# STRUKTURA JEDNOSTAVNE KONVOLUCIJSKE NEURONSKE MREŽE

Konvolucijska neuronska mreža je specijalizirana vrsta neuronskih mreža i dizajnirana je za rad s 2D slikovnim podacima, iako ih se može koristiti i kod 1D ili 3D podataka. Središte konvolucijskih neuronskih mreža čine konvolucijski slojevi.[2]

## ULAZNI SLOJ

Općenito, ulazni sloj prima neovisne varijable ulaznih primjera. Ulazi su podaci koji se prosljeđuju neuronskoj mreži na temelju kojih ona donosi određenu odluku ili predviđanje temeljeno na podacima. Svaka ulazna vrijednost se prosljeđuje jednom neuronu u neuronskoj mreži. Za razliku od ostalih slojeva u bilo kojoj neuronskoj mreži, on ne izvodi nikakve operacije nad ulaznim primjerima već on služi samo zato da primi ulazne podatke takvima kakvi jesu i da ih proslijedi mreži. [13]

Broj neurona u ulaznom sloju određen je samom strukturom podataka koji se prosljeđuju neuronskoj mreži. On odgovara broju dimenzija, ili broju jedinstvenih značajki, ulaznih podataka.

Jedna od glavnih razlika između tradicionalne neuronske mreže i konvolucijske neuronske mreže je ta da su neuroni u drugoj mreži organizirani u tri dimenzije (visina, širina i dubina). U konvolucijskom neuronskim mrežama, ulazni sloj sadrži vrijednosti piksela ulazne slike.[4]

Ulazni sloj konvolucijske neuronske mreže najčešće prima slikovne podatke. Slikovni podaci su, kao što je maloprije spomenuto, predstavljeni korištenjem 3D matrice. Tu matricu je, prije nego se proslijedi mreži, potrebno transformirati u vektor, odnosno u samo jedan stupac. Neka ulazna slika ima dimenzije 28x28=784. Tu sliku je potrebno transformirati u vektor 784x1. To znači da će ulazni sloj u konvolucijsku neuronsku mrežu imati onoliko piksela koliko ih ima ulazna slika preko sve svoje tri dimenzije.[14][15]

## KONVOLUCIJSKI SLOJ

Računalo čita slike kao matrice s vrijednostima piksela koje imaju dimenzije NxNx3 (visina, širina, dubina). Slike se najčešće sastoje od 3 kanala (RGB) i zato imaju 'dubinu'. U kontekstu konvolucijskih neuronskih mreža, konvolucija je linearna operacija koja uključuje množenje skupa težina (filtera) sa skupom odgovarajućih ulaza (matrica s vrijednostima piksela.[5][10]

Ulaz u konvolucijski sloj se naziva još i ulaznim kanalom (eng. *input channel*). Izlaz iz konvolucijskog sloja se naziva izlaznim kanalom (eng. *output channel*), matricom značajki (eng. *feature map*) ili matricom aktivacija (eng. *activation map*).[1][2]

### FILTERI

Svaki konvolucijski sloj u konvolucijskoj neuronskoj mreži posjeduje definirani broj filtera. Ti filteri se sustavno primjenjuju na ulazni kanal i tako sloj kreira matricu značajki.[1][2][5]

Primjena više filtera na ulazni kanal rezultira jednim izlaznim kanalom (matricom značajki). Ulazni kanal je obično 3D slika (broj redaka, stupaca i kanala). Tako i filteri moraju imati 3 dimenzije s jednakim brojem kanala, ali manjim brojem stupaca i redaka u odnosu na ulaznu sliku. Filter sadrži težine (u tradicionalnim neuronskim mrežama bi to bile uobičajene težine veza) čije optimalne vrijednosti mreža treba naučiti tijekom procesa učenja. Težine filtera definiraju strukturu ili obrazac (eng. *feature*) koji će filter detektirati na ulaznom kanalu.[5]

Slika se matematički može izraziti kao matrica sljedećih dimenzija:

