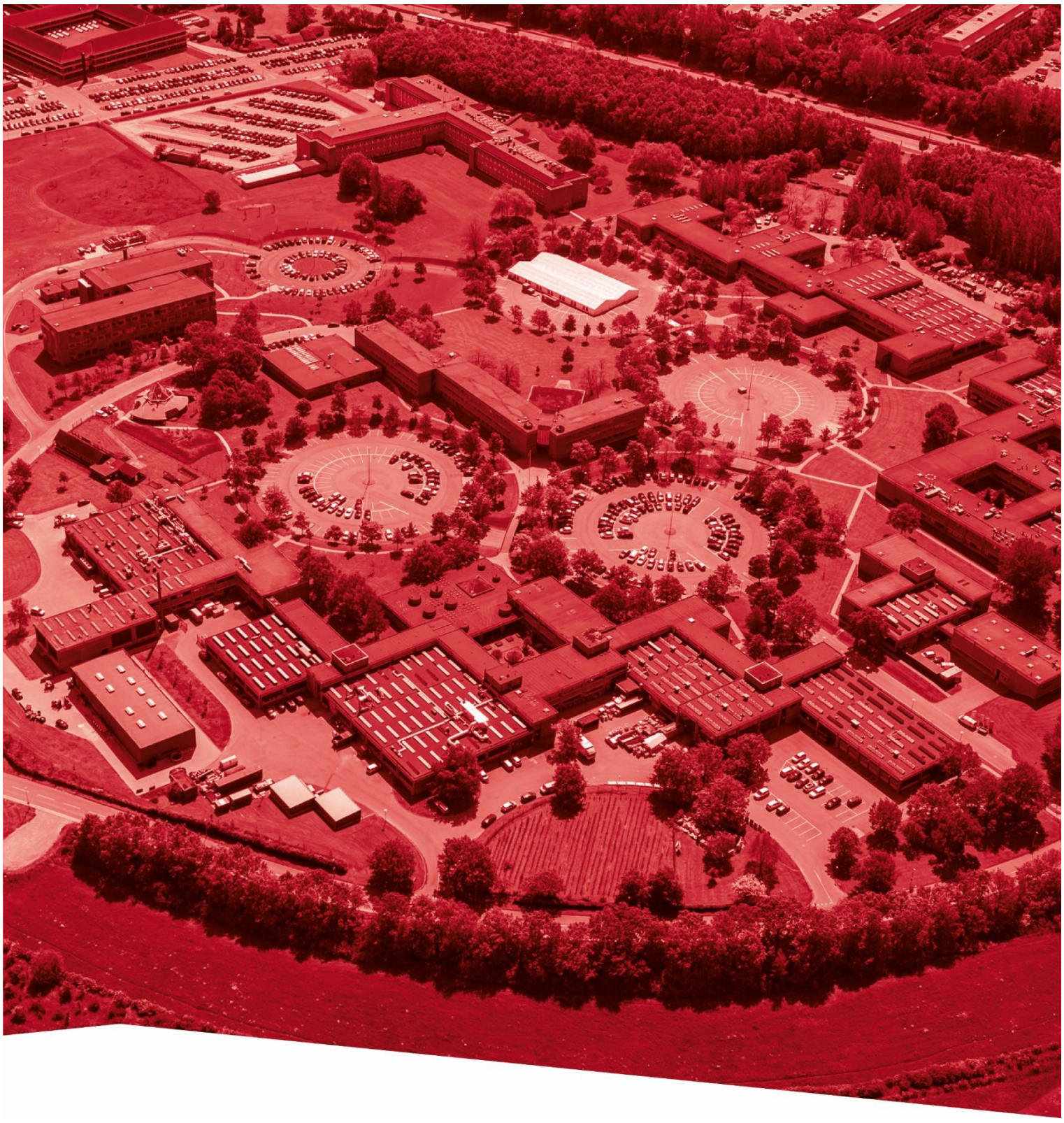


Effektiv annotering af AI træningsdata

TEKNOLOGISK INSTITUT

TEKNOLOGISK
INSTITUT

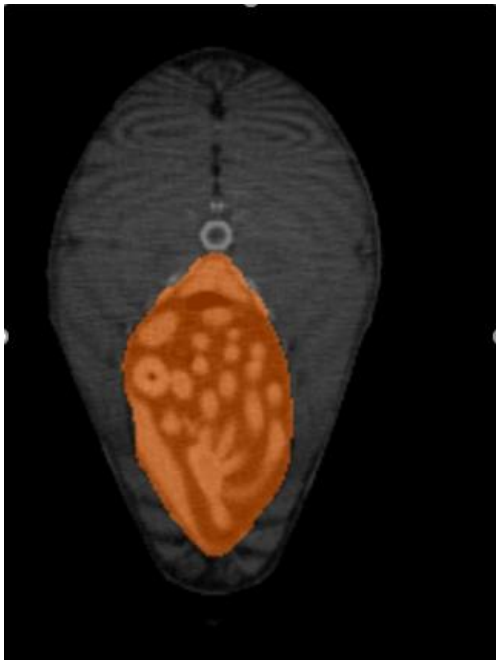
Effektiv annotering af AI træningsdata

Udarbejdet af:
Teknologisk Institut
Gregersensvej 9
2630 Taastrup
DMRI

December 2021
Forfatter: PAHD

1. Baggrund

De seneste år er brugen af dybe neurale netværk, kendt som *Deep Learning*, blevet mere og mere udbredt til at analysere billeder i forskellige anvendelser; eksempelvis til identifikation af forekomst af specifikke objekter på billedet (fx er der en kat på billedet?) eller mere generelt en klassifikation af billeder (fx i kategorier med katte, hunde, heste etc.). Udover at klassificere kan *Deep Learning*-teknologien også bruges til at finde placeringen på billedet af specifikke objekter eller features (maskering), se fx figur 1.

