjeru danom na Slici 3, vjerojatnost pneumonije iz radiološke slike je 85%, a žarište je označeno crvenom bojom zbog lakše potvrde dijagnoze od strane liječnika.

Zaključak

Tehnika dubokog učenja omogućila nam je da izradimo algoritme koji imaju podjednako dobru ili bolju točnost od ljudskih stručnjaka u prepoznavanju patologija na dijagnostičkim slikama. Ovakvi modeli danas još uvijek ne uzimaju u obzir punu kliničku sliku, ali razvojem modela za procesiranje prirodnog jezika i strukturiranih podataka samo je pitanje dana kad će biti u stanju sagledati potpunu sliku. Primjena ove tehnologija automatizirat će veliki dio rutinskog posla koji danas opterećuje ljudske stručnjake, a njihovo vrijeme će moći biti bolje iskorišteno u potrazi za liječenjem pacijenata i za traženje novih oblika patologija kako bismo napravili još bolje algoritme. ■

Reference

- Deng, Jia, et al. „Imagenet: A large-scale hierarchical image database.“ *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. „Imagenet classification with deep convolutional neural networks.“ *Advances in neural information processing systems*. 2012.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. „Deep learning.“ *Nature* 521.7553 (2015): 436-444.
- Mahajan, Dhruv, et al. „Exploring the limits of weakly supervised pretraining.“ *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
- Rajpurkar, Pranav, et al. „Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning.“ *arXiv preprint arXiv:1711.05225* (2017).
- Schaller, Robert R. „Moore’s law: past, present and future.“ *IEEE spectrum* 34.6 (1997): 52-59.
- Zhou, Bolei, et al. „Learning deep features for discriminative localization.“ *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Abstract

The deep learning technique of neural networks has started a real revolution in the development of algorithms that have equal or even better accuracy than human experts on a whole range of cognitive audio, visual and textual tasks and this rapid progress is called the new spring of artificial intelligence. The main feature of this technique is complete independence from the domain in which it is applied, so that the same algorithm is applied when recognizing cats in images from the Internet as well as recognizing pathologies on radiological images. In this review article, we will briefly describe the main characteristics and processes of deep learning, and give an example of the diagnosis of pneumonia based on a radiological image of the lungs.