Gdje je:

: visina slike

: širina slike

: broj kanala

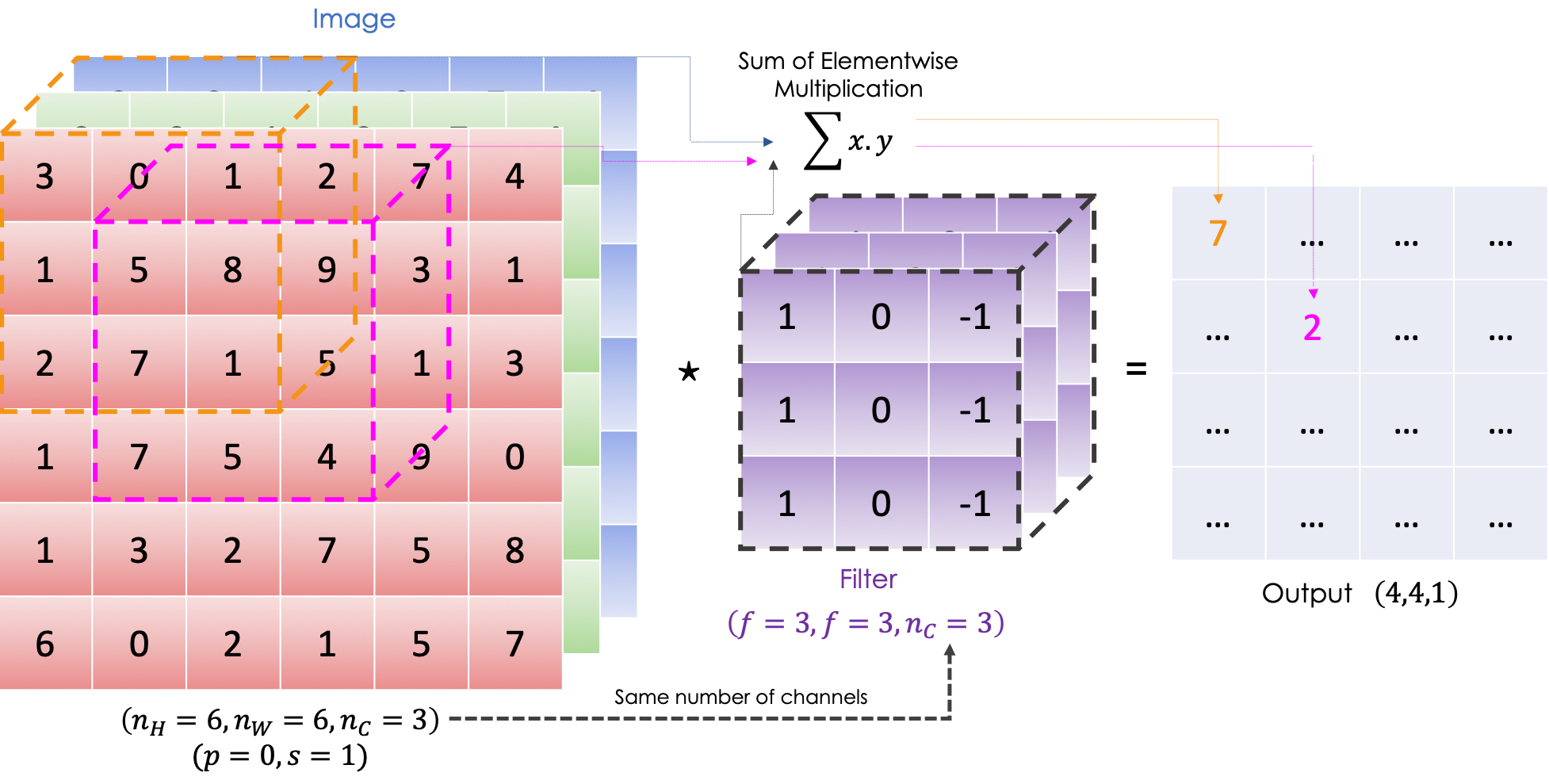
Za RGB sliku, broj kanal je jednak 3 s obzirom da RGB slika posjeduje crveni, zeleni i plavi kanal.[7]

