

Ansiktsigenkänning med hjälp av faltningsnät

Tim Dahlström

xx.xx.2023

Referat

{WIP}

Sökord: maskininlärning, faltningsnät, CNN, ansiktsigenkänning.

Innehåll

1. INTRODUKTION	1
2 BAKGRUND.....	2
2.1 Maskininlärning	2
2.2 Artificiella neuroner.....	3
2.3 Djupinlärning	4
2.4 Ansiktsigenkänning.....	6
2.5 Användningsområde.....	7
3 STRUKTUREN AV CNN	8
3.1 Grundstruktur.....	8
3.2 Indatalager.....	10
3.3 Faltninglager (Convolutional layer).....	11
3.4 Poolinglager (Pooling layer).....	14
3.5 Fullt kopplade lager	15
4 CNN INOM ANSIKTSIGENKÄNNING.....	16
4.1 Fördelar/nackdelar med CNN.....	16
4.2 AlexNet	16
4.3 ResNet50	16
SAMMANFATTNING/DISKUSSION	17
KÄLLOR.....	18

1. Introduktion

Den vanligaste identifieringsmetoden har länge varit lösenord, men under det senaste årtiondet har det blivit allt vanligare att använda sig av biometri för identifiering av personer. Biometri innebär tekniker för att mäta fysiologiska samt beteendemässiga egenskaper elektroniskt av biologiska data. Till de vanligaste identifieringsmetoderna inom biometri hör iris-, röst-, fingeravtrycks- samt ansiktigenkänning. Dessa identifieringsmetoder medför många fördelar över den mera traditionella metoden att använda lösenord, till exempel så räcker det med att personen som ska identifieras är närvarande för identifikationsprocessen [1].

Ansiktigenkänning är en identifieringsmetod som vuxit i popularitet under det senaste årtiondet. Detta är på grund av de stora framstegen som skett inom maskininlärningsområdet. I dagens läge har ansiktigenkänning ett brett användningsområde som omfattar allt från människa—datorinteraktion till identitetsidentifiering vid åtkomstkontroll och gränskontroll [2].

Denna avhandling kommer att behandla faltningsnätverk eller faltningsnät (eng: Convolutional Neural Network, CNN) och dess implementation samt roll inom ansiktigenkänning. I kapitel 2 av avhandlingen kommer både definitionen och utvecklingen av maskininläring, djupinläring samt ansiktigenkänning att presenteras. Sedan i kapitel 3 presenteras den grundläggande strukturen för faltningsnätverk. Inom kapitel 4 i avhandlingen kommer jag att presentera två välkända implementationer av faltningsnätverk i ansiktigenkänning, AlexNet och ResNet50. Varav den första populariserade användningen av faltningsnätverk inom datorseendet och den andra introducerade en mera komplex och djup arkitektur för att uppnå bättre resultat. Jag kommer att diskuteras även om de fördelar som faltningsnätverk har över sina företrädare när det handlar om ansiktigenkänning och bildklassificering.

2 Bakgrund

2.1 Maskininlärning

Maskininlärning (engelska: Machine Learning) är ett delområde av artificiell intelligens (AI). Maskininlärning fick sin start år 1943 då forskarna Warren McCulloch och Walter Pitts försökte ta reda på om det vore möjligt för två datorer att kommunicera med varandra utan medverkan av människor. Deras modell som baserade sig på elektriska kretsar bevisade att det verkligen var möjligt för datorer att kommunicera med varandra [3].

Begreppet maskininlärning myntades först år 1959, då forskaren Arthur L. Samuels publicerade sin forskning om ett datorprogram som själv via inlärning kunde förbättra sin spelskicklighet i spelet dam. Sedan dess har tekniken inom lagring och processorkraften utvecklats i mångfald, vilket har lett till att maskininlärningen har frambringat allt nyare och innovativa idéer som till exempel självkörande bilar [4].

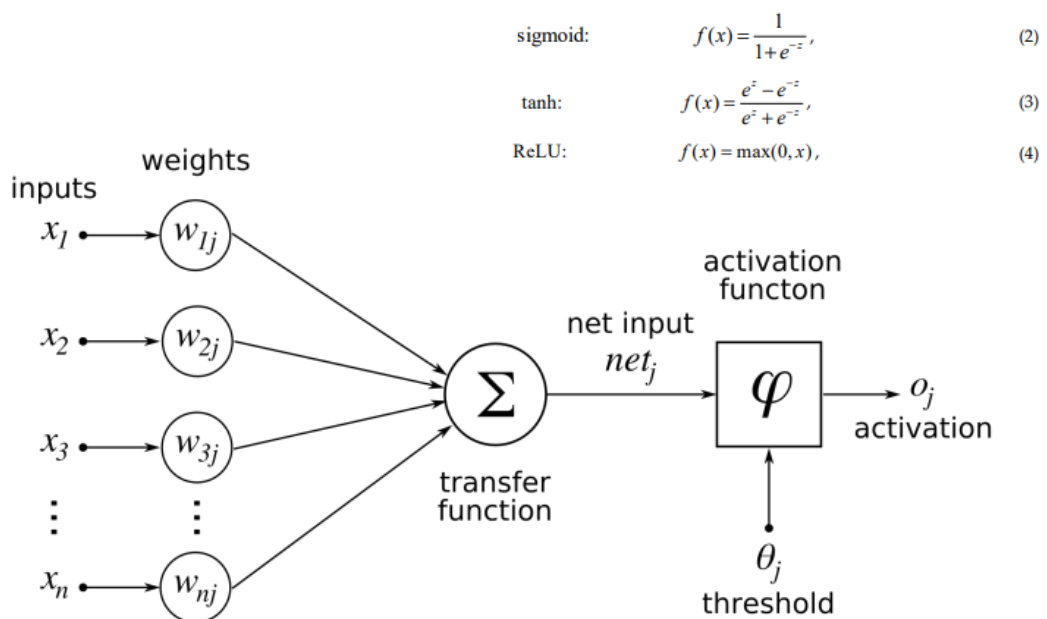
Inom maskininlärning finns tre olika inlärningsformer: övervakad inlärning (eng. supervised learning), oövervakad inlärning (eng. unsupervised learning) och förbättrande inlärning (eng. reinforcement learning). Övervakad inlärning innebär att algoritmen syftar till att analysera data enligt kända egenskaper, så att algoritmer lär sig direkt från indata (eng. input data) samt utdata (eng. output data). Däremot handlar oövervakad inlärning om att algoritmen bör finna okända mönster och egenskaper i data, utan att algoritmen får ta del av någon explicit återkoppling. Förbättrande inlärning är en typ av inlärning där algoritmen tar del av en serie olika belöningar eller bestraffningar, och därefter lär sig vad som är lönsamt eller icke lönsamt [5, p. 694-695].

{Skriv om Mitchells definition av inlärning - WIP}

2.2 Artificiella neuroner

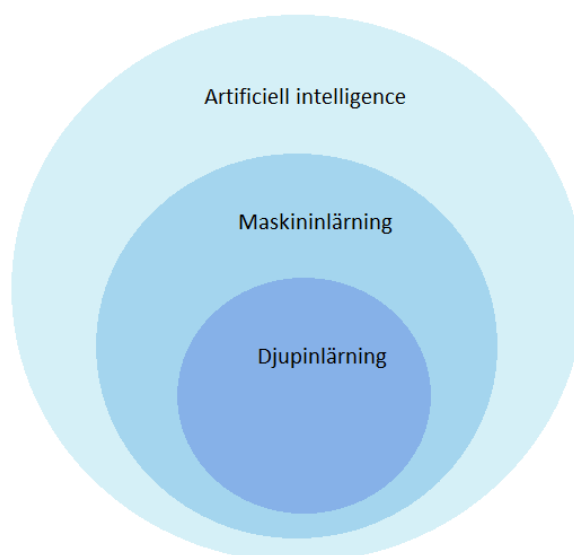
Från ett informationsbehandlingssynsätt, kan en biologisk neuron definieras som en grundläggande enhet som kan bearbeta och överföra information genom elektriska eller kemiska signaler. Den biologiska neuronerna som är byggstenen i det nervsystem som finns i människans kropp och andra djur har stora liknelser med artificiella neuroner som används inom artificiella neuronnätverk (eng. Artificial Neural Network, ANN). Den första liknelsen är att på samma sätt som den biologiska neuronerna blir stimulerad av indata så tar den artificiella neuronerna också emot indata och har förmågan att föra data vidare. Den andra liknelsen mellan de två typerna av neuroner är att båda kan tränas över vilken typ av indata de tar emot och sedan för vidare. På så vis kan man säga att det är möjligt att träna neuroner på att filtrera ut signaler som kan uppskattas som onödiga [6]. Artificiella neuronnätverk tillämpas i dagens läge främst inom områden som mönsterigenkänning, taligenkänning och signalbehandling [7].

