

Adviesrapport

Onderzoek naar methoden voor data inwinning

– Ministerie van Infrastructuur en Waterstaat

HAS Hogeschool
Onderwijsboulevard 221
Postbus 90108
5200 MA 's-Hertogenbosch
Telefoon: 088 890 36 00

Documenttitel: Onderzoek naar methoden voor data inwinning
Projectcode: 21200440

Opdrachtgever: Ministerie van Infrastructuur en Waterstaat
Contactpersoon: Robby Veders

Projectleider: Marien de Bakker

Projectteam: Niek Broekhoven
Daniël Poot

Plaats: 's-Hertogenbosch
Datum: 24 juni 2022

Voorwoord

Voor u ligt het adviesrapport 'Onderzoek naar methoden voor data inwinning'. Het adviesrapport is opgesteld ten behoeve van een onderzoek voor het ministerie van Infrastructuur en Waterstaat. Het advies is opgesteld door twee vierdejaars studenten Applied Geo Information Science aan de HAS Hogeschool te 's-Hertogenbosch. Het onderzoek beslaat netto 20 weken en vindt plaats van 07-02-2022 tot en met 08-07-22. Het onderzoek wordt geleid door Marien de Bakker en heeft als expert Maurits Dorlandt toegewezen.

Tijdens het onderzoek hebben we hulp gekregen van verschillende experts op hun eigen vakgebied. Graag willen we hen bedanken. In het bijzonder dank aan Marien de Bakker en Maurits Dorlandt voor de goede en prettige begeleiding tijdens het project.

Daniël Poot en Niek Broekhoven
's-Hertogenbosch, 24 juni 2022

Samenvatting

Voor het register externe veiligheidsrisico's dient er nieuwe data en bestaande data aangevuld te worden, zodat het register gevuld wordt. Vanuit deze stelling wordt de best passende methode voor data inventarisatie gezocht door middel van evaluatie per criterium op de methode. De data in deze casus betreft propaantanks door Nederland verspreid. Deze inventarisatie dient beschreven te worden aan de hand van de voor- en nadelen van de methoden, eventueel onderbouwd aan de hand van pilot uitvoeringen.

Het doel van het project is om de data voor het register externe veiligheidsrisico's zo efficiënt mogelijk op te halen en hierbij een reproduceerbaar proces neer te zetten. Dit proces is een pilot en moet op kunnen worden geschaald naar heel Nederland. Hierbij is het van belang dat de data van kwaliteit is, zodat het kan worden gepubliceerd op Atlas Leefomgeving. Hieruit volgt de hoofdvraag van het onderzoek:

Welke methode is efficiënt, betrouwbaar en realiseerbaar als mogelijkheid om slim data in te winnen van datasets die nuttig zijn voor het register externe veiligheidsrisico's?

Om dit te beantwoorden zijn de volgende methoden onderzocht:

Fysiek data inwinnen is een vorm van data inwinnen waarbij men naar buiten gaat om te zien waar propaantanks staan en deze propaantanks te inventariseren. Tijdens het data inwinnen met voorbereiding is er eerst gekeken naar een gebied waar mogelijk propaantanks aanwezig zijn. Hier zijn 18 propaantanks in een uur geïnventariseerd.

Machine learning (onderverdeeld in classificatie en object herkenning) is een vorm van kunstmatige intelligentie waarbij een computer getraind wordt aan de hand van datasets om bijvoorbeeld objecten te herkennen. Tijdens het proces van classificatie herkenning voor een machine learning model is de methode dieper onderzocht door de licht spectrum banden rood, groen en blauw (RGB) te onderzoeken. Tijdens het proces van herkenningstechnieken met machine learning worden objecten herkend door het model te trainen aan de hand van training objecten. In een gebied van 16 km² zijn 2 van de 12 propaantanks herkend.

Crowd sourcing is een vorm van het verzamelen van informatie of kunde bij een grote groep mensen. Hier staat de inzet van de mens centraal. Om de haalbaarheid van deze methode te achterhalen zijn meerdere vragenlijsten opgesteld die onderzochten hoe de burger of student tegenover participatie staat. De bezitters van propaantanks willen in de meeste gevallen niet meewerken. Studenten staan positief tegenover deelname aan een mapathon.

Data opvragen bij derden is als alternatief op verschillende data inwin methodieken ook een mogelijkheid om de gasleveranciers van propaan zelf te benaderen voor een locatie dataset van hun afnemers. Hiervoor is door gebrek aan medewerking van de partijen geen resultaat geboekt.