Filter, koji će se označavati s *K*, najčešće ima jednake dimenzije visine i širine (3x3, 5x5 ili 7x7) i koje se označavaju s *f*. Takav filter, s neparnim brojem za dimenzije, omogućava da se svaki piksel slike nalazi u sredini filtera. Filter mora imati jednak broj kanala kao i slika.[7]

Dimenzije filtera se tako označavaju na sljedeći način:

Konvolucijski produkt slike i filtera je 2D matrica značajki. Toj matrici je svaki element rezultat sume množenja po odgovarajućim elementima (eng. *elementwise multiplication*). Filteri sadrže težine koje su inicijalizirane nasumičnim brojčanim vrijednostima. Filter u jednom trenutku računa točkasti produkt (eng. *dot product*) na kocki (jer posjeduje visinu, širinu i dubinu) ulazne matrice koja ima jednake dimenzije kao i filter. Ta 'kocka' koju filter promatra u danom trenutku se naziva *patch* ili 'receptivno polje' (eng. *receptive field*).[7][8]

Primjer jednog konvolucijskog produkta prikaza je na slici:



Konvolucijska operacija se matematički može opisati kao:

Gdje je:

* *I*: ulazna slika
* *K*: filter
* *x*: broj retka u matrici značajki
* *y*: broj stupca u matrici značajki
* *i*: redni broj retka ulazne matrice
* *j*: redni broj stupca ulazne matrice
* *k*: redni broj kanala ulazne matrice

Dimenzije matrice značajki se dobiju na sljedeći način:

Gdje je pod (eng. *floor*) funkcija od *x*.

Postoje neke specijalne vrste konvolucije:

1. *Valid* konvolucija:
2. *Same* konvolucija: veličina izlaza je jednaka veličini ulaza:
3. 1x1 konvolucija: ; ovo je korisno u nekim slučajevima kako bi se smanjio broj kanala bez da se smanje dimenzije ().

Konvolucijski sloj izvodi konvolucijske operacije na svojim ulazima tako što istovremeno primjenjuje više filtera. Izlazi iz konvolucijskog sloja se prosljeđuju aktivacijskoj funkciji koja se u ovom primjeru označava s .

Ovdje je potrebno napomenuti da literatura nije usuglašena pripadaju li aktivacijske funkcije konvolucijskom sloju ili čine zasebni, aktivacijski sloj. U ovom poglavlju će se opisati na koji se način aktivacijske funkcije primjenjuju na rezultate konvolucije dok će aktivacijske funkcije biti obrađene u zasebnom poglavlju.

Na *l*-tom sloju vrijede sljedeće anotacije:

* Ulaz: dimenzija gdje je slika koja ulazi u neuronsku mrežu.
* padding: i pomak:
* broj filtera: (broj kanala u tom sloju) gdje svaki (oznaka za filter) ima dimenzije
* prag n-te konvolucije:
* aktivacijska funkcija:
* izlaz: s dimenzijama

Sada vrijedi:

:

I nadalje:

S tim da:

Parametri u l-tom sloju koji se mogu naučiti su:

* filteri s parametara
* prag s parametara

### ZERO PADDING

Svaki konvolucijski sloj posjeduje definirani broj filtera određenih dimenzija. Sustavnom primjenom filtera, dimenzije matrice značajki su reducirane u odnosu na dimenzije ulaznog kanala.[2]

Kao što je matematički opisano u prethodnom poglavlju, filter se primjenjuje na ulazni kanala tako što se gornji lijevi kut filtera preklapa s gornjim lijevim kutom ulaznog kanala. Nakon što izračuna točkasti produkt, filter se pomiče za jedno mjesto (ili u slučaju slike, piksel) u desno te ponovno računa točkasti produkt. Cijeli proces se ponavlja sve dok se gornji desni kut filtera ne poklopi s gornji desnim kutom slike. Znači, filter se primjenjuje samo unutar slike, ne izvan nje i zbog toga dolazi do redukcije dimenzija.

