



ERFARINGER MED DEEP LEARNING OG JORD-OBSERVATIONSDATA



Karsten Østergaard Noe
Alexandra Institutttet



Morten Henriksen Birk
FieldSense

Både hos Alexandra Institutttet og hos FieldSense har vi i løbet af de seneste år arbejdet med brug af maskinlæringsteknikken deep learning til at analysere jordobservationsdata. I denne artikel vil vi gerne dele ud af vores erfaringer med at kombinere rasterlag og polygonlag med deep learning-teknikker, hvorved man kan optræne neurale netværk der kan automatisere arbejdsopgaver i forbindelse med kortlægning og ændringsudpegning. Vi vil give eksempler på, hvordan komplekse informationer kan udvindes fra satellit- eller flyfotos via kunstig intelligens, når store mængder af data, og lige så vigtigt den nødvendige ground truth data, er tilgængelig. I samarbejde med Alexandra Institutttet har FieldSense f.eks. udviklet kunstig intelligens til at identificere markgrænser og generere matrikeldata ud fra satellitdata. Desuden vil vi fortælle om en metode til visuel søgning efter objekter i meget store rasterbilleder.

Keywords: Remote sensing, jordobservation, flyfotos, ortofotos, satellit, droner, deep learning, segmentering, ændringsudpegning, maskinlæring.

UDNYTTELSE AF RUMSYSTEMER TIL ØGET VÆKST

Alexandra Institutttet er en almennyttig virksomhed, der arbejder med forskningsbaseret it-innovation. I 2007 blev Alexandra Institutttet udpeget til Godkendt Teknologisk Service (GTS)-Institut af Uddannelses- og Forskningsministeriet. Hermed følger en forpligtelse til at hjælpe danske virksomheder med at anvende den nyeste forskning og teknologi.

Det er netop Alexandra Instituttets status som GTS-institut, der har dannet basis for deltagelsen i projektet "Udnyttelse af rumsystemer til øget vækst", som er et samarbejde med Force Technology. Projektet er forløbet i forlængelse af den nationale rumstrategi i 2017 og 2018.

Projektet har både dækket downstream- og upstream-aktiviteter. Sigtet med projektet har været at styrke potentielle danske rumvirksomheder, herunder særligt SMV'er, med viden og teknologier til at udnytte det forventede potentiale indenfor rumområdet. I denne artikel vil vi fortælle om nogle af downstream-aktiviteterne i projektet, som handler om at bruge kunstig intelligens til at fortolke jordobservationsdata.

FRIE GRUNDDATA OG KUNSTIG INTELLIGENS

EUs Copernicus-satellitprogram har givet verden adgang til en ny og værdifuld datakilde. Det er bl.a. denne datakilde, der danner grundlag for FieldSense's platform, og det er også den datakilde, vi har udnyttet til at segmentere marker som beskrevet nedenfor.

Herudover har vi i Danmark fri adgang til et væld af frie grunddata som for eksempel ortofotos (forår og sommer), højdekort, skråfotos og diverse polygonlag. Det, at vi har jordobservations-billed-

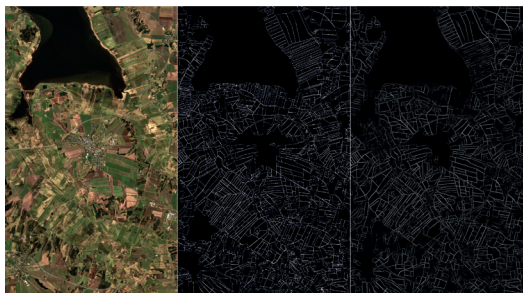
data til rådighed sammen med polygondata, som kan bruges som "ground-truth"-kilde til maskinlæring, har potentialet til at muliggøre automatisering af billedfortolknings-opgaver, der i dag er tidskrævende, manuelle processer. Et eksempel på dette er segmentering og ændringsudpegning af bygninger på ortofotos, som vi vil fortælle mere om nedenfor.

MARKGRÆNSER FRA SENTINEL-SATELLITBILLEDER

FieldSense A/S er en teknologivirksomhed med base i Aarhus, der leverer beslutningsstøttesystemer til landmænd og rådgivere, samt specialiserede løsninger til større virksomheder i landbrugsindustrien. Virksomheden har blandt andet udviklet FieldSense platformen: et abonnement-baseret beslutningsstøttesystem der, baseret på multispektrale satellitbilleder og lokale vejrstationer, gør brugere i stand til at overvåge udviklingen af afgrøder, oprette variable tildelingskort til gødskning, og monitorere vejrforholdene for de enkelte afgrøder. Yderligere specialiserer FieldSense sig i skalerbar processing af data, hvor blandt andet kunstig intelligens er med til at automatisere distributionen af store kvantiteter af data internt og til tredjeparter.