Om deze methoden tegen elkaar af te wegen is een multi criteria analyse opgesteld op basis van drie criteria, waaruit resulteert dat een mapathon en het fysiek inwinnen met voorbereiding het best aansluiten op de wens en urgentie op basis van de criteria. Een kanttekening hierbij is dat de realiseerbaarheid nog niet is meegenomen, waardoor de toevoeging van een machine learning model op de lange termijn toch belangrijk is, aangezien dit voor het waarnemen van veranderingen in de toekomst een lucratievere methode is.

Om in het geval van de mapathon en data opvragen bij derden een beter resultaat te boeken had kunnen worden onderzocht hoeveel resultaat op een mapathon wordt geboekt door het organiseren van een mapathon.

Daarnaast had kunnen worden onderzocht hoe een leverancier tegenover een beloning staat als data wordt gedeeld.

Dit alles leidt tot een advies, waarin naar voren komt dat een onderzoek naar aanwezigheid van data over propaantanks bij omgevingsdiensten aan de basis ligt. Door middel van 150 á 250 samples kan het ministerie d.m.v. uitbesteding een machine learning model laten opzetten. Hierna kan het model meermaals worden uitgevoerd om veranderingen op te merken en te laten valideren door mapathons op te zetten bij de genoemde studenten. Hiernaast wordt aanbevolen vervolgonderzoek te doen naar overige objecten als gas-meet en

regelstations. Mocht dit door machine learning te herkennen zijn, kan net als voor windturbines in grote lijnen de methode uit dit rapport worden gevolgd.

Inhoud

1. Inleiding	8
1.1. Achtergrond	8
1.2. Probleemstelling	9
1.3. Doel/hoofdvraag	9
1.4. Randvoorwaarden	10
1.5. Leeswijzer	11
2. Methoden & Materiaal	12
2.1. Datacollectie	12
2.2. Datakwaliteit	12
2.3. Software	12
2.4. Fysiek data inwinnen	13
2.4.1. Met voorbereiding	14
2.4.2. Zonder voorbereiding	15
2.5. Machine Learning	16
2.5.1. Remote sensing met machine learning (classificatie)	16
2.5.2. Herkenningstechniek met machine learning (object herkenning)	17
2.6. Crowd sourcing	18
2.6.1. Literatuur onderzoek	18
2.6.2. Enquête onderzoek	18
2.7. Data opvragen bij derden	20
2.7.1. Gasleveranciers	20
2.8. Multi Criteria Analyse	21
3. Resultaten	22
3.1. Resultaten fysiek data inwinnen	22
3.1.1. Fysiek data inwinnen met voorbereiding	22
3.1.2. Resultaat	23
3.1.3. Fysiek data inwinnen zonder voorbereiding	23
3.1.4. Resultaat	24
3.2. Machine learning	25
3.2.1. Classificatie met deep learning	25
3.2.2. Object herkenning met deep learning	25
3.3. Crowd sourcing	27

3.4. Multi Criteria Analyse	28
3.5. Vergelijkingsmatrix	30
3.6. Beantwoorden van deelvragen.....	36
4. Discussie	38
4.1. Methoden	38
4.1.1. Fysiek data inwinnen.....	38
4.1.2. Remote sensing met machine learning (classificatie).....	38
4.1.3. Herkenningstechniek met machine learning (object herkenning).....	38
4.1.4. Crowd sourcing.....	39
4.1.5. Data opvragen bij derden.....	39
4.2. Combinaties discussie.....	40
4.2.1. Machine learning en mapathon	40
4.2.2. Machine learning met fysiek inwinnen.....	40
5. Conclusie.....	41
6. Advies en aanbevelingen.....	42
6.1. Advies	42
6.2. Aanbevelingen.....	43
6.2.1. Windmolens.....	43
6.2.2. Gas meet- en regelstations	43
7. Literatuur	44
8. Bijlagen.....	47
8.1. Bijlage datacollectie	47
8.2. Bijlage datakwaliteit	49
8.3. Bijlage Deep learning proces.....	51
8.4. Bijlage Deep learning analyse beschrijving	53
8.4.1. Stappen voor het gebruik van deep learning	54
8.4.2. Deep learning classificatie analyse beschrijving.....	54
8.4.3. Analyseren met een spectraal profiel (analyse beschrijving proces).....	63
8.4.4. Deep learning object detection analyse beschrijving	65
8.5. Bijlage Bereidheid melden propaantank enquête	72
8.6. Bijlage animo meting mapathon	75
8.7. Berekening Fysiek data inwinnen	78
8.8. Bijlage Multi Criteria analyse beschrijving.....	80
8.9. Mapathon	82

Terminologie

In dit rapportage worden een aantal termen gebruikt die afgekort zijn of onder het vakjargon vallen. Deze termen worden hieronder uitgelegd:

CPU = Processor (Central processing unit)

Deep learning = onderdeel van machine learning, een gelaagd neurale netwerk leert van grote hoeveelheden data.

GPU = Grafische kaart (graphics processing unit)

Ground truthing = Op locatie ingewonnen data / waarnemingen

Machine learning = Kunstmatige intelligentie, het trainen van een computer om van verwerkte data te leren.

REV = Register externe veiligheidsrisico's

RRGS = Register Risicosituaties gevaarlijke stoffen

VRAM = Video Random Access Memory (intern geheugen grafische kaart)

1. Inleiding

1.1. Achtergrond

Het Register Externe Veiligheidsrisico's (REV) is een database van het Ministerie van Infrastructuur en Waterstaat. De database bevat data over de Externe Veiligheidsrisico's. Het REV is een nieuw register dat opgesteld wordt om veiligheidsrisico's met de bijbehorende data in kaart te brengen en is een vervanging van het al bestaande RRGs, Register risico situaties gevaarlijke stoffen. Hierbij gaat het om risico's rondom opslag, transport en gebruik van gevaarlijke stoffen. In de Atlas Leefomgeving wordt deze informatie zichtbaar gemaakt via kaartlagen voor alle burgers. Hierop kan gezien worden waar zich risicovolle activiteiten plaatsvinden en wat dat betekent voor hun leefomgeving, zoals het hebben van een propaantank in de buurt.

Bovengrondse propaantanks hebben een aantal gemeenschappelijke kenmerken. Ze zijn namelijk altijd mintgroen of wit en hebben een inhoud van 1000, 1750, 2700 of 5000 liter. Voorbeelden hiervan zijn te vinden in Figuur 1.



Figuur 1: Voorbeeld propaantanks

Het bij- en up to date houden van deze database kost relatief veel tijd en energie. Volledigheid en betrouwbaarheid van de data maken het mogelijk goed inzicht te bieden op de veiligheidsrisico's binnen Nederland. De duurzaamheid van het register heeft vooral te maken met het slim inwinnen van data*, waarbij de kwaliteit van de informatie gegarandeerd wordt. Transparantie is hierbij voor de overheid belangrijk. Om te garanderen dat de gevisualiseerde informatie betrouwbaar en actueel is, wordt deze op één plek verzameld. Het REV stelt deze data beschikbaar, zodat iedereen over dezelfde actuele informatie beschikt.

*Onder het slim inwinnen van data wordt verstaan dat de data wordt ingewonnen op een minder arbeidsintensieve manier ten opzichte van de huidige situatie. De uitdaging ligt bij het gebruiken en evalueren van nieuwe technieken. Hierbij is belangrijk dat de betrouwbaarheid van de data en de efficiëntie van het inwinnen centraal staan.

Het REV is belangrijk om risico's omtrent veiligheid te tonen bij een noodsituatie, zoals brand of vrijkomen van giftige stoffen. Bij deze risico's/activiteiten horen aandachtsgebieden. Door middel van deze aandachtsgebieden wordt het voor iedereen mogelijk om rekening te houden met eventuele gevaren binnen een gebied. Dit speelt een grote rol bij de afweging om bepaalde activiteiten binnen een gebied wel of niet toe te staan.

1.2. Probleemstelling

Het REV staat klaar om informatie te ontvangen en te ontsluiten. Momenteel zit er echter nog nauwelijks of geen gevalideerde data in het REV. Dit heeft te maken met de tijd die het zal kosten om de bronsystemen om te zetten naar de nieuwe informatieafspraken van de omgevingswet.

Naast het aanvullen van het REV kost het samenvoegen van de verschillende datasets en up to date houden van het REV ook veel tijd. Daarom is het essentieel om onderzoek te doen naar verschillende “slimme inwin”* methoden om de datasets van het REV te verrijken met gevalideerde data.

In december 2021 bevatte de landelijke signaleringskaart externe veiligheidsrisico's (EV, 2022) in totaal 4467 propaantanks. Door de heer Boudewijn de Hoop (omgevingsdienst Rivierenland) is hierop een schatting gemaakt dat dit ongeveer 25% van de aanwezige propaantanks is in Nederlands. De overige 75% resulteert in een schatting van 13.401 niet in het RRGs opgenomen propaantanks.

Voor het register externe veiligheidsrisico's dient er nieuwe data en bestaande data aangevuld te worden, zodat het register volledig wordt. Het aanvullen van de oude informatie en nieuwe datasets creëren voor het verrijken van de data is erg tijdrovend. Om het register aan te vullen dienen er verschillende nieuwe datasets van externe veiligheidsrisico's gecreëerd te worden, zoals propaantanks, gas meet- en regelstations en windturbines. Bij alle drie de datasets zijn er zowel objecten met vergunningsplicht en zonder vergunningsplicht. Vooral de objecten zonder vergunningsplicht zijn slecht of niet gedocumenteerd en lastig in kaart te brengen.

1.3. Doel/hoofdvraag

De hoofdvraag bij deze casus luidt: “Welke methode is efficiënt, betrouwbaar en realiseerbaar als mogelijkheid om slim data in te winnen van datasets die nuttig zijn voor het Register Externe Veiligheidsrisico's?”

Deze hoofdvraag is onderverdeeld in deelvragen:

1. Wat zijn kenmerken van propaantanks die vastgelegd dienen te worden in de dataset van propaantanks
2. Welke methoden zijn er voor het slim inwinnen van data?
3. Welke methode is passend voor het inwinnen van data voor de propaantank dataset?
4. Welke methode is haalbaar voor het slim inwinnen van propaantank data voor heel Nederland?
5. Welke methodes vallen onder de mogelijkheid om slim data in te winnen voor overige datasets die nuttig zijn voor het register externe veiligheidsrisico's?
6. Hoe worden de datasets gevalideerd?

Het doel van het project is om de data van propaantanks voor het REV zo efficiënt mogelijk op te halen en hierbij een reproduceerbaar proces neer te zetten. Dit proces is een pilot en moet op kunnen worden geschaald naar heel Nederland. Hierbij is het van belang dat de data van kwaliteit is, zodat het kan worden gepubliceerd op Atlas Leefomgeving. Het betreft hierbij data herkenning van propaantanks.

1.4. Randvoorwaarden

Voor dit project ligt de focus op de data van propaantanks. Voor het project dient de methodiek geprioriteerd te worden aan de hand van het uitvoeren van een pilot om de methode te valideren. Op basis hiervan wordt er naar de andere datasets gekeken en beschreven of dezelfde techniek wel of niet een mogelijke oplossing is voor het inwinnen van data. Deze inventarisatie dient beschreven te worden aan de hand van de voor- en nadelen van de methoden, eventueel onderbouwd aan de hand van pilot uitvoeringen. Concluderend dient hier een advies voor het ministerie van Infrastructuur en Waterstaat uit te komen.

Het projectgebied bij dit onderzoek beslaat het Nederlands grondgebied, omdat dit de nationale afbakening voor het Ministerie van Infrastructuur en Waterstaat is. Ook is het uitgangspunt van dit onderzoek en eventueel vervolgonderzoeken om de risico objecten voor geheel Nederland in te winnen.

Het projectgebied voor machine learning wordt afgebakend naar de regio van omgevingsdienst midden-West Brabant **Fout! Verwijzingsbron niet gevonden.** De keuze voor regio midden-West Brabant is gedaan, omdat er op deze manier ground-truthing data beschikbaar is om de resultaten van machine learning te toetsen.



Figuur 2 Projectgebied en deelgebied

Voor dit project ligt de focus op de data van propaantanks. Voor het project dient de methodiek geprioriteerd te worden aan de hand van het uitvoeren van een pilot om de methodiek te valideren. Op basis hiervan wordt er naar de andere risico objecten gekeken en beschreven of dezelfde techniek wel of niet een mogelijke oplossing is voor het inwinnen van data. Deze inventarisatie dient beschreven te worden aan de hand van de voor- en nadelen van de methoden, eventueel onderbouwd aan de hand van pilot uitvoeringen.

Het projectgebied bij dit onderzoek beslaat het Nederlands grondgebied, omdat dit de nationale afbakening voor het ministerie van infrastructuur en waterstaat is. Ook is het uitgangspunt van dit onderzoek en eventueel vervolgonderzoeken om de risico objecten voor geheel Nederland in te winnen.

Het projectgebied voor de methode machine learning wordt afgebakend naar de regio van omgevingsdienst midden-West Brabant Figuur 2 Projectgebied en deelgebied. De keuze voor regio midden-West Brabant is gedaan, omdat er gevalideerde waarnemingen van propaantanks beschikbaar zijn om de resultaten van machine learning te toetsen.

1.5. Leeswijzer

In hoofdstuk 1 wordt beschreven wat de casus van het project is. Hoofdstuk 2 introduceert de hoofdvraag en deelvragen, waarbij het doel van het project wordt gedefinieerd. In hoofdstuk 3 wordt na een korte inleiding over datakwaliteit toegelicht hoe elke methode is uitgevoerd met eventuele referenties naar bijlagen. Na het behandelen van het fysiek inwinnen, machine learning en crowd sourcing binnen dit hoofdstuk, wordt een vervolg gegeven in hoofdstuk 4. Dit hoofdstuk presenteert van elke methode de resultaten uit het onderzoek, waarna een multi criteria en een vergelijkingsmatrix het hoofdstuk afsluiten en een inzicht bieden in de combinatie van resultaten. Hoofdstuk 5 geeft een toelichting op de resultaten en vervolgt met het beantwoorden van de opgestelde deelvragen. Hoofdstuk 6 richt de pijlen op het concluderen van de resultaten door middel van een antwoord op de hoofdvraag. Hierna volgt het advies in hoofdstuk 7. Dit hoofdstuk verbindt met de rest van het verslag en geeft de lezer een inzicht hoe een vervolg te geven aan het inventariseren van propaantanks door Nederland. In hoofdstuk 8 wordt de literatuur beschreven, waarna wordt afgesloten met de bijlagen in hoofdstuk 9.

Verwijzingen

In dit rapportage worden verwijzingen naar figuren en bijlage gedaan door middel van kruisverwijzingen. Door op de verwijzingen te klikken terwijl de knop "ctrl" ingedrukt word, word de lezer direct naar het juiste Figuur of bijlage verwezen.

2. Methoden & Materiaal

Tijdens het onderzoek zijn er verschillende methoden onderzocht om slim data in te winnen. Deze methoden worden in dit hoofdstuk kort en gestructureerd uitgelegd. Hier wordt beschreven wat er gebeurd is om een methode te onderzoeken. De methoden die uitgelegd worden zijn: Fysiek data inwinnen, Deep learning, Crowd sourcing en data opvragen bij derden. Deze methoden worden gekeurd aan de hand van drie criteria; betrouwbaarheid, kosten en duur in een multi criteria analyse.

2.1. Datacollectie

Om een juiste analyse uit te voeren en een deep learning model te kunnen maken wordt er gebruik gemaakt van verschillende databronnen. Onder andere de 8cm luchtfoto's van beeldmateriaal is aangeleverd door PDOK. Ook is er een dataset aangeleverd met de propaantank inventarisatie van omgevingsdienst Midden West-Brabant. Deze data is beschreven aan de hand van de metadata en eventueel de structuur van de data tabellen. De beschrijvingen van de data zijn te vinden in Bijlage datacollectie.

2.2. Datakwaliteit

De data die gebruikt worden, dienen gekeurd te worden op verschillende datakwaliteit kenmerken om te bepalen of de data bruikbaar en van goede kwaliteit is. De data worden gekeurd op data kwaliteit kenmerken die van toegevoegde waarde zijn voor het gebruiksdoel van de data. De stappen in het datakwaliteit proces zijn beschreven aan de hand van tabellen met de gebruikte attributen en functie. Eventuele opmerkingen voor meer context zijn toegevoegd en een korte keuring van het datakwaliteit kenmerk. De datakwaliteit onderzoeken zijn beschreven in Bijlage datakwaliteit.

2.3. Software

Tijdens het project is gebruik gemaakt van verschillende software. Per methodiek staat beschreven van welke software gebruik is gemaakt en of er eventuele open source (gratis te gebruiken) alternatieven zijn.

2.4. Fysiek data inwinnen

Fysiek data inwinnen is een vorm van data inwinnen waarbij men naar buiten gaat om te zien waar propaantanks staan en deze propaantanks te inventariseren.

Voor de methode fysiek data inwinnen is er onderscheid gemaakt tussen twee methoden;

- met voorbereiding (filtering van route op buitengebieden en route uitgezocht voor het rondrijden)
- zonder voorbereiding (tijdens het monitoren gekozen voor bepaalde routes in verwachting van propaantanks)

Dit is gedaan om te toetsen of filtering van gebieden en geen filtering van gebieden een significant verschil geeft op de uitkomsten. Het filteren van deze gebieden is uitgevoerd door de "standplaats" (kaartlaag) uit de BAG dataset te filteren op de vlakken die geen overlap hebben met de Enexis gasleiding data. Dit zijn de percelen die niet op het gasnet liggen aangesloten.