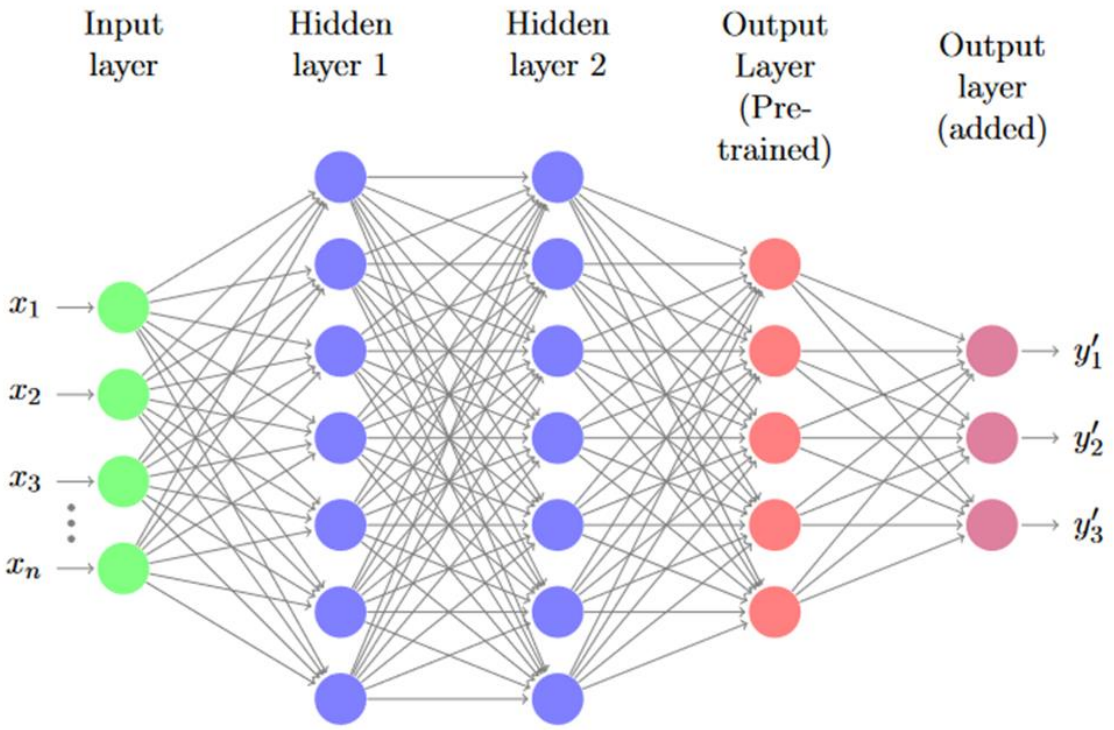


Figur 1. Et eksempel på output fra et *Deep Learning*-netværk, hvor de indre organer i et tværsnitsbillede af en fisk er fundet og maskeret (i rødbrunt). Billedet er taget fra CompleteSCAN-projektet, hvor Teknologisk Institut, DMRI har CT-scannet hele laks og udviklet *Deep Learning*-algoritmer til automatisk at finde og fjerne indvoldene, hovedet og finnerne for at kunne bestemme filetudbytterne.

Den stigende udbredelse af *Deep Learning* skyldes dels, at teknologien er yderst effektiv til at lave billedanalyser, som det med traditionelle billedanalyseteknikker ville være yderst kompliceret og svært at udvikle, dels at værktøjer til at designe, træne og eksekvere *Deep Learning*-netværk de seneste år er blevet udviklet og gjort tilgængelige fra de store Tech-virksomheder (fx TensorFlow fra Google og PyTorch fra Facebook).

For at træne et *Deep Learning*-billedanalysenetværk fra bunden skal der bruges referencedata i form af mange (hundrede-)tusinde annoterede billeder. Dvs. billeder, hvor det er angivet, hvad der er på billedet, og ofte også hvor på billedet det befinder sig. Heldigvis vil man ofte kunne tage udgangspunkt i et

netværk, som allerede er trænet til et andet formål, og nøjes med at ændre de sidste lag af netværket eller tilføje lag og gentræne det med billeder, som er relevante for det nye formål.



Figur 2. Eksempel på modifikation af et prætrænet netværk.

Selvom man benytter sig af et prætrænet netværk, skal der ofte bruges mange billeder, og generelt indebærer brugen af *Deep Learning*, at man bruger mindre tid på at udvikle algoritmer, men mere tid på at forberede referencematerialet, i form af annoterede billeder, til at træne netværket.

Der er derfor brug for metoder og værktøjer, der kan effektivisere arbejdet med at generere de annoterede billeder til *Deep Learning*, så de økonomiske omkostninger til udvikling og implementering af løsninger baseret på *Deep Learning* kan reduceres. I det SAF-finansierede projekt "Nye målemetoder til kødindustrien" i 2021undersøgtes nye teknologier og metoder for deres industrielle potentiale. Pga. den biologiske variation i kødindustrien, som er et grundvilkår, alle løsninger skal kunne håndtere, er *Deep Learning* yderst velegnet til mange billedbaserede løsninger til kødindustrien (fx til produktidentifikation, procesovervågning, kvalitetsvurdering og maskinstyring). Derfor undersøgtes metoder og værktøjer til at effektivisere annoteringen i førnævnte SAF-projekt. Indeværende rapport er en sammenfatning af vores vidensindsamling, erfaringer og konklusioner på dette område.

2. Metoder

2.1. "*Brute Force*" manuel annotering af de mange billeder

Den første og umiddelbart letteste metode til at effektivisere annoteringen er at søge at reducere lønomkostningen til annotering ved at reducere timelønnen. Der er to tilgange til dette:

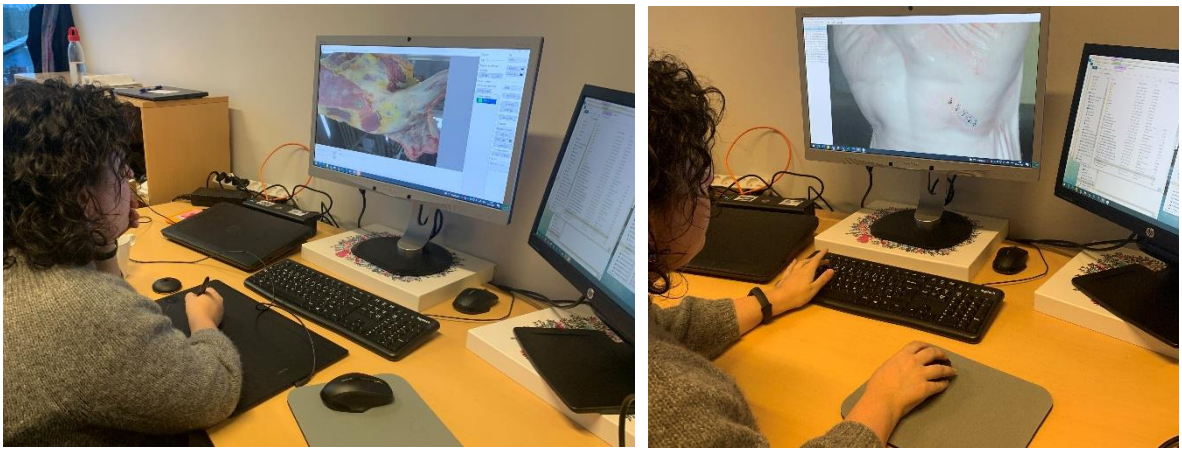
1. Man kan source annoteringsopgaven til et lavtlønsområde.
2. Man kan forsøge at opnå, at der annoteres frivilligt/gratis.

2.1.1. Sourcing

Erfaringer med at source annoteringsopgaven er blandede. Da kvaliteten af annoteringen som regel er vigtig, kan kommunikationsvanskeligheder og misforståelser i sidste ende være dyre og potentielt udgøre en større omkostning i form af kvalitetsforringelser og deraf følgende fejl og tidsforsinkelser end besparelsen i timelønnen. Alt andet lige er misforståelser hyppigere, når der er fysisk, kulturel og sproglig afstand mellem opgavestillerne og dem, der udfører annoteringen.

Dertil er det tilfældet, at når referencedata (dvs. billeddata, hvor man ved, hvad der er på billedet, og eventuelt hvor det er på billedet) granskes i forbindelse med annoteringen, opstår der ofte bevidsthed om mulige fejl og faldgruber, fx i form af bias i referencedata. Som et eksempel på sidstnævnte kan nævnes et tilfælde, hvor et netværk var trænet til at finde en bestemt hunderace på billeder. Da billederne med hunderacen tilfældigvis alle havde et datomærke på billedet, blev netværket trænet i at genkende billeder med datomærkning. En lignende situation opstod for DMRI, da vi ville træne et netværk til at finde fejl på brystflæsk. Da fejlene i træningsdata fortrinsvis var på venstresider, blev netværket i realiteten trænet til at finde venstresider. Når der er afstand mellem opgavestillere og dem, der annoterer, bliver chancerne for hurtigt at opdage sådanne fejl meget mindre, da dialogen som regel er mere omstændelig.

God annotering kræver omtanke og omhyggelighed. Det er omstændeligt og gentaget arbejde, der fordrer disciplin for at holde kvaliteten. Det er også en fordel med gode IT-hjælpemidler, fx digitale penne, gode computerskærme og i øvrigt gode arbejdsforhold, herunder lysforhold. I vores arbejde med *Deep Learning* hælder vi derfor til at udføre annoteringen in-house, hvor vi kan have en god tæt og løbende dialog og opfølgning. Vi oplever ofte, at dygtige studentermedhjælpere opdager både fejlkilder og muligheder i data.



Figur 3. Manuelle annoteringer med digital pen, mus og tastatur.

2.1.2. Gratis annotering

Vi har alle oplevet at stille vores arbejdskraft gratis til rådighed til annotering, når vi under besøg på en hjemmeside har skullet bevise, at vi ikke er en robot: Når du fx skal vælge billeder med trafiklys, er du måske i gang med at lave referencedata til *Deep Learning*-algoritmer til selvkørende biler.

En anden mere eller mindre oplagt måde at få gratis annotering på er gennem såkaldt "*gamification*", altså ved at gøre det sjovt at annotere. Se fx Gamification of Annotations — GIRL CRUSH CODE.

Vi har ikke selv eksperimenteret med "gratis" annotering på DMRI.

2.2. Iterativ gentræning

Et *Deep Learning*-netværk bliver typisk gradvist bedre til sin opgave, når det trænes med flere billeder, der afspejler den variation, som netværket udsættes for, når det kommer i drift. Slutapplikationen kræver ofte en høj nøjagtighed. Fx skal et netværk til veterinær kontrol ikke overse den relevante bemærkning, men det må på den anden side ikke for ofte "finde" fejl, der ikke findes i virkeligheden og dermed give anledning til unødvendig håndtering. Dette ville både kunne have negative økonomiske konsekvenser og kunne give anledning til en "ulven kommer-effekt", så sande positive ignoreres. Dvs. at raten af falske negative og falske positive skal være lave samtidigt, hvilket typisk er en udfordring og stiller store krav til slutapplikationen. Men selvom et netværk er trænet med et utilstrækkeligt antal annoterede billeder til at nå den nødvendige slutpræcision, vil det alligevel kunne generere et output, som i mange tilfælde er korrekt, eller tæt på at være korrekt. Dette kan være meget nyttigt og udnyttes til at effektivisere annoteringsarbejdet. Ved at træne netværket med få billeder, som er manuelt annote-

rede fra bunden, kan output ofte relativt hurtigt verificeres, og eventuelt korrigeres, og derefter bruges til at gentræne netværket. Denne proces kan gentages, så netværket opnår en stigende præcision og en øget evne til at håndtere varians. Værktøjer, der kan lette gennemsyn af billeder og eventuelt kan lette editeringen, når annoteringen skal rettes, kan med fordel bruges. Fx kan gennemsynet af billeder guides af parametre, der indikerer netværkets sikkerhed i sin vurdering (fx softmax funktionsværdien, som ofte bruges i klassificeringsnetværk), så billeder med lille sikkerhed enten pilles ud eller tilrettes manuelt. Den manuelle tilretning kan fx understøttes af andre *machine learning*-teknikker, så det går hurtigere.

2.3. Annotering af sjældne begivenheder

I nogle tilfælde skal et netværk trænes til at klassificere kategorier, der er sjældent forekommende. Dette kunne fx være et netværk, der skal overvåge produkter for sjældne fejl. I praksis indebærer dette, at der skal optages mange billeder, for kun en meget lille brøkdel af dem vil repræsentere den sjældne kategori. Derfor skal der gennemses rigtigt mange billeder for at finde nok (fx med fejlen) til at træne netværket til at genkende den sjældne kategori. Hvis det ovenikøbet er vanskeligt at finde fejlen på billedet, fx hvis der er tale om små forureninger på en stor heterogen overflade, kan det blive en næsten uoverstigelig opgave at finde billeder nok til selv at starte den iterative proces beskrevet foroven. I disse tilfælde kan to strategier overvejes:

1. Er det muligt at syntetisere (lave kunstige) billeder med den sjældne kategori/fejl. Enten ved at skabe kategorien/fejlen i virkeligheden og tage billeder af den, eller ved at lave billedmanipulationer så kategorien/fejlen skabes virtuelt på billeder?
2. Kan der benyttes en anden *machine learning*-teknik (fx *boosted decision trees, random forest, nearest neighbor, t_SNE*), der kan bruges til at analysere et stort datasæt? Teknikken behøver ikke have stor præcision, men skal kunne finde eksempler på kategorien (sande positive), uden alt for mange falske positive.

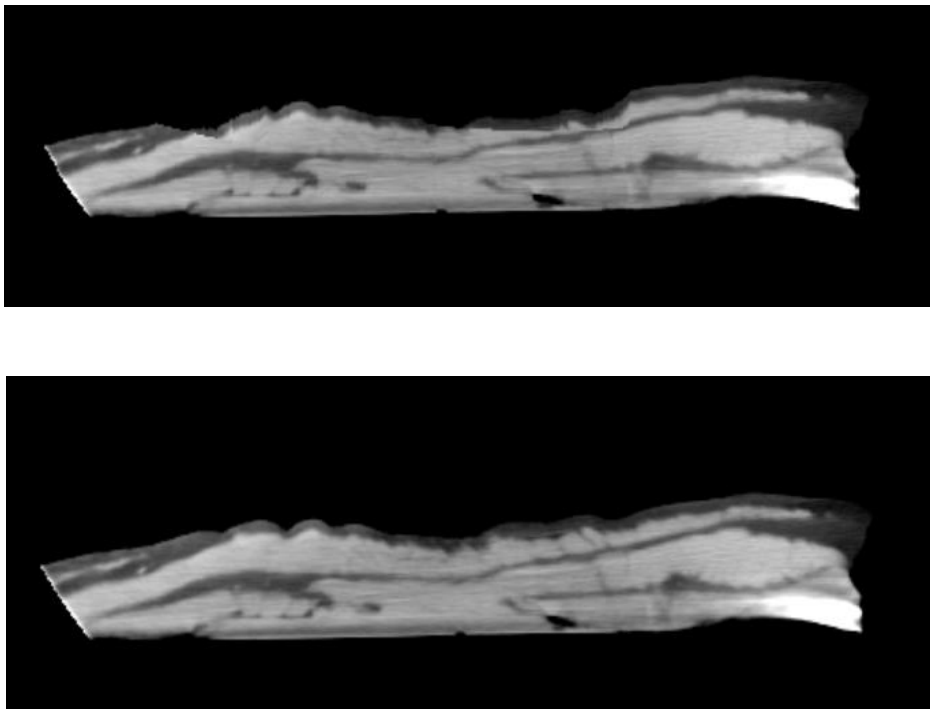
De to strategier kan eventuelt kombineres. Fx i tilfælde af sjælden forekomst af en forurening kan forureningsmaterialet påføres et emne, billeder tages, der laves en algoritme baseret på en *random forrest*-analyse af pixelværdierne, der, i en vis udstrækning, kan skelne mellem forurening og baggrund. Algoritmen benyttes på et stor datasæt, og output gennemgås til at finde billeder, der kan bruges til at træne et *Deep Learning*-netværk.

En sidste mulighed for annotering af sjældne kategorier kunne være brug af såkaldt *unsupervised learning*. Tanken er at bruge netværksteknikker til at lave en automatisk genereret kategorisering i grupper med den forventning, at den sjældne fejlkategori afviger tilstrækkeligt fra de normale kategorier, til at den kan udpeges af netværket på baggrund af de parametre, som netværket har udviklet til at lave sin kategorisering. Hos DMRI har vi forsøgt at gøre dette, både ved hjælp af t-SNE-analyse til dimensionsreducering for at finde afvigelser, og vha. semi-unsupervised auto-encodere til anomali detektion. Dog

uden succes, da de anomalier, vi har forsøgt at finde, ikke kan skelnes "tydeligt" fra det perfekte tilfælde. Der findes dog eksempler i litteraturen med heldigere udfald (se fx Unsupervised Clustering Methods for Identifying Rare Events in Anomaly Detection (waset.org)).

2.4. Dyre referencebilleder

Et andet problemområde er, at det i visse tilfælde er dyrt at tilvejebringe de referencebilleder, der skal benyttes til at træne et netværk. Hos DMRI har vi fx arbejdet med at prøve at bruge autoencoder-netværk til at repræsentere anatomiens underliggende struktur til at kunne prædiktere subkutane fedtlag på delstykker af gris vha. forskellige røntgenprojektioner. I dette tilfælde er inputbillederne til netværket røntgenbilleder, mens output er billeder, der beskriver fedtlagets tykkelse. Referencen er dermed billeder af den korrekte fedttykkelse, og disse kan tilvejebringes vha. CT-billeder. Problemet her er ikke arbejdet med selve annoteringen af fedtlaget på CT-billederne – disse kan opnås vha. automatiske analysealgoritmer, som vi tidligere har udviklet – men at det er dyrt at CT-scanne tilstrækkeligt mange delstykker til at opnå referencebilleder nok til at træne et netværk. Vi har derfor arbejdet med at syntetisere mere data ud fra relativt få CT-billeder. Vi har dels syntetiseret billeder af røntgenprojektioner ud fra CT-billederne, og vi har lavet syntetiske modifikationer af CT-billederne til kunstigt at øge antallet at CT-billeder (se figur 4).



Figur 4. Eksempel på syntetiseret billeddata. Foroven er det oprindelige tværsnitsbillede af et stykke brystflæsk. Forneden er der ud fra det øverste billede syntetiseret et nyt billede, som bruges til at forøge mængden af (dyre) CT-billeder.

En problemstilling her er, om den underliggende anatomiske struktur, som auto-encoder netværket skal fange, går tabt i syntetiseringen. For at undgå dette kan et GAN-netværk eller et andet auto-encoder-netværk eventuelt benyttes til billedsyntesen (se fx [2111.03388] A *Deep Learning* Generative Model Approach for Image Synthesis of Plant Leaves (arxiv.org)).

I øvrigt er teknikker til at øge variationen i træningsdata ved at tage annoterede billeder og lave ændringer (fx farvejusteringer, spejling, rotation, tilføjelse af støj) velkendte og kan reducere omkostningerne ved tilvejebringelse af annoterede referencebilleder i mange situationer.

2.5. Genbrug af referencedata

Den sidste metode, som vi har undersøgt på DMRI, er brug af billeddata fra en kontekst i en anden kontekst. Et hypotetisk eksempel kunne være brug af annoterede billeder fra en eksisterende database med forskellige dyr til at træne et netværk til at finde billeder med grise ved modtagelsen på et slagteri (uden annoterede billeder på slagteriet). Teknikken kaldes *domain adaptation* og er fx relevant, når det er dyrt, svært eller ligefrem umuligt at tilvejebringe et referencedatasæt til den specifikke anvendelse, men til gengæld muligt at få et referencedatasæt i en anden men lignende situation. En særlig relevant anvendelse af domain adaptation er, når en *Deep Learning*-applikation, udviklet på ét sted med de gældende lokale baggrunds- og lysforhold samt kameravinkler, skal implementeres et nyt sted, hvor forholdene er lidt anderledes. Det er tænkeligt, at domain adaptation kan reducere omkostningerne for genimplementering ved at nedsætte behovet for nye annoteringer og gentræning.

3. Sammenfatning

Deep Learning er en meget effektiv teknologi med stor relevans for kødindustrien, da den faciliterer billedanalyse af emner med stor (biologisk) varians. For at reducere omkostningerne forbundet med udvikling og implementering af løsninger baseret på *Deep Learning* er der behov for metoder, der effektiviserer annoteringen af referencedatasæt. I projektet "Nye målemetoder til kødindustrien" er der arbejdet med at tilegne sig og rapportere metoder, der effektiviserer annoteringen og dermed øger anvendeligheden af teknologien i kødindustrien. Det er håbet, at nærværende rapport kan informere og inspirere, når der udvikles nye *Deep Learning*-baserede løsninger til kødindustrien.

TEKNOLOGISK
INSTITUT