Figur 1. Datagrundlag for marksegmentering. Til venstre ses RGB billedet fra et skyfrit Sentinel-2 produkt. Til højre ses de polygoner vi bruger til at træne vores netværk.



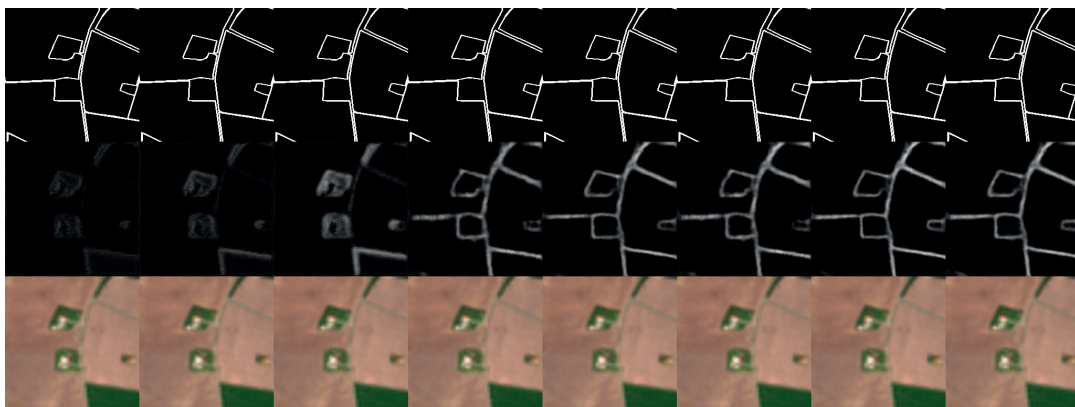
Figur 2. Markgrænser fundet med deep learning (til højre). Ground truth ses i midten.

Størstedelen af data i FieldSense platformen er knyttet til specifikke marker, og det er derfor essentielt for produktet, at vi ved hvor hvert enkelt mark befinder sig, således at hver enkelt måling kan associeres til den mark hvorfra den stammer. I Danmark er informationen om de enkelte markgrænser tilgængelig takket være de markblokke, der udstilles af Landbrugsstyrelsen. Dette er dog ikke tilfældet i udlandet, det er derfor helt essentielt for skaleringen af produktet, at vi automatisk og effektivt kan detektere de enkelte markgrænser i alle områder hvor vi indsamler data.

I FieldSense anvender vi i høj grad billeder fra Sentinel-2, så vi har alle historiske L2A produkter tilgængelige for ethvert område hvor vi opererer. Da vi i forvejen har denne data ønsker vi at vores løsning skal kunne detektere markgrænserne udelukkende fra denne, relativt lavt opløselige, datakilde. Vores løsning er baseret på flere neurale

netværk, der alle tager de 4 bånd RGB+NIR fra Sentinel-2 satellitterne som input. Deep Learning er ofte utroligt effektivt til at udlede komplekse informationer fra billeder, men det kræver store mængder af træningsdata og, eftersom vi arbejder med supervised learning, tilhørende labels til vores træningsdata. De store mængder af data har vi tilgængeligt, da vi har flere terabytes Sentinel-2 data tilgængeligt for Danmark alene. De tilsvarende labels får vi fra markblokkene der er tilgængelige fra Landbrugsstyrelsen.

Givet ovenstående input og tilhørende labels, kan vi træne et neuralt netværk til at generere et billede hvor grænserne til de enkelte marker tegnes ud fra RGB + NIR båndene fra et Sentinel-2 produkt. Altså er det neurale netværks opgave at konvertere de 4 bånd fra Sentinel-2 produktet til et billede med et enkelt bånd, der for hver pixel angiver sandsynligheden for at denne udgør en del af en markgrænse. Dette er illustreret i figur 2, hvor vi ser henholdsvis input billedet, de markgrænser der reelt er på billedet, og det billede som det trænedes neurale netværk genererer. Eftersom input billederne er enormt store og ofte af variabel størrelse, kan vi ikke give billedet direkte som input til det neurale netværk. I stedet inddeles input billedet i små dele der hver er 256x256 pixels, og hver af disse anvendes som input til netværket, der vil producere et billede i tilsvarende størrelse. Vi kan efterfølgende samle resultaterne fra det neurale netværks kørsel på alle de ge-

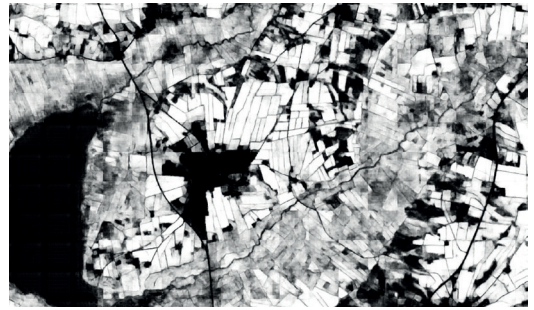


Figur 3. Udvikling af træning over tid.

nererede udklip til et stort raster billede med samme dimensioner som det originale input. Her er det dog nødvendigt at vi kører netværket på udklip der overlapper hinanden, da vi ellers vil se tydelige linjer langs kanterne af hvert udklip. Figur 3 viser for et enkelt af disse udklip hvordan netværket igennem træningen bliver bedre til at udpege de pixels der udgør en markgrænse.

Ud fra de detekterede markgrænser kan vi segmentere inputbilledet i en række polygoner. Det er dog kun en delmængde af disse polygoner der faktisk udgør marker, og vi er derfor nødt til at frasortere alle de polygoner der ikke udgør en mark. Til at løse dette problem træner vi endnu et neuralt netværk, der denne gang har til opgave at afgøre for hver enkelt pixel om denne udgør en del af en mark eller ej. Dette neurale netværk vil altså, givet det samme input data som tidligere, ikke afgøre om hver enkelt pixel ligger på grænsen af en mark, men derimod om det ligger inden for grænsen af en mark. Dette netværk vil generere et output som vist på figur 4, hvor det eksempelvis er tydeligt at vand og byområderne har meget lave marksandsynligheder, og derfor vil disse områder kunne blive sorteret fra.

Med de to ovenstående neurale netværk er det blevet muligt fuldt automatisk at detektere de enkelte marker i et billede. Det skal dog bemærkes, at vi, fordi vi har valgt en datakilde med en opløsning på 10 meter, ikke har plads til mange



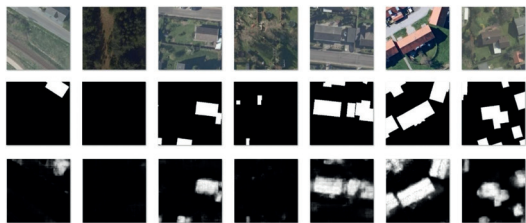
Figur 4. Marksandsynlighedskort, hvor sort angiver en lav sandsynlighed for at en pixel ligger inden for en mark og hvid angiver en høj sandsynlighed.

fejl i vores detekterede markgrænser. En fejl på få pixels vil altså hurtigt resultere i fejl på relativt store afstande. Sådanne fejl kan gøre det langt mere besværligt at anvende vores data til eksempelvis graderet gødskning, og vi ønsker derfor meget høj præcision af de detekterede grænser. Vi har derfor anvendt en teknik, hvor vi bruger yderligere et neuralt netværk til at forfine de detekterede markpolygoner. Netværket er trænet til at kigge på hver enkelt af de detekterede marker en ad gangen, for herefter at rette små fejl i grænserne til uden brug af andet data end det input vi hele tiden har brugt.

På figur 5 ses resultatet af marksegmenteringen, hvor billedet er blevet inddelt i polygoner for hver enkelt mark. Til venstre ser vi hvordan resultatet ser ud inden forfiningen har rettet markerne til, og til højre ser vi det endelige resultat.



Figur 5. Forfining af marksegmentering. Området i nederste højre hjørne er forstørret.



Figur 6. Eksempler på output fra bygningssegmenteringsnetværk (nederste række). Ground truth ses på midterste række.

SEGMENTERING AF BYGNINGER

På baggrund af erfaringerne fra marksegmenteringsprojektet har Alexandra Instituttet været så heldige at få til opgave for Styrelsen for Dataforsyning og Effektivisering (SDFE) at udforske brug af lignende teknologi til segmentering af bygninger fra ortofotos. Den tænkte anvendelse af dette er ændringsudpegning - altså afgørelse af, hvor der er opført eller nedrevet bygninger i tidsrummet mellem to ortofotos. I dette pilotprojekt har SDFE's GeoDanmark-database over bygninger udgjort

ground truth-datasættet. Rasterdatasættet har bestået af 55 quick-ortofotos (RGB-NIR, 16543x5512 pixels) af primært parcelhus-områder. For at undgå problemer med, at perspektivet i billederne flytter tagene relativt til bygningernes fundament har vi valgt kun at anvende den midterste tredjedel af billederne, hvor synsvinklen er "top-down". Som beskrevet i afsnit 2 blev rasterbillederne delt op i træningseksempler på 256x256 pixels. Der blev trænet et neuralt netværk i stil med den, som blev anvendt til marksegmentering. På figur 6 kan man se eksempler på træningsdata, samt hvilke output det trænedede netværk giver for disse eksempler.

På figur 7 kan man se et eksempel på evalueringen af det trænedede neurale netværk på et quick-ortofoto (som ikke var del af træningsdatasættet). Generelt kan man sige, at netværket genkender langt de fleste bygninger, men der er dog stor forskel på, hvor pænt segmenteringen følger bygningernes footprint, som givet i GeoDanmark-databasen. Da det på sigt er visionen at segmenteringsnetværket skal kunne gene-



Figur 7. Anvendelse af deep learning til segmentering af bygninger (øverste). GeoDanmark polygoner nederst.



Figur 8. Forfining af bygningssegmentering.

rere korrekte bygningspolygoner, arbejder vi på at inkludere flere lag i det neurale netværk, så vi kan opnå "forfining" af segmenteringen i stil med det, vi har beskrevet i afsnit 2 for marksegmenteringen. Et foreløbigt resultat af denne forfining kan ses på figur 8.

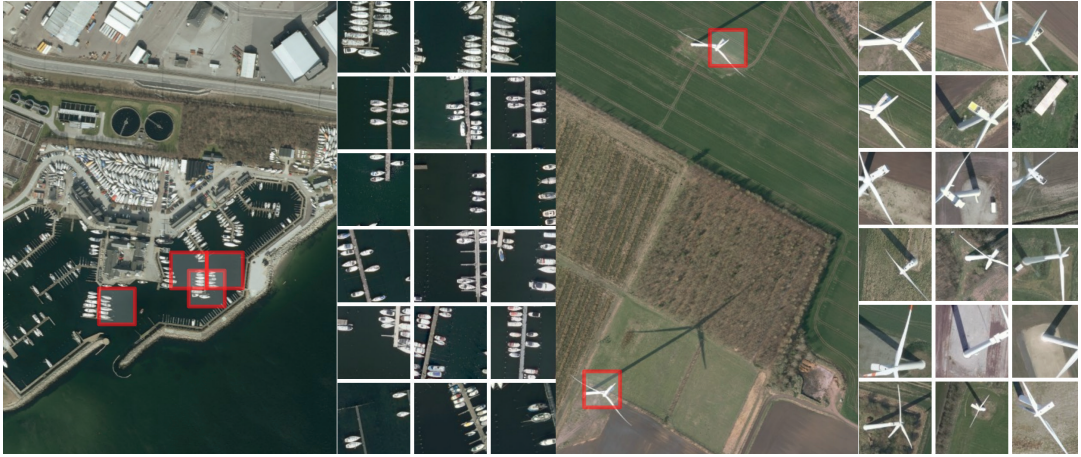
Vi har med det trænede netværk også set områder, der fejlagtigt blev segmenteret som bygning. Her var f.eks. en gentaget fejl, at veje med visse typer asfalt blev fortolket som bygninger. Vores bud på en forklaring af dette er, at disse veje er svære at skelne fra bygninger med tagpap. For at afhjælpe dette og lignende problemer er vi på nuværende tidspunkt i gang med et opfølgingsprojekt med SDFE, hvor der til træningen udover RGB- og NIR-båndene fra ortofotos også anvendes et højdekort, som er blevet genereret ud fra billeder fra samme overflyvning via fotogrammetri. Foreløbige resultater indikerer, at dette afhjælper de fleste af de problemer vi har set hidtil med det nuværende neurale netværk.

DEEP LEARNING OG SØGNING I MEGET STORE RASTER-BILLEDER

De store datamængder i f.eks. Danmarks ortofoto-datasæt giver nogle praktiske udfordringer, når

billedmaterialet skal fortolkes og analyseres. Hvordan bærer vi os for eksempel ad hvis vi gerne effektivt vil søge efter objekter i meget store rasterdatasæt? Det har det amerikanske firma Descartes Labs givet et bud på med deres løsning "geovisual search". Teknologien gør det muligt at klikke på et objekt i et rasterdatasæt, og en effektiv algoritme finder andre områder i datasættet, der ligner. Således kan man klikke på en vindmølle, og algoritmen vil finde andre vindmøller i et rasterdatasæt, der dækker hele USA.

For at kunne udforske mulighederne ved denne teknologi satte vi i Alexandra Institutet os for i samarbejde med DHI-GRAS at udvikle vores egen variant af søgeteknologien og arbejde med bl.a. et dansk ortofoto-datasæt. Ligesom det amerikanske forbillede gør søgeteknologien brug af deep learning-teknikker. Vi har taget afsæt i et neuralt netværk af typen ResNet, som er blevet trænet til at kunne skelne mellem billeder af 1000 forskellige typer "hverdagsobjekter", som for eksempel hunde, katte, biler, cykler og så videre. Som et resultat af optræningen, har dette netværk "lært" at skabe en meget kompakt repræsentation af input-billedet bestående af 2048 tal. Denne kompakte repræsentation kalder vi en "descriptor" for input-billedet,



Figur 9. Eksempler på søgninger med geovisual search.

og den indeholder den information, der er nødvendig for at kunne skelne mellem de 1000 objekttyper. Her kommer det smarte; denne kompakte repræsentation kan vi bruge til at søge i - vi kan simpelthen ved at sammenligne descriptors for to billeder få et mål for, hvor ens to billeder er i den forstand, om de indeholder samme type objekt!

Udgangspunktet for ”geovisual search” er således, at rasterbilledet deles op i et stort antal udklip, som hver især køres gennem et ResNet, så vi får en descriptor per udklip, som kan bruges til at søge i. I vores tilfælde med danmarks-ortofoto-datasættet har vi delt datasættet op i 48 mio. udklip, og som udgangspunkt fylder descriptors for disse 393 GB. Ved hjælp af datakomprimering kan alle descriptors være i hukommelsen på en normal computer og med lidt effektiv parallelisering af beregningerne kan vi sammenligne descriptor for billedet med et objekt, vi søger efter, med alle 48 mio. descriptors og finde de 100, der ligner bedst, på under 80 millisekunder!

Vi har valgt at lave en web-baseret demonstration af denne søgning, som alle frit kan tilgå og afprøve. Den kan findes på adressen denmark3d.alexandra.dk/geosearch. Den fungerer sådan, at man zoomer ind på et interessant objekt og klikker på dette. Herefter vil algoritmen finde de 100 udklip, som ligner bedst ifølge de beregnede descriptors. På figur 9 er vist et par eksempler på søgnin-

ger - i dette tilfælde henholdsvis en lystbådehavn og en vindmølle.

Hvis man leger lidt med demonstratoren vil man opdage, at den har en forkærlighed for at finde objekter med samme orientering som det objekt, man har klikket på. Dette er en konsekvens af det netværk, vi har taget udgangspunkt i, der som nævnt er trænet i at skelne mellem hverdagsobjekter, der vender ”rigtigt”. Netværket er således ikke trænet i at genkende katte og hunde, der vender på hovedet! Dette forventer vi vil være teknisk muligt at afhjælpe ved at optræne et netværk, som er rotationsinvariant, som vi så beregner descriptors ud fra. En anden ting, man hurtigt kommer til at savne, er muligheden for at søge efter objekter i forskellig skala. Da alle descriptors i demonstratoren svarer til udklip i samme skala (224x224 pixels á 20 cm) vil dette definere den skala, der kan foretages søgning på. Det vil naturligvis være teknisk muligt at tillade søgning i andre skalaer på bekostning af at udklippe og beregne descriptors i forskellig skala.

Vi håber at interesserede læsere vil eksperimentere med demonstratoren og endelig ikke tøve med at tage fat i forfatterne hvis I ser potentielle anvendelser af teknologien. Udover den anvendelse, som er skitseret i demonstratoren har vi også en række ideer til andre anvendelser af teknologien. F.eks. kan man forestille sig at det er muligt at lave æn-

dringsudpegning eller areal-anvendelses-estimerer på basis af de beregnede descriptors, men det må komme an på senere eksperimenter!

FREMTIDEN FOR DEEP LEARNING INDEN FOR GIS OG JORDOBSERVATION

Deep learning er et meget alsidigt og værdifuldt værktøj, som har revolutioneret tekniske discipliner så forskellige som billedanalyse (computer vision) og talegenkendelse. Det er vores opfattelse, at vi kun har set begyndelsen på brugen af dette

værktøj inden for jordobservation. Man siger nogen gange, at "hvis det eneste værktøj, du ejer, er en hammer, begynder ethvert problem at ligne et søm". Der er meget hype omkring deep learning, og vi kommer utvivlsomt til at se anvendelser af deep learning, hvor andre teknikker med fordel kunne have været anvendt i stedet. Men vi kommer også til at se værktøjer baseret på deep learning, som kommer til at vende op og ned på vores opfattelse af, hvad der er praktisk muligt. Det bliver spændende at følge!