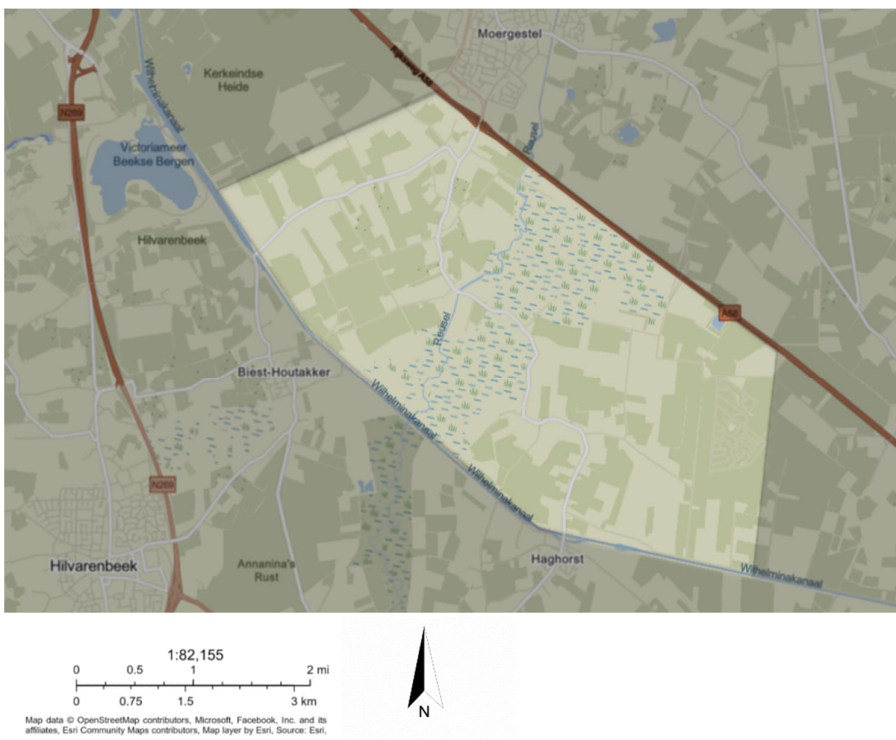
Om een idee te krijgen van het fysiek inwinnen van data is er gekozen om een middag fysiek data inwinnen als pilot uit te voeren. Hierdoor kunnen er resultaten geboekt worden waar een berekening op gedaan wordt. Door middel van rondrijden door buitengebieden en het gebruik van de ESRI field maps applicatie om locatie data van propaantanks in te winnen in een digitale omgeving. Esri field maps is een applicatie binnen het Esri platform waarbij digitale kaart omgevingen en datasets opgeslagen kunnen worden voor gebruik op een telefoon applicatie. Hierbij kan er gemakkelijk door middel van de telefoon in het veld data ingewonnen worden. Deze applicatie valt binnen de Esri licentie en is een betaalde service.

Eventuele andere alternatieven voor opensource zijn Qfield en Trimble terraflex, waarbij Trimble terraflex een gelimiteerde open source service heeft.

Voor beide methoden is er één uur lang data ingewonnen. Hierdoor is er een idee ontwikkeld van de kosten, manuren en realiseerbaarheid van deze methoden.

2.4.1. Met voorbereiding

Tijdens het data inwinnen met voorbereiding is er eerst gekeken naar een gebied waar mogelijk propaantanks aanwezig zijn. Dit is gedaan door te kijken welke gebieden niet aangesloten zitten op het gasnet van Enexis in West/midden Brabant. Het geïnventariseerde gebied is het buitengebied van Biest-Houtakker, Moergestel en Haghorst zichtbaar op Figuur 3 Studiegebied Biest-Houtakker - Fysiek data inwinnen. Dit gebied bevindt zich tussen Tilburg, Boxtel en Oirschot.



Figuur 3 Studiegebied Biest-Houtakker - Fysiek data inwinnen

2.4..2. Zonder voorbereiding

Ook is er gekeken naar het inventariseren van propaantanks door het buitengebied in te rijden zonder directe planning. Hierbij is dus niet naar het netwerk van Enexis met gasleidingen gekeken om specifieke gebieden uit te filteren.

Tijdens het monitoren is er gebruik gemaakt van dezelfde materialen en software.

Het monitoren is gedaan tussen Udenhout en de Loonse en Drunense duinen, omdat het een afgelegen buitengebied is met veel boerderijen. De kaart in Figuur 4 Studiegebied Udenhout/Loonse Drunense Duinen - Fysiek data inwinnen visualiseert de in kaart gebrachte propaantanks en het gebied waar het monitoren is uitgevoerd.



Figuur 4 Studiegebied Udenhout/Loonse Drunense Duinen - Fysiek data inwinnen

2.5. Machine Learning

Machine learning is een vorm van kunstmatige intelligentie waarbij een computer getraind wordt aan de hand van datasets om bijvoorbeeld objecten te herkennen.

Om propaantanks te herkennen zijn er twee verschillende machine learning methodieken die gebruikt kunnen worden. De methodieken die uitgevoerd worden als een prototype hierbij zijn:

1. "object classification with deep learning (Esri, 2022) Deep learning classificatie analyse beschrijving.
2. "object recognition with deep learning (Esri, z.d.-b)) Deep learning object detection analyse beschrijving

Beide methodieken zijn uitgevoerd door middel van ArcGIS Pro 2.9 ((Esri, 2021)) met een geïnstalleerd package "Deep learning essentials" versie 2.8 (Esri, z.d.-a)).

Bij het gebruik van Machine learning en deep learning wordt er een herkenningstechniek toegepast door een computer, op de ingevoerde rasterdata (rasterdata van vlakdekkend Nederland). Dit is een proces dat lang kan duren en daarom is het essentieel om de invoer data te beperken tot het gebied waar mogelijk propaantanks kunnen staan. Hierdoor kan de invoer data dus gefilterd worden op een aantal verschillende locaties waar geen propaantanks verwacht worden Figuur 10 Deep learning proces - Filtering gebied. Hierbij kan gedacht worden aan gebieden waar water is; meren, rivieren, beken etc. Ook zijn locaties zoals stedelijke gebieden en percelen die aangesloten zijn op het gasnet uit te sluiten. Deze gebieden kunnen zoals in Figuur 10 Deep learning proces - Filtering gebied worden gefilterd van de invoer data voor het Machine- of Deep learning model om rekentijd te besparen.

De resultaten per machine learning model zijn getoetst door de "summarize within" functie. Deze functie berekent het aantal polygonen (herkende propaantanks) die binnen een polygoon (ground truthing) vallen. Wanneer een van deze herkende propaantanks binnen een ground truthing polygoon vallen is dit een juist herkende propaantank. De overige zijn valse positieven en dienen verwijderd te worden.

2.5.1. Remote sensing met machine learning (classificatie)

Tijdens het proces van classificatie herkenning voor een machine learning model Deep learning classificatie analyse beschrijving is de methode dieper onderzocht door de licht spectrum banden rood, groen en blauw (RGB) te onderzoeken. Dit is getest door een spectraal profiel (Esri, z.d.-d) uit te zetten op de luchtfoto data, aangeraden door de heer Willem Vlot (W. Vlot, Esri AI specialist, Esri GIS-Tech Beurs, 12 april 2022). Een spectraal profiel visualiseert de waardes in de RGB banden die waargenomen worden voor verschillende objecten die gespecificeerd zijn Figuur 20 Spectraal profiel RGB. In het spectrale profiel is waar te nemen dat veel andere "wit" gekleurde objecten zoals; daken, busjes etc. overeenkomen met het spectrale profiel van de propaantanks. Hierdoor komen er dus ook veel "valse positieven" uit het gebruik van een machine learning op basis van object classificatie.

De classificatie is uitgevoerd door middel van de esri "image classification wizard" een methode die stapsgewijs het classificatie proces doorloopt (Esri, z.d.-e) . Hierbij is gebruik gemaakt van informatie die vergaard is tijdens de ESRI GIS-Tech beurs van de heer Willem Vlot en uit gesprekken met de heer Loek van Oijen, deep learning en AI specialist bij NEO. Ook is er gebruik gemaakt van verschillende informatieve YouTube video's en literatuur van de esri website (Bennett, 2018; Singh, 2020).

2.5.2. Herkenningstechniek met machine learning (object herkenning)

Tijdens het proces van herkenningstechnieken met machine learning, beschreven in Deep learning object detection analyse beschrijving, er een aantal stappen beschreven en beargumenteerd betreft het gebruik van parameters. Hierbij zijn verschillende keuzes beargumenteerd en gereflecteerd, maar is er ook advies geboden voor eventuele veranderingen in parameters (variabelen), door het gebruik van eventuele betere specificaties van een pc dan tijdens dit onderzoek beschikbaar zijn geweest. De specificaties die tijdens dit onderzoek gebruikt zijn waren:

- I7 processor
- Nvidia K1000M grafische kaart
- 32GB Ram geheugen

De vereisten voor het uitvoeren van de machine learning object recognition zijn met deze specificaties wel toereikend als er gebruik wordt gemaakt van de CPU, maar dit zorgt voor een groot gebruik van het RAM-geheugen waardoor alle 32GB ongeveer vrij dient te zijn. Het uitvoeren van het machine learning model is hierbij langzaam, ongeveer 8 uur voor een 2 bij 2 kilometer vak.

Het geheugen van de grafische kaart is daarentegen niet toereikend genoeg en dient minimaal 4GB VRAM te zijn en wordt er aangeraden te werken met 8GB VRAM (Esri, 2020). Deze specificaties zijn dus nodig bij het uitvoeren van een machine learning model object recognition bij het gebruik van de GPU.

Voor het maken van een deep learning model zijn er 125 training samples (training objecten) in kaart gebracht, waar er ongeveer 150 nodig zijn. Deze objecten zijn met de hand ingewonnen met als ground truthing (gevalideerde waarnemingen) referentie de dataset van omgevingsdienst midden- en west-Brabant. Na het intekenen van training samples zijn deze samples gelabeld en geëxporteerd tot plaatjes die gebruikt worden om het deep learning model te trainen. Na het exporteren is het deep learning model getraind en uitgevoerd (Esri, z.d.-d).

Tijdens het maken van het model is er gebruik gemaakt van verschillende literatuur van de Esri website, gesprekken met professionals op de Esri GIS-Tech beurs zoals Willem Vlot en het volgen van een deep learning workshop door de heer David Yu, data analyst en deep learning specialist bij ESRI.

2.6. Crowd sourcing

Crowd sourcing is een vorm van het verzamelen van informatie of kunde bij een grote groep mensen. Hier staat de inzet van de mens centraal. Crowd sourcing is onderzocht door literatuur onderzoek en gebruik van enquêteren via social media en aan huis.

2.6.1. Literatuur onderzoek

Er is onderzocht hoe toepasbaar een methode in de vorm van crowd sourcing is voor het inventariseren van objecten als propaantanks. Uit dit onderzoek is gebleken dat crowd sourcing voor dit onderzoek een uitkomst zou kunnen bieden afhankelijk van de combinatie van andere methoden Tabel 3 Multi Criteria Analyse. Echter is het voor crowd sourcing in het algemeen een uitdaging om voldoende respons te verzamelen voor een grootschalige inventarisatie als deze. Uit de literatuur blijkt dat het van belang is de mogelijke deelnemers te motiveren (Cotfas & Diosteanu, 2010). Afhankelijk van de gekozen onderstaande methoden kan hiervoor een passende motivatie worden gevonden. Niet elke doelgroep raakt namelijk gemotiveerd van een beloning in de vorm van een prijs of geld. Zo blijkt het van belang te zijn dat een persoon voldoening en plezier uit de participatie moet halen (*Motivations of crowdsourcing contributors*, 2018).

De mogelijke methoden die uit dit onderzoek naar voren komen zijn:

- Het organiseren van een mapathon voor studenten (<https://www.missingmaps.org/host>).
 - Door middel van een game met leaderboard waar een relevante prijs aan verbonden is, zoals een budget voor gedeeltelijke afbetaling studieschulden.
- Integratie van een app voor wandelaars en fietsers die propaantanks en andere elementen kunnen aangeven.
 - Beloningssysteem door samenwerking met partijen als de anwb (sparen voor korting of producten)
- Enquetes voor de mogelijke bezitters van propaantanks uitzetten.
 - Door middel van social media kanalen
 - Door middel van directe benadering via de propaanleverancier
- Samenwerking met leveranciers van propaan(tanks).

2.6.2. Enquête onderzoek

Om het animo en de haalbaarheid van de methoden “directe benadering van de bezitters van propaantanks” in de bovenstaande lijst te peilen, is een onderzoek gestart door middel van een enquête. Deze enquêtes geven inzicht in de haalbaarheid van een methode.

Voor het enquête onderzoek is er gebruik gemaakt van survey123: <https://arcg.is/naa8>. Survey123 is een applicatie binnen de ESRI software waarmee gemakkelijk enquêtes opgezet kunnen worden. Deze service is een betaalde licentie service. Naast survey123 zijn er meerdere mogelijkheden zoