NIBIO POP

NIBIO
NORSK INSTITUTT FOR BIOØKONOMI

VOL. 6 - NO. 35 - 2020

Store mengder med eksempeldata er grunnlaget for at en maskin kan lære seg å skille mellom ulike bildeelementer

Maskinlæring til bildeanalyse

Stadig oftere møter vi begreper som stordata, kunstig intelligens, maskinlæring, nevrale nettverk og dyp læring. Slike «smarte teknologier» har for lengst blitt en del av hverdagen vår – iblant også uten at vi vet det. Denne artikkelen gir en kort innføring i hva maskinlæring er og hvordan denne teknologien kan brukes til automatisk bildeanalyse.

INNLEDNING

Det kan ikke være tvil om at vi lever i en datatidsalder. Spesielt de siste 20 årene har teknologien utviklet seg nesten eksponentielt. Samtidig har også samfunnet vårt forandret seg fundamentalt. Aldri før har vi produsert så store mengder med data – over 2,5 trillioner byte (et tall med 18 nuller) hver eneste dag. Men denne utviklingen skaper også nye utfordringer. De enorme datamengdene som produseres hvert eneste sekund ligger langt utenfor det vi mennesker er i stand til å forholde oss til. Mens vi før klarte å håndtere mye data manuelt, er vi i dag nødt til å satse i økende grad på automatisering. Automatiserte arbeidsprosesser bidrar til å holde styr på dataene og til å analysere komplekse sammenhenger som vi ellers kanskje aldri hadde oppdaget.

For å kunne utføre oppgaver automatisk må hele prosessen vanligvis programmeres på forhånd. Oppskriften som beskriver de enkelte stegene i en slik programmering kalles en «algoritme». For svært komplekse problemstillinger er det vanskelig å skrive en slik algoritme, siden dette allerede forutsetter en viss forståelse for hvordan problemet kan løses. Kunstig intelligens og maskinlæring kan bistå i slike situasjoner og hjelpe oss med å løse de vanskelige oppgavene.

I dag møter vi kunstig intelligens i ulike sammenhenger. Mange av oss bruker f. eks. mobile applikasjoner for å identifisere planter eller fugler med smarttelefonen. Slike apper er trent opp ved bruk av flere hundretusen bilder for å lære hvilke karakteristika som skiller de ulike artene fra hverandre. Også i landbruket finnes det stadig mer av denne teknologien. Allerede i dag bruker mange bønder melkerobot og fôringsautomat, men i økende grad også ulike sensorer, skannere, GPS, satellitt og droner for å overvåke vær, klima, dyr, jord og plantevekst. Presisjonslandbruk er et område i stor vekst som tar i bruk mange av disse teknologiene for å effektivisere ressursbruken og redusere negative miljøpåvirkninger (Korsæth m.fl. 2019). De store datamengdene som samles gjennom slike systemer kan analyseres gjennom maskinlæring f.eks. for å estimere avlingsmengder eller for å oppdage plantesykdommer og spredning av ugress.

HVA ER MASKINLÆRING?

Alle teknologier som på en eller annen måte etterligner menneskelig intelligens samles under begrepet kunstig intelligens (AI*, artificial intelligence*). Maskinlæring (ML, *machine learning*) er en underkategori av kunstig intelligens (Fig. 1). Ved ML er maskinen programmert til å lære å gjenkjenne eller skille signaler (f.eks. bilder) gjennom erfaring. For å lære, trenger ML-algoritmer store mengder med eksempler av det vi er interessert i. Slike eksempler kaller vi treningsdata. I prinsippet kan maskinlæring brukes på alle slags data som f.eks. bilder, tabeller, tekst og lyd. I denne artikkelen ser vi nærmere på bruk av ML til å gjenkjenne eller skille mellom bilder.

Fig. 1: Sammenheng mellom begrepene kunstig intelligens, maskinlæring og dyp læring.

Maskinlæring er et samlebegrep for mange ulike metoder og algoritmer. Disse varier i oppbygging og funksjon. Vi skiller blant annet mellom klassiske ML-algoritmer og dyp læring.

Klassiske ML-algoritmer som *Support Vector Machines* eller *Random Forest* er velprøvde og pålitelige metoder, men krever at det foreligger beskrivelser av dataenes sentrale egenskaper («*features*»), utarbeidet manuelt på forhånd. Hvis vi for eksempel skal beskrive en katt, kan vi gjøre dette gjennom ulike egenskaper som pelsfarge, formen på ører og øyne, størrelse eller vekt. Hver egenskap danner en egen kolonne i datatabellen:

Eksempel	Pelsfarge	Ører	Høyde (cm)	Vekt (kg)	Klasse
1	grå	stående	23	3,6	Katt
2	svart	stående	25	3,8	Katt
3	svart	hengende	52	11,0	Hund
4	brun	stående	33	6,5	Hund
5	hvit	stående	27	4,0	?

Etter å ha fått nok eksempler vil maskinen forhåpentligvis oppdage et mønster i datasettet. Så er pelsfargen alene åpenbart ikke egnet for å skille mellom hund og katt. Katteører er alltid stående, mens en hund kan ha både stående og hengende ører. Tar vi med høyde og vekt har vi nok informasjon for å kunne identifisere eksempel 5 som «katt» med stor sannsynlighet. Likevel er det sjeldent at en modell oppnår 100 % nøyaktighet, fordi de fleste datasett inneholder noen særtilfeller som avviker fra normalen («*outliers*»). For eksempel, kunne det siste eksempelet like godt være en liten hund.

Fordelen med de klassiske ML-algoritmene er at de trenger relativt lite treningsdata. Dermed kan de også anvendes på små datasett. Disse metodene er derimot kun i liten grad egnet til digital bildeanalyse, ettersom de først og fremst ble utviklet for å håndtere data i tabellform.

Dyp læring (DL, *deep learning*) inntar en særskilt posisjon innen maskinlæring. DL bruker kunstige nevrale nettverk som minner litt om det nettverket av neuroner vi finner i hjernen vår. DL er vanligvis førstevalget når målet er å gjenkjenne objekter eller personer i bilder. Utgangspunkt for disse systemene er at et bilde i prinsipp er en todimensjonal matrise av fargeverdier (piksler). Hver for seg har pikslene liten informasjonsverdi, fordi de ikke «vet om» nabopikslene sine. For å kunne finne mønster, er det nødvendig å betrakte pikslene i forhold til hverandre og slå dem sammen til mer meningsfulle enheter. Den største utfordringen innen bildeanalyse er å håndtere de ulike måtene én og samme klasse kan representeres på. Grunnen til dette er at bilder viser en stor variasjon, både innenfor hver klasse (ulike typer, raser, farger), i perspektivet (kameravinkel, avstand), i bakgrunnen (innendørs, utendørs) eller i belysningen (sol, skygge).

Fordelen med dyp læring er at maskinen selvstendig henter ut all informasjon fra bildene som er nødvendig for å identifisere noe som f.eks. som en katt. Det betyr at vi slipper å bruke tid på å beskrive de sentrale egenskapene manuelt. Ulempen med å overlate dette til maskinen er at den gjerne velger komplekse og abstrakte mønster som kun maskinen selv er i stand til å tolke. Generelt oppfører en DL-modell seg som en «svart boks», dvs. at vi ikke vet hva som skjer i maskinen og hvordan den kom fram til svaret. Derfor egner DL seg best i tilfeller hvor vi ikke har behov for å forstå løsningsprosessen. En vesentlig begrensning i bruken av DL er videre at maskinen trenger enorme mengder treningsdata før vi kan forvente å få gode resultater. Avhengig av problemstillingen kan det kreve flere tusen til hundretusen eksempler som treningsdata.

Supervised learning
Maskinen trenes med data som er merket med riktig svar. Formålet er å tilordne data til en klasse (klassifikasjon) eller til å predikere et kontinuerlig tall (regresjon)
Typiske algoritmer: Support Vector Machines, Neural Networks, Random Forest

Unsupervised learning
Maskinen lærer helt selvstendig gjennom umerkede treningsdata uten å få noe mer informasjon i tillegg. Formalet er å strukturere data eller oppdage mønster i store datamengder.
Typiske algoritmer: k-means Clusteranalyse, PCA

HVORDAN EN MASKIN LÆRER: MASKINLÆRING MED LANDSKAPSBILDER

I en test med landskapsfotografier har vi brukt maskinlæring for å gjenkjenne busk- og tre-vegetasjon. Modellen i denne studien er et «*Convolutional Neural Network*» (CNN) som faller innenfor kategorien dyp læring med nevrale nettverk. Dette er et praktisk eksempel som kan bidra til å forklare hvordan en maskin faktisk lærer.

Analysen begrenset seg til å skille mellom to mulige klasser. Den ene klassen bestod av det vi er interessert i, nemlig busk- og trevegetasjon. Den andre klassen omfattet alt annet vi finner i landskapsbildene som f.eks. gress, vann, bygninger, fjell, asfalt, dyr og mennesker. Vi samlet et stort datasett med over 50 000 små eksempelbilder (50x50 pikselstørrelse). Disse var utsnitt fra større fotografier, klippet ut slik at de enten viste busk- og trevegetasjon eller helt andre fenomener – aldri en blanding. Eksempelbildene ble klassifisert i «positive» eksempler (som viste busk- og trevegetasjon) og «negative» eksempler (som viste andre objekter).

Før treningen starter, deles hele datasettet tilfeldig i tre ulike deler (Fig. 2). Spesielt viktig er det å holde tilbake et uavhengig testsett som brukes først etter at treningen er avsluttet.

TRENINGSDATA

Dataene som brukes for å trene og evaluere en ML-modell deles vanligvis i tre ulike datasett. Mest vanlig er supervised learning hvor hvert eksempel er merket med riktig svar

Trening
Eksempler som brukes for å trene en ML-modell

60 til 80 % av eksemplene

Validering
Datasett som modellen bruker parallelt med treningsdataene for å optimalisere modell-parametrene
10 til 20 %

Test
Brukes etter avsluttet trening for å teste hvor godt den ferdige modellen fungerer på helt nye data
10 til 20 %

Fig. 2: Eksempeldataene deles vanligvis i tre datasett: Treningsdata, valideringsdata og testdata.

Sammen med treningsdataene får modellen også vite hva fasiten er, altså hva hvert eksempelbilde faktisk viser. Basert på denne informasjonen, prøver modellen å finne ut hvilke egenskaper i bildene som skiller de to klassene fra hverandre. Første steg for data-

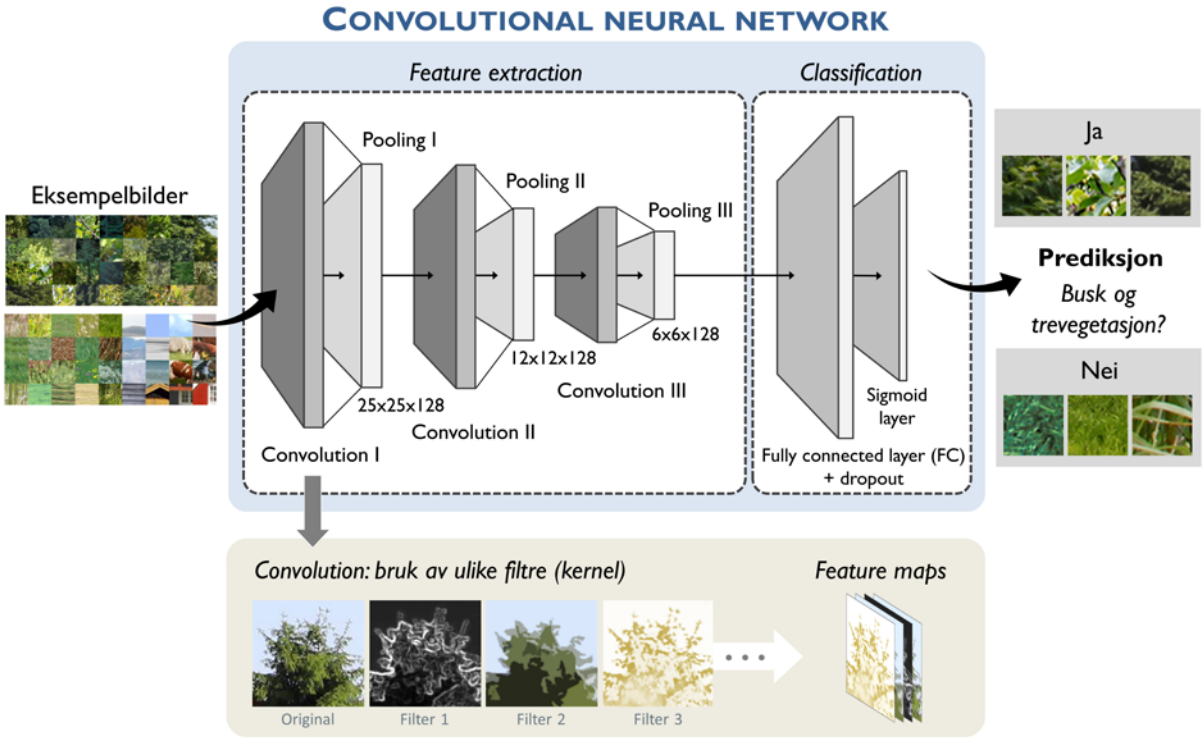

Fig. 3: Oppbygging av en standard CNN-modell. Illustrasjonen viser treningsfasen der modellen får et stort antall eksempelbilder for å lære hvordan ulike klasser kan skilles fra hverandre.

maskinen er derfor å hente ut så mye informasjon som mulig fra eksemplene («*feature extraction*»). I en CNN-modell skjer dette automatisk ved at dataene behandles med et stort antall forskjellige filtre som samler informasjon om farge, kanter og tekstur. Modellen i Figur 3 består av tre «*convolutions*» eller konvolusjoner. I hver konvolusjon samler maskinen en rekke egenskaper ved bruk av ulike filtre. Resultatet er en stabel med «*feature maps*» som igjen brukes som input i den etterfølgende konvolusjonen (med nye filtre). Dette gjør at egenskapene som maskinen henter ut øker i kompleksiteten med hver konvolusjon.

I neste steg prøver maskinen å kombinere den informasjonen som er samlet inn på en (for oss ukjent) måte for å kunne skille mellom de ulike klassene. Læringen skjer gjennom en gjentagende prosess hvor modellen etter hver runde sammenligner sine resultater med fasiten. Hvis prediksjonen ikke er god nok, justerer modellen noen av sine innstillinger (modellparametre) og prøver å løse oppgaven på nytt. Metoden kan derfor beskrives som «prøving og feiling». Figur 4 viser en typisk treningskurve og hvordan modellen gradvis blir bedre. Modellen starter med en relativ lav nøyaktighet på 64 %, men læringsprosessen gjentar seg til modellen har oppnådd maksimal nøyaktighet. Treningsfasen stopper når kurven flater ut, altså når flere omganger ikke fører til en ytterligere forbedring. I vårt tilfelle var dette etter 90 omganger og en maksimal nøyaktighet på rundt 95 % for treningsdataene.

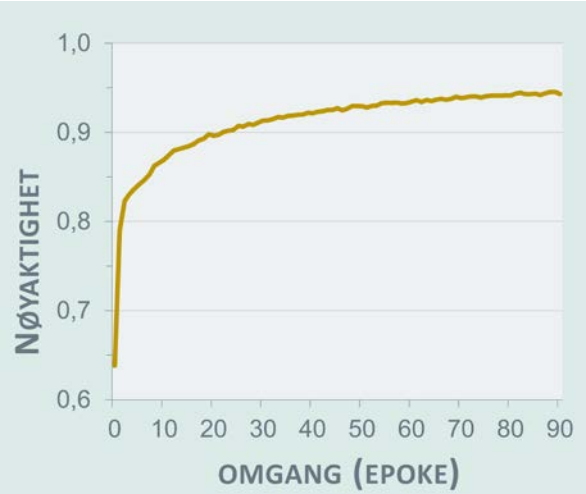

Fig. 4: Grafen viser hvordan modellens nøyaktighet forbedrer seg gradvis etter hver gjennomgang.

Så snart modellen er ferdig trent og oppnår gode resultater på treningsdataene kan den anvendes på helt nye data som modellen ikke har sett før (testdata). Dette er det mest kritiske øyeblikket i prosessen siden det forteller oss noe om modellens generaliseringsevne. Først når modellen får gode resultater også på uavhengige testdata, vet vi om den virkelig fungerer.

CNN-modellen vi brukte i testen klarte til slutt å lære seg selvstendig å skille mellom de to klassene med 95 % nøyaktighet på enkelte 50x50 piksel store testbilder og gjennomsnittlig 87 % nøyaktighet når

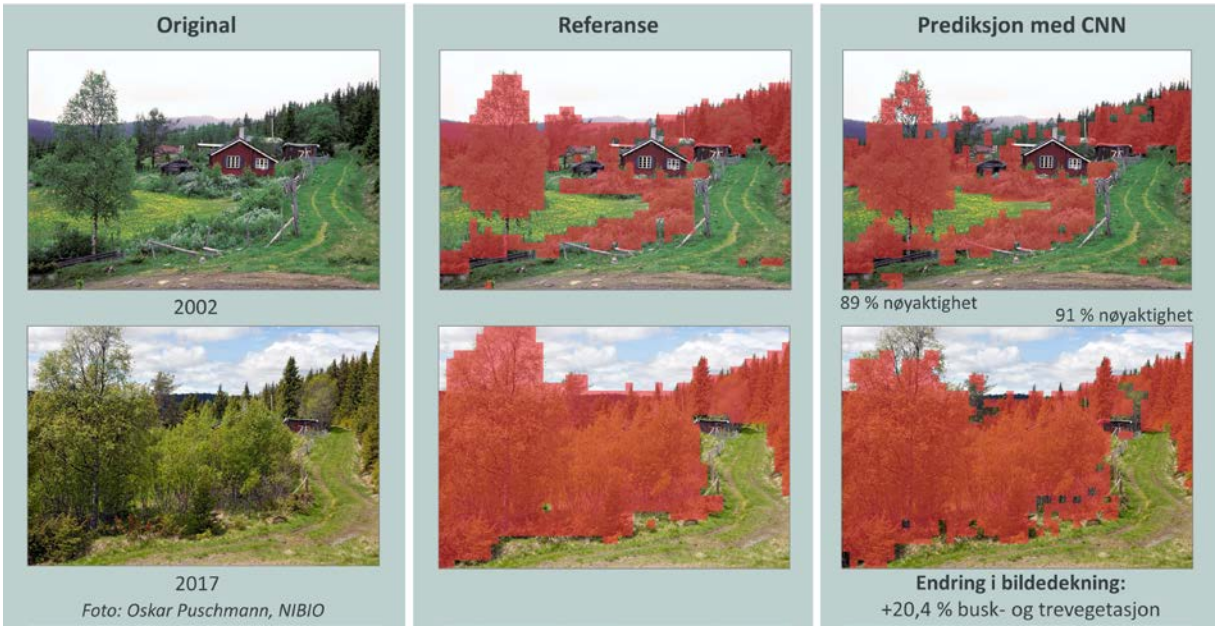

Fig. 5: Bruk av den trente CNN-modellen til å automatisk klassifisere busk- og trevegetasjon i landskapsbilder. Resultatene til høyre viser at modellen hadde noen problemer i overgangen mellom vegetasjon og himmel.

hele landskapsbilder ble klassifisert. Anvendt på refotograferinger som NIBIO bruker som del av prosjektet «Tilbakeblikk» og i 3Q-overvåkingen blir det derved mulig å kvantifisere relative endringer i vegetasjonsdekningen i foto på en automatisert måte (Bayr & Puschmann 2019). Figur 5 viser som eksempel et endringsbilde tatt i Saupeset, Buskerud. En manuell klassifikasjon tjener som referanse (midten) for å kunne evaluere modellens nøyaktighet. Bildene til høyre viser resultatet av modellen. Dette forteller oss at bildedekningen av busker og trær har økt med 20 % mellom 2002 og 2017. Selv om anslaget ikke sier noe om den faktiske arealdekningen, så gir resultatet en god indikasjon på gjengroing i landskapet.

Resultatene gjorde det også tydelig at bildekvaliteten spiller en stor rolle for hvor nøyaktig modellen klarer å predikere busk- og trevegetasjon (Fig. 6). Eldre fotografier og bilder som ble tatt under dårlige lysforhold har ofte lavere oppløsning. Dette gjør at kanter og teksturen til objekter forsvinner og med dette informasjon som er svært viktig for å kunne gjenkjenne busk- og trevegetasjon. Den samme effekten ser vi i bakgrunnen av Figur 5 hvor modellen ikke klarte å klassifisere de skogkledde fjellene i det fjerne.

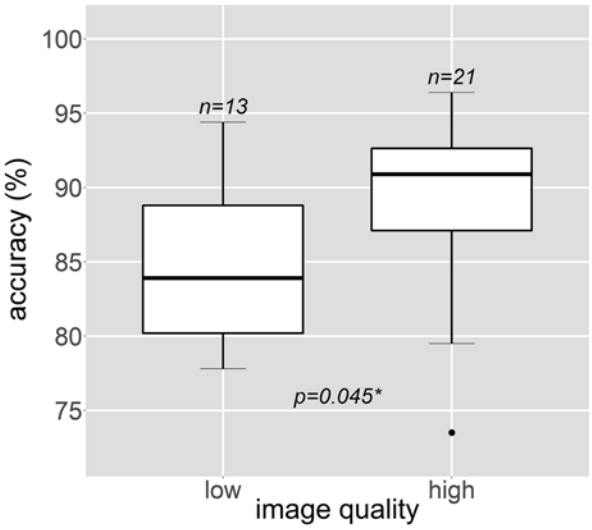

Fig. 6: Bilder som ble tatt under gode lysforhold er ofte skarpere og ga bedre resultater.

EN DIGITAL FREMTID

Utviklingen innen datateknologi har gjort enorme framskritt de siste 20 årene og det er rimelig å anta at flere store teknologiske gjennombrudd er rett rundt hjørnet. Allerede i dag bidrar kunstig intelligens og maskinlæring til å håndtere og analysere store datamengder på en effektiv måte innen mange ulike områder. Automatisk bildeanalyse er en av de viktigste bruksområder for maskinlæring. Spesielt i landbruket har kunstig intelligens et potensial for å bidra til gode og framtidsrettete løsninger for en mer bærekraftig matproduksjon og ressursforvaltning. Så lenge vi lykkes i å bruke kunstig intelligens på en ansvarlig måte, kan det åpne opp for mange spennende muligheter og bidra til å finne gode løsninger på noen av de store samfunnsutfordringene.

REFERANSER

Bayr U. & O. Puschmann (2019): Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. *Ecological Informatics 50.* https://doi.org/10.1016/j.ecoinf.2019.01.012

Korsæth A., Johansen Lindgaard H., Veidal A. & L.J. Asheim (2019): Utbredelse og potensiell økonomisk og miljømessig nytteverdi med presisjonsjordbruk i Norge. NIBIO Rapport 5(41). http://hdl.handle.net/11250/2591261

FORFATTER:

Ulrike Bayr, Divisjon for kart og statistikk,
Avdeling Landskapsovervåking
ulrike.bayr@nibio.no

NIBIO-**POP** 6(35)2020
ISBN 978-82-17-02638-9
ISSN 2464-1170

Fagredaktør: Hildegunn Norheim, divisjonsdirektør kart og statistikk
Ansvarlig redaktør: Per Stålnacke, forskningsdirektør, NIBIO

nibio.no