{Skriv om neuroner (vikter, bias, aktiveringsfunktion)}



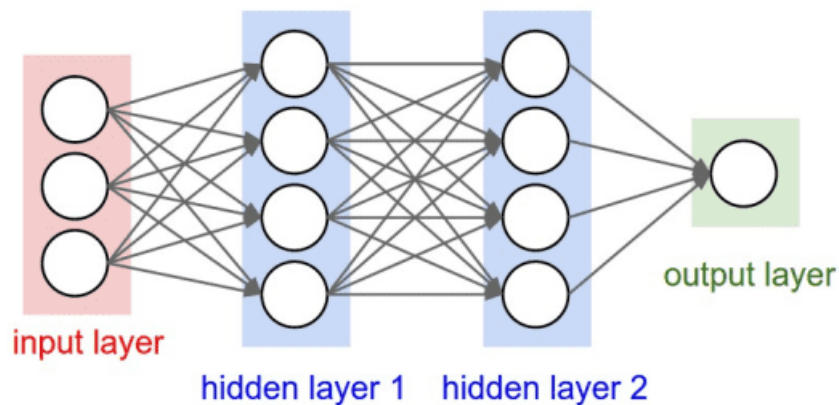
2.3 Djupinlärning

Djupinlärning (eng. Deep learning) är ett delområde av maskininlärning som är ytterligare ett delområde av artificiell intelligens (figur 1.). Djupinlärning grundar sig på neuronnät som använder sig av mer än ett gömt lager, d.v.s. fler än ett lager mellan indatalagret (eng. Input layer) och utdatalagret (eng. Output layer) [6][9]. Fördelen med att använda sig av ett neuronnät bestående av flera lager som behandlar icke-linjär information är att det möjliggör effektiv egenskapsinlärning samt mönsterigenkänning hos nätverket [8]. Traditionella maskininlärningsmetoder kräver oftast någon form av förbehandling av indata. Förbehandlingen kan till exempel innebära att man kategoriserar eller minskar dimensionaliteten på indata. Däremot brukar metoder som tillämpar sig av djupinlärning inte kräva någon form av förbehandlingen av indata. Vilket i sin tur innebär att att problem som involverar ostrukturerad data, till exempel bilder och ljud, passar bra för metoder som använder sig av djupinlärning.



Figur 1: Relationen mellan djupinlärning, maskininlärning och artificiell intelligens.

Termen djupinlärning togs i bruk först år 2006 då forskaren Geoffrey Hinton publicerade sin forskning om Deep Belief Networks (DBN), en typ av feedforward nätverk var data rör sig endast åt ett håll inom nätverket. Ett exempel på hur en fundamental struktur av ett feedforward nätverk kan se ut presenteras i figur 2.



Figur 2: Ett feedforward nätverk med två dolda lager. [11]

Hintons forskning introducerade en ny implementation av oövervakad inlärning som tillåter skapandet av neuronnät bestående av flera lager. Var de olika lagren består av egenskapsdetektorer som har funktionen att identifiera egenskaper hos icke annoterade data, d.v.s. data som inte har blivit klassificerad. Dessa egenskapsdetektorer ökar i komplexitet gradvis för varje lager inom neuronnätet. Genom förträning av neuronnätet gör man det möjligt att behålla rimliga värden hos vikterna på neuronerna inom modellen. Efter att neuronnätet har tränats klart kan man sedan tillämpa bakåtspridning (eng. backpropagation) för att finslipa parametrarna av modellen. Bakåtspridning innebär att man använder avvikelser mellan utdata och det önskade resultatet för att justera på inställningarna inuti nätverket. Avvikelsena skjuts bakåt i nätverk och ändrar på nätverkets ursprungliga inställningar ända tills nätverkets resultat är nöjaktiga [10].

2.4 Ansiktsigenkänning

Ansiktsigenkänning har revolutionerat vårt sätt att identifiera människor och dess egenskaper genom användningen av avancerade algoritmer och maskininlärnings metoder. Tekniken fungerar väl på grund av att människans ansikte har en unik struktur och unika egenskaper som varierar från person till person, vilket i sin tur gör det lättare att avskilja och identifiera människor. Dock finns det en hel del utmaningar som ansiktsigenkänningsystem bör lösa. Till dessa utmaningar hör bland annat belysning, udda ansiktsuttryck, synligheten av ansiktet och upplösningen av bilden. För att ett system som använder sig av ansiktsigenkänning skall kunna prestera så bra som möjligt bör bilderna vara av hög kvalitet, d.v.s. ansiktet bör vara synligt och inte dolt av något objekt som till exempel solglasögon. Om ansiktsigenkänningsystemet använder sig av maskininläring bör man ta hänsyn till vilken typ av data man använder för att träna modellen. Till exempel om träningsdata inte är tillräckligt mångsidigt kan modellens prestation lida. Ifall modellen tränas endast över data av personer av samma kön eller etnicitet kommer modellen prestera sämre då den identifierar en person av annat kön eller annan etnicitet [11].

De tre grund funktionerna av ett funktionerande ansiktsigenkänningsystem kan beskrivas som; (1) identifiering av ansikte i en bild, (2) extraheringen av egenskaper hos ett ansikte, (3) jämförandet av ansiktets egenskaper med andra ansikten i en databas [1].

{utveckla, face detection, feature extraction, face recognition}

{beskriv de tre olika tillvägagångsätt för ansiktsigenkänning:}

{Local approach}

{Holistic approach}

{Hybrid approach}

2.5 Användningsområde

{Access control}

{Border control}

{Marketing}

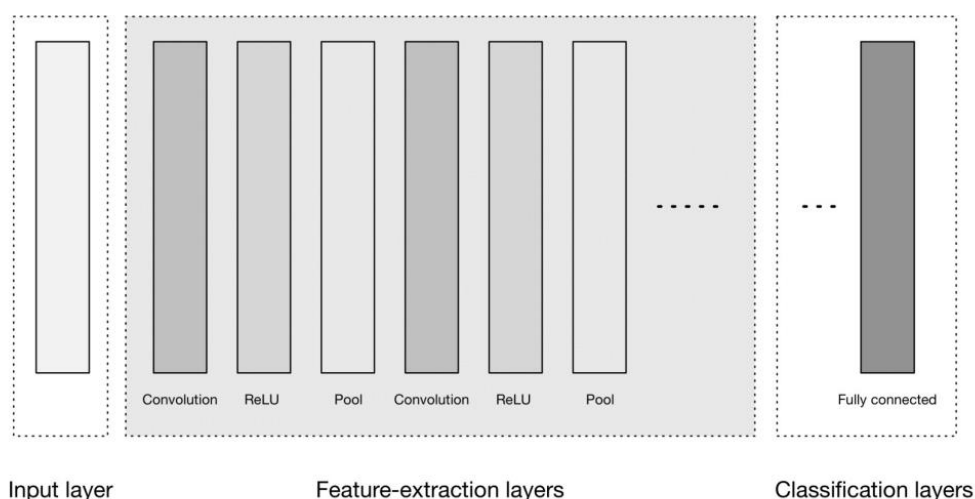
{National security}

3 Strukturen av CNN

Faltningsnät är en klass av artificiella neurala nätverk som tillämpas främst över visuell data så som bilder och videoklipp men kan även tillämpas över ljud och signal data. Huvudtanken bakom faltningsnät är att de har förmågan att lära sig och identifiera egenskaper inom data genom faltning (eng. convolution). I och med detta fungerar faltningsnät bra för att identifiera ansikten, individer, trafikskyltar och andra objekt utav rådata. Tekniken används även inom språkteknologi för tolkning av tecken inom texter, till exempel för tydandet av handstil. I dagens läge är faltningsnät ett av de mest forskade områden inom datorseendet, i första hand på grund av att faltningsnät är utmärkta för att hitta egenskaper i rådata oberoende av egenskapens position och rotation. [6]

3.1 Grundstruktur

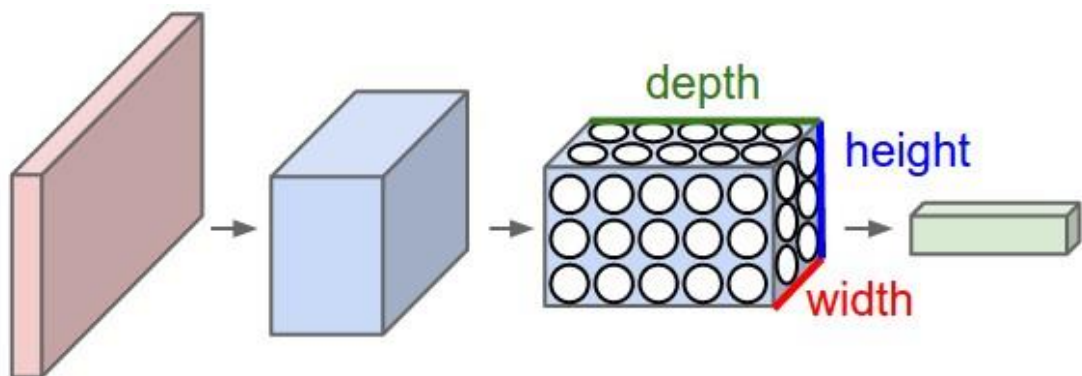
Strukturen av ett faltningsnät kan variera betydligt beroende på ändamålet av implementationen. Dock så följer varje faltningsnät en grundstruktur som presenteras i **figur 3**.



Figur 3: Strukturen av ett faltningsnätverk på högnivå. [6]

Grundstrukturen av faltningsnät kan delas upp i tre grupper: (1) indatalager, (2) lager för egenskapsextrahering och (3) klassificeringslager.

I indatalagret tas en eller flera bilder in som rådata för att vidare bearbetas i modellen. Lagren för egenskapsextrahering är en grupp av två olika unika lager för faltningnät, faltningslager och poolinglager. Dessa lager ansvarar för att hitta och identifiera mera avancerade egenskaper hos den data som lagren tar in. Slutligen består faltningsnät av ett klassificeringslager var vi använder oss av ett eller flera lager av fullt kopplade neuroner. Syftet med klassificeringslagret är att man skall koppla ihop all de egenskaper som modellen har hittat in till en endimensionell vektor för att därmed göra en klassificering [6]. Till exempel om man vill ta reda på om ett djur på en bild föreställer en hund eller en katt så kommer den slutliga klassificering ske inom klassificeringslagret. Där modellen tar i beaktan alla de egenskaper den identifierat i en bild och sedan anger ett definierat klassvärde.



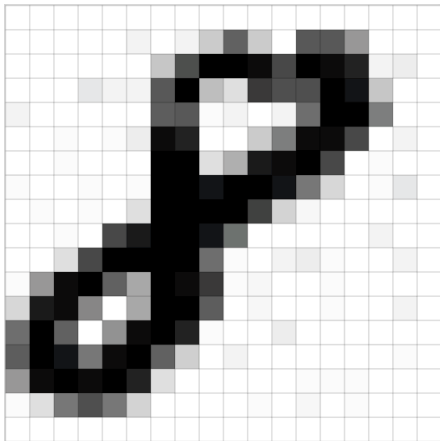
Figur 4: CNN modell vars neuroner är ordnade i ett tredimensionellt fält (höjd, bredd och djup). Det röda lagret består av indata bilden i modellen, vars höjd och djup representerar antalet pixlar i bilden och djupet representerar de tre färgkanalerna röd, grön och blå. [12]

I traditionella artificiella neuronnät beskriver man oftast modellens djup enligt hela nätverkets struktur, d.v.s. hur många gömda lager nätverket använder. Däremot i ett faltningsnät definierar djupet den dimensionalitet som data har i modellen. Man kan säga att neuronerna i ett faltningsnät kan ordnas i antingen en tvådimensionell eller en tredimensionell struktur beroende på om nätverket använder indata i gråskala eller RGB-färgrymden. I **Figur 4** ser vi en visualisering av ett faltningsnät där indatalagret består av en tredimensionell struktur. Höjden och bredden i faltningsnätet avspeglar den samma strukturen som indatat har, d.v.s. pixelhöjden och pixelbredden. Ytterligare så avspeglar djupet i indatalagret de tre färgkanalerna röd, grön och blå som representerar bilder med färg. Varje lager i modellen tar och omvandlar den tredimensionella indata till utdata bestående av en tredimensionell volym av neuronaktiveringar [12].

Om faltningsnätet använder sig av mini-batch träning kan man beskriva nätverket att ha ytterligare en till dimension. Mini-batch inträning innebär att man tränar en modell med en större sats av data på en och samma gång. Detta innebär att modellen kommer att behöva en till dimension för att representera indexet för den data som behandlas. Ifall modellen behandlar bilder med färg inom RGB-färgrymden så kommer dimensionalitet av data som indatalagret och egenskapsextraheringslagren utmata att vara fyrdimensionell, (höjd x bredd x djup x index). På samma sätt kommer också utdata från klassificeringslagret vara tvådimensionellt $[b \times N]$, var b representerar antalet bilder som använts vid träningen av modellen och N representerar antalet olika klasser vi vill använda vid klassificering [6].

3.2 Indatalager

Indatalagret (eng. Input layer) är det första lagret inom ett faltningsnät. Lagret antar den indata som modellen kommer att bearbeta. Inom bildbehandling representerar bilder av tensorer bestående av pixelvärden.



```

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 12 0 11 39 137 37 0 152 147 84 0 0 0
0 0 1 0 0 0 0 41 160 250 255 235 162 255 238 206 11 13 0
0 0 0 16 9 9 150 251 45 21 184 159 154 255 233 40 0 0
10 0 0 0 0 0 145 146 3 10 0 11 124 253 255 107 0 0
0 0 3 0 4 15 236 216 0 0 38 109 247 240 169 0 11 0
1 0 2 0 0 0 253 253 23 62 224 241 255 164 0 5 0 0
6 0 0 4 0 3 252 250 228 255 255 234 112 28 0 2 17 0
0 2 1 4 0 21 255 253 251 255 172 31 8 0 1 0 0 0
0 0 4 0 163 225 251 255 229 120 0 0 0 0 0 11 0 0
0 0 21 162 255 255 254 255 126 6 0 10 14 6 0 0 9 0
3 79 242 255 141 66 255 245 189 7 8 0 0 5 0 0 0 0
26 221 237 98 0 67 251 255 144 0 8 0 0 7 0 0 11 0
125 255 141 0 87 244 255 208 3 0 0 13 0 1 0 1 0 0
145 248 228 116 235 255 141 34 0 11 0 1 0 0 0 1 3 0
85 237 253 246 255 210 21 1 0 1 0 0 6 2 4 0 0 0
6 23 112 157 114 32 0 0 0 0 2 0 8 0 7 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```

Figur 5: Exempel på hur bilder representeras som data.

I **figur 5** ser vi hur en svartvit bild med upplösning 18x18 kan representeras som en tensor bestående av pixelvärden. Bilder inom RGB-färgrymden representeras på samma sätt förutom att de utgör i stället en tredimensionell tensor som representerar färgkanalerna röd, grön och blå. Var varje pixelvärde i tensoren beskriver en färg respektive dess färgkanal.

3.3 Faltningsslager (Convolutional layer)

Det lager som kallas för faltningsslaget är avgörande för att identifiera egenskaper hos data i ett faltningsslagret och betraktas som själva hörnstenen i nätverket. Lagret får sitt namn från den matematiska operationen faltning som sker i lagret.

Faltning innebär att man bildar en summafunktion av två andra funktioner och brukar vanligtvis betecknas med en asterisk: $s(t) = (x * w)(t)$, och definieras som

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a)$$

Var funktionen x består endast av diskreta värden snarare än kontinuerliga värden. I ett faltningsslagret representerar funktionen x indata, funktionen w kärna eller filter (eng. kernel eller filter) och $s(t)$ funktionskarta (eng. feature map eller activation

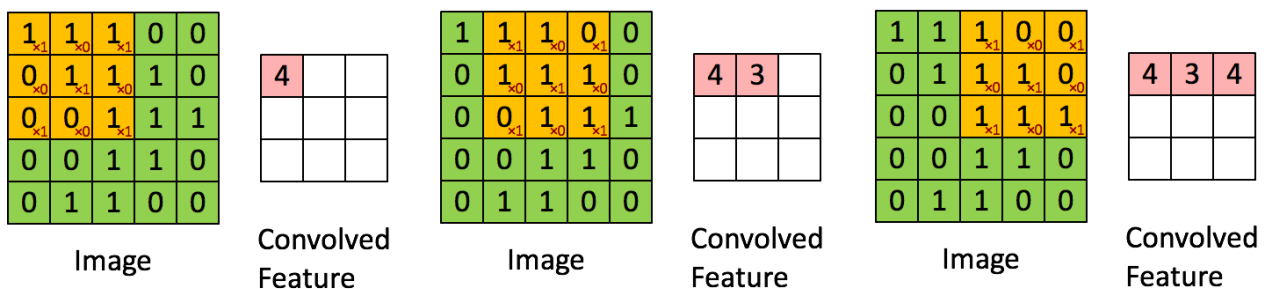
map). Eftersom alla dessa komponenter består av flerdimensionella vektorer, beskriver man dem som tensorer inom maskininlärning [14].

Faltningsslaget har förmågan att hitta egenskaper i en bild genom att skjuta kärnor, som även kallas för filter, över indata med en specificerad steglängd (eng. stride). Skalärprodukten mellan kärnans reella tal och indata beräknas vilket bildar en egenskap. Dessa egenskaper sparas och skapar en funktionskarta. Kärnan i ett faltningsslag har en mindre bredd och höjd än indata i faltningsslaget och konfigureras enligt faltningsslagets parametrar [6].

1	0	1
0	1	0
1	0	1

Figur 6: En kärna (filter) med dimensionen $3 \times 3 \times 1$. [13]

I **figur 6** ser vi en kärna med dimensionen $3 \times 3 \times 1$. Den här kärnan faltas sedan över en tensor av dimensionen $5 \times 5 \times 1$, där faltningen sker med steglängden 1 i **figur 7**. Vi ser hur kärnan förflyttas en pixel i taget över bredden och höjden av indata och sedan beräknas skalärprodukten av faltningen för varje steg. Vilket i sin tur skapar en funktionskarta över egenskaper. Ett faltningsslag kan anta indata både som rådata till exempel en bild eller som funktionskartor från tidigare faltningsslag.

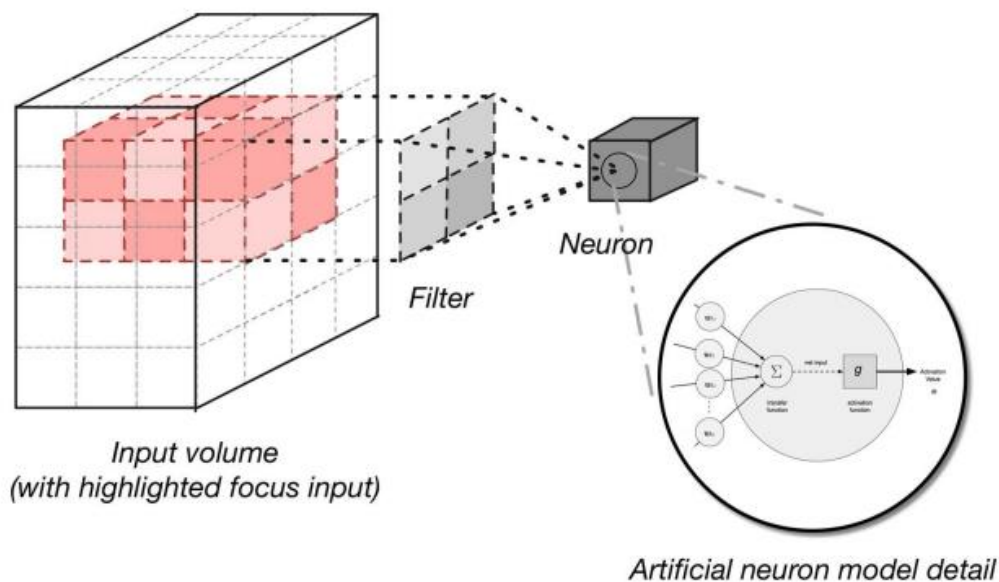


Figur 7: Faltningen av indata och en kärna som sedan bildar en funktionskarta. [13]

Inom maskininlärning beskrivs det att en neuron aktiveras då när data passerar genom den. På samma sätt inom ett faltningsslag beskriver man att en kärna

aktiveras då den släpper igenom data. Genom träning lär sig faltningsnätverket använda olika typer av kärnor som aktiveras då de hittar specifika egenskaper i indata. I faltningsnätverk som tillämpar bildklassificering använder främst aktiveringsfunktionen likriktad linjär enhet (eng. rectified linear unit, ReLU) vilken definieras som $f(x) = \text{argmax}(0, x)$. Detta medför att modellen antar 0 för alla värden av x då $x < 0$ och x för alla värden $x > 0$.

ReLU aktiveringsfunktionen används inom faltningsnät eftersom funktionen ändrar endast på pixelvärdet av indata utan att ändra på dimensionaliteten av utdata. Genom att inte använda negativa värden undviker vi risken för att en icke relevant neuron skall negativt påverka utdata av nätverket. Fördelen också med att använda sig av ReLU inom faltningsnätverk är att aktiveringsfunktionen påverkar nätverkets beräkningskostnad linjärt. Som en följd av detta bidrar användningen av ReLU till att förhindra den exponentiella tillväxten av beräkningar som krävs för att driva ett faltningsnätverk. I **figur 8** ser vi aktiveringen av en neuron i faltningslagret över en tredimensionell volym av indata.



Figur 8: Genereringen av en aktiveringsvolym inom ett faltningslager. [6]

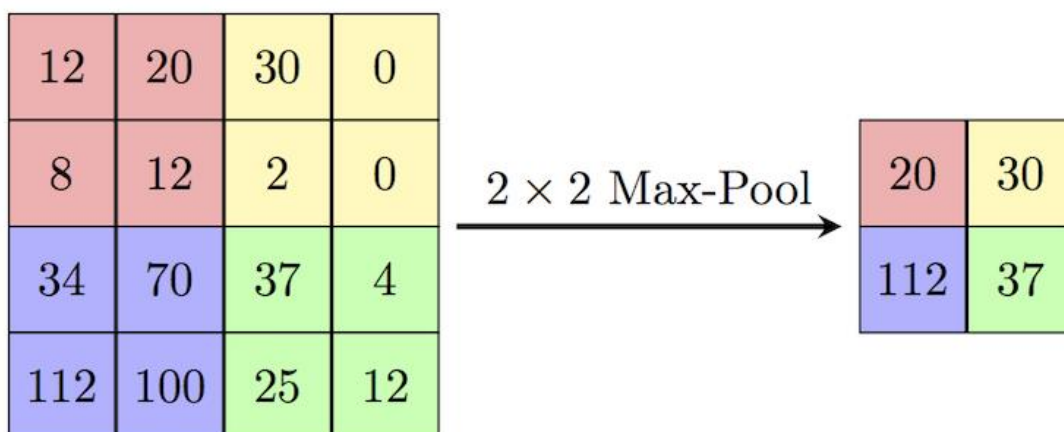
3.4 Poolinglager (Pooling layer)

Poolinglager (eng. pooling layer) används mellan faltningslager för att minska på dimensionaliteten i höjd och bredd rymden. På grund av att bilder består av en stor mängd data finns det en risk för att nätverket undergår överträning (eng. overfitting). Vilket innebär att nätverket fungerar bra över dess inträningsdata men inte över den data som man testar nätverket med. Man kan indirekt säga att nätverket har memorerat träningsdata men misslyckas i att bearbeta ny data som nätverket inte tidigare har sett. Syftet med poolinglager är att motverka detta problem [6].

Poolinglager använder sig av $\max()$ funktionen för att minska på dimensionaliteten av indata. Inom faltningsnät beskrivs denna term med namnet *max pooling*. Max pooling går ut på att man speglar ett filter över indata och sedan behåller endast maxvärdet av det speglade området. Likt kärnor i faltningslagret flyttar man filtret över bredden och höjden av indata med hjälp av en specificerad steglängd [6][12].

Den vanligaste dimensionaliteten på filter som används för max pooling är 2×2 med steglängden 2, på så vis minskar man indatavolymen med 75% [12]. I **figur 9** utförs max pooling med hjälp av ett 2×2 filter med steglängden 2.

Dimensionaliteten av indata minskar därmed i figur 9 från 4×4 till 2×2 .



Figur 9: Max pooling med ett 2×2 filter. [15]

3.5 Fullt kopplade lager

Problemet med att klassificera utdata från ett faltningsnät är att faltningslagren representerar utdata i en abstrakt form. Vi ser i **figur 8** att neuronerna tittar endast på en del av indata, detta innebär att neuronerna är endast kopplade till de närliggande neuronerna som har iakttagit delvis samma indata. För att göra en korrekt klassificering bör nätverket kunna sammankoppla alla högnivåegenskaper och lära sig olika icke-linjära kombinationer av dessa egenskaper. Ett lätt sätt att undvika detta problem är att lägga till ett utdatalager var alla neuroner är fullt sammankopplade (eng. fully connected layer) [12]. Ett fullt kopplat lager består av en given mängd neuroner där alla neuronerna i lagret är kopplade till varandra. Vilket låter nätverket ta i beaktan alla identifierade egenskaper och skapa en sannolikhetskala eller poängsättning för klassificeringen [6].

{Utveckla: affine function, non-linear function}

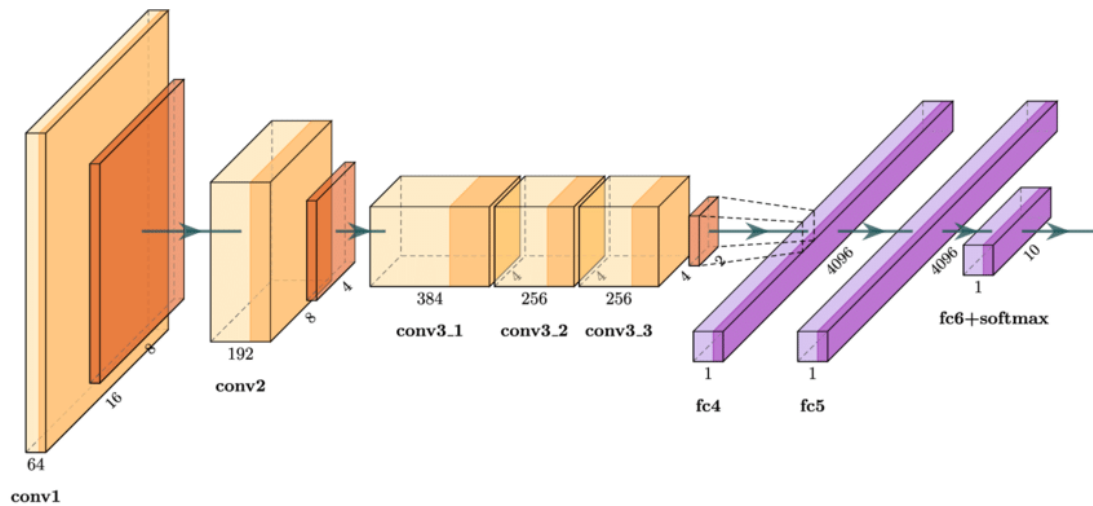
4 CNN inom ansiktsigenkänning

4.1 Fördelar/nackdelar med CNN

{WIP}

4.2 AlexNet

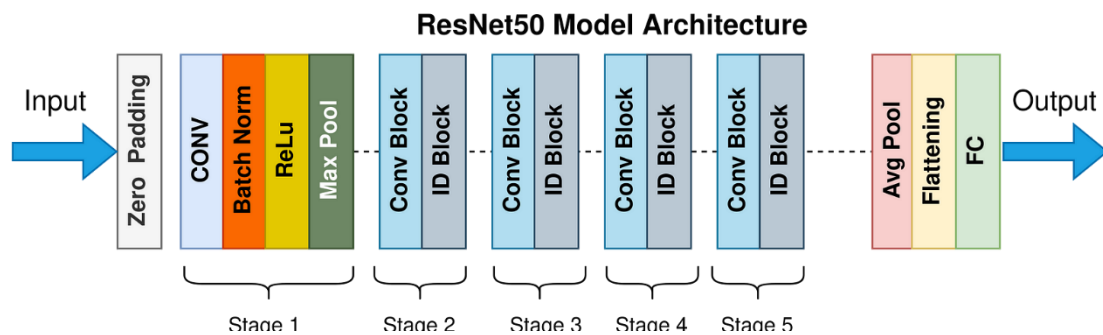
{WIP}



Figur 10: placeholder

4.3 ResNet50

{WIP}



Figur 11: placeholder

Sammanfattning/diskussion

{WIP}

Källor

- [1] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face Recognition Systems: A Survey," *Sensors*, vol. 20, no. 2, Art. no. 2, Jan. 2020, Available:
- [2] C. Zhao, X. Li, and Y. Cang, "Bisecting k-means clustering based face recognition using block-based bag of words model," *Optik - International Journal for Light and Electron Optics*, vol. 126, no. 19, pp. 1761–1766, Oct. 2015, doi: [10.1016/j.ijleo.2015.04.068](https://doi.org/10.1016/j.ijleo.2015.04.068).
- [3] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, doi: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259).
- [4] IBM, "What is Machine Learning?" <https://www.ibm.com/topics/machine-learning> (accessed Feb. 27, 2023).
- [5] S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence: a modern approach*, 3rd ed. Upper Saddle River: Prentice Hall, 2010.
- [6] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*, 1st edition. Sebastopol, CA: O'Reilly Media, 2017.
- [7] "neuronnätverk - Uppslagsverk - NE.se." <https://www.ne.se/uppslagsverk/encyklopedi/l%C3%A5ng/neuronn%C3%A4tverk> (accessed Feb. 28, 2023).
- [8] Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, och M. S. Nasrin, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches".
- [9] Michael Nielsen. *Neural Networks and Deep Learning*. Springer Science Business Media, 2013.
- [10] Y. LeCun, Y. Bengio, och G. Hinton, "Deep learning", *Nature*, vol. 521, nr 7553, s. 436–444, maj 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [11] A. Kumar, A. Kaur, och M. Kumar, "Face Detection Techniques: A Review", *Artificial Intelligence Review*, vol. 52, aug. 2019, doi: [10.1007/s10462-018-9650-2](https://doi.org/10.1007/s10462-018-9650-2).

- [12] "CS231n Convolutional Neural Networks for Visual Recognition".
<https://cs231n.github.io/convolutional-networks/> (åtkomstdatum 24 mars 2023).

- [13] Ujjwalkarn, "An Intuitive Explanation of Convolutional Neural Networks",
Ujjwal Karn, 10 augusti 2016.
<https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
(åtkomstdatum 30 mars 2023).

- [14] I. Goodfellow, Y. Bengio, och A. Courville, *Deep Learning*. MIT Press,
2016. Åtkomstdatum: 02 april 2023. [Online]. Tillgänglig vid:
<https://www.deeplearningbook.org/>

- [15] Computer Science Wiki.
<https://computersciencewiki.org/index.php/File:MaxpoolSample2.png>
(åtkomstdatum 03 april 2023).