Redukcija dimenzija matrice značajki u odnosu na ulazni kanal se naziva 'efekt ruba' (eng. *border effect*). Taj efekt ne predstavlja veliki problem kod velikih slika i malih filtera, iako se gube potencijalno važne informacije koje se nalaze na rubovima slike. Problem je puno veći kada se filter primjenjuje na slike malih dimenzija, kada je više konvolucijskih slojeva povezano u seriji ili kod dubokih mreža s desecima ili stotinama skrivenih slojeva. U jednom trenutku mreža može ostati bez podataka u matrici značajki nad kojom bi mogla izvoditi operacije ili predviđanja.[2][7][11]

Pikseli, koji se nalaze na rubovima ulaznog kanala (ili slike), bit će izloženi samo rubovima filtera i rjeđe se koriste od piksela koji se nalaze po rubovima slike. Kada bi se filter počeo primjenjivati tako da se njegov rub nalazi izvan okvira slike, pikseli, koji se nalaze na rubovima ulaznog kanala, imaju veću priliku utjecati na rezultat konvolucije. Drugim riječima, dobiva se više prilika kako bi se detektirale neke strukture. Još jedna od posljedica je da matrica značajki ima iste dimenzije kao i ulazni kanal.[2][7]

Kako bi se riješio problem 'odbacivanja' piksela na rubovima ulaznog kanala, oko rubova ulaznog kanala se dodaje neka vrsta popune (eng. *padding*), dodaju se dodatni pikseli oko cijelog okvira ulaznog kanala. Prema konvenciji se taj *padding* sastoji od nula. Zato se naziva *zero padding*. *Padding* se označava s *p* i označava broj elemenata koji su dodani na sve četiri strane ulaznog kanala.[7][11]

S obzirom na to koristi li se *padding* ili ne, postoje dvije vrste konvolucije: *same* konvolucija i *valid* konvolucija. *Valid* konvolucija znači da se konvoluira originalni ulazni kanal brez dodanog obruba. *Same* konvolucija znači da se oko ruba slike stavlja dodatni obrub s nulama tako da matrica značajki ima iste dimenzije kao i ulazni kanal. U drugom slučaju, 'debljina' obruba se računa na sljedeći način:

Gdje su:

* p: 'debljina' obruba
* f: dimenzija visine ili širine filtera.[1][11]

### POMAK

Kao što je već opisano, filter se po ulaznom kanalu pomiče s lijeva na desno, od gore prema dolje po jedan piksel horizontalno i vertikalno. Broj piksela koliko se filter pomiče može varirati, ali je uvijek simetričan u horizontalnom i vertikalnom smjeru. Taj broj se naziva 'pomak' (eng. *stride*).[2][8]

Osnovni pomak je (1,1) za pomicanje filtera po visini i širini ulaznog kanala i ima zadovoljavajuće performanse u većini slučajeva. Ako se ne koristi pomak (1,1), onda se najčešće koristi pomak (2,2). Na taj način se filter pomiče dva stupca u desno i dva reda dolje. Posljedica većeg pomaka je da će matrica značajki imati puno manje dimenzije. Na primjer, ako se 3x3 filter primjenjuje na 6x6 sliku, matrica značajki će imati dimenzije 3x3.[2]

## SLOJ SAŽIMANJA

Sloj sažimanja (eng. *pooling layer*) slijedi nakon jednog ili konvolucijskih slojeva. Taj sloj, kao što mu i samo ime govori, reducira dimenzije izlaza iz konvolucijskog sloja. Drugim riječima, uzorkuje prostornu dimenzionalnost (visinu i širinu, ali ne i broj kanala) te tako smanjuje broj parametara i kontrolira prenaučenost mreže. Postoji nekoliko razloga zašto konvolucijska neuronska mreža sadrži takav skriveni sloj.[4][5][6][8]

Konvolucijski slojevi sustavno primjenjuju određeni filter (ili više njih) na ulazne kanale kako bi kreirali matrice značajki koje ukazuju na prisutnost odgovarajućih struktura na ulaznom kanalu. Ti slojevi su se pokazali jako efikasnima i slaganje konvolucijskih slojeva u seriju omogućuje da se detektiraju sofisticirane strukture kao što su konkretni oblici i specifični objekti.[1]

Međutim, matrice značajki pohranjuju precizne lokacije detektiranih struktura. To znači da bi mali pomak u lokaciji rezultirao potpuno različitom matricom značajki. Takav mali pomak u lokaciji se događa kod manjih modifikacija na, na primjer, ulaznim slikama.[3]

Taj problem se rješava uzorkovanjem, odnosno sažimanjem. Izlaz iz sloja sažimanja je matrica značajki manjih dimenzija koja i dalje sadrži važne informacije, ali ne sadrži informacije koje vrlo vjerojatno nisu potrebne. Takve matrice značajki su korisne jer će male promjene u lokaciji rezultirati jako sličnom, ako ne i istom, matricom značajki čime je postignuta generalizacija pretpostavljenih struktura. Ova sposobnost sažimanja se zove 'neosjetljivost mreže na lokalnu translaciju' (eng. *net's invariance to local translation*).[3][6][8][9]

Slojevi sažimanja također sadrže filtere, odnosno receptivno polje. Njihova dimenzija je skoro uvijek 2x2. Ti filteri su često puno manji od filtera konvolucijskog sloja i nikada ne sadrže vlastite brojčane vrijednosti. Moglo bi ih se nazivati okvirima. Također, pomak filtera u sloju sažimanja je postavljen na brojčanu vrijednost koja izbjegava preklapanje, a to je najčešće 2.[3][8]

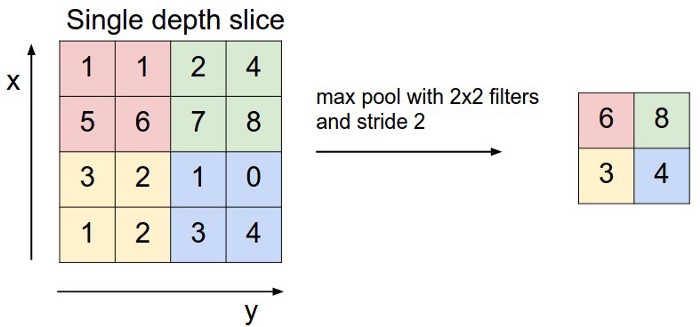
To znači da će sloj sažimanja skoro uvijek reducirati matricu značajki za 2, odnosno svaka će se dimenzija prepoloviti (visina i širina). Broj elemenata ulazne matrice značajki će se smanjiti na četvrtinu.[3]

Postoje dvije moguće operacije u sloju sažimanja: sažimanje maksimalnom vrijednošću (eng. *max pooling*) i sažimanje prosječnom vrijednošću (eng. *average pooling*).[3][5]

Sažimanje maksimalnom vrijednošću uzima maksimalnu vrijednost iz okvira. To se čini tako što filter klizi po matrici značajki i, sa svakim pomakom, uzima najveću vrijednost iz okvira.[5]

Sažimanje prosječnom vrijednošću uzima prosječnu vrijednost iz okvira i pohranjuje ju u matricu značajki reduciranih dimenzija.[3]

Na dolje prikazanoj slici je opisano sažimanje maksimalnom vrijednošću:



Na slici je prikazana matrica značajki dimenzija 4x4. Filter sažimanja ima dimenzije 2x2 i pomak je postavljen na vrijednost 2.

Prva regija matrice značajki je obojena u crveno. Maksimalna vrijednost tog područja je 6 i ona se sprema u prvo polje sažete matrice značajki.

Filter se pomiče za dva mjesta u desno. Maksimalna vrijednost tog područja je 8 o ona se sprema u drugo polje.

Kako je filter došao do ruba matrice, vraća se na početak i spušta se za dva polja dolje. Cijeli proces se ponavlja i određene vrijednosti se pohranjuju na odgovarajuća mjesta u matrici.

Isti ovaj proces se odvija i kod sažimanja prosječnom vrijednošću. Jedina razlika je što se umjesto najveće vrijednosti uzima prosječna vrijednost područja koja se potom sprema na odgovarajuće mjesto u matrici.

Matematički se cijeli proces može opisati na sljedeći način:

Neka vrijedi sljedeća anotacija:

* Ulaz: dimenzija gdje je slika koja ulazi u neuronsku mrežu.
* popuna: (rijetko se koristi) i pomak:
* veličina filtera za sažimanje:
* funkcija sažimanja:
* izlaz: s dimenzijama

Vrijedi:

S tim da:

Sloj za sažimanje ne posjeduje parametre koje konvolucijska neuronska mreža može naučiti.[7]

## POTPUNO POVEZANI SLOJEVI I IZLAZNI SLOJ

Potpuno povezani slojevi u konvolucijskoj neuronskoj mreži su standardni skriveni slojevi tradicionalne neuronske mreže. Ovi slojevi također sadrže neke od ne-linearnih aktivacijskih funkcija. Ako je taj sloj izlazni sloj, često će kao aktivacijsku funkciju imati softmax ili sigmoidnu aktivacijsku funkciju.[8]

Potpuno povezani slojevi se koriste pri kraju neuronske mreže. Oni koriste rezultate konvolucijskog sloja ili sloja sažimanja te ih koristi kako bi generirao određene odluke mreže.[8][9]

Postoji nekoliko vrsta potpuno povezanih slojeva u konvolucijskog neuronskoj mreži. To su:

1. Sloj za izravnavanje (eng. *flatten layer*): uzima izlaz iz prethodnih slojeva (koji su u obliku matrice) te ih 'izravnava', odnosno transformira u vektor vrijednosti.
2. Potpuno povezani sloj: na svoje ulaze prima vrijednosti prethodnog sloja za izravnavanje pomnožene s težinama veza i na njima primjenjuje aktivacijsku funkciju. Ovaj sloj se može smatrati standardnim skrivenim slojem u tradicionalnim umjetnim neuronskim mrežama.
3. Potpuno povezani izlazni sloj: na svojim izlazima daje konačne odluke neuronske mreže za određene ulazne primjere.[9]

Još jedan način na koji se može gledati potpuno povezane slojeve je taj da potpuno povezani slojevi čine tradicionalnu neuronsku mrežu koja je integrirana s konvolucijskom neuronskom mrežom.

[1] [www.deeplizard.com](http://www.deeplizard.com)

[2] <https://machinelearningmastery.com/padding-and-stride-for-convolutional-neural-networks/>

[3] <https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/>

[4] <https://arxiv.org/pdf/1511.08458.pdf>

[5] <https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17>

[6] <https://missinglink.ai/guides/convolutional-neural-networks/convolutional-neural-network-tutorial-basic-advanced/>

[7] <https://towardsdatascience.com/convolutional-neural-networks-mathematics-1beb3e6447c0> --> tu su formule

[8] <https://machinelearningmastery.com/crash-course-convolutional-neural-networks/>

[9] <https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/>

[10] <https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/>

[11] <https://towardsdatascience.com/gentle-dive-into-math-behind-convolutional-neural-networks-79a07dd44cf9>

[12] <https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529>

[13] <https://missinglink.ai/guides/neural-network-concepts/complete-guide-artificial-neural-networks/>

[14] <https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529>

[15] <https://stackoverflow.com/questions/52272592/how-many-neurons-does-the-cnn-input-layer-have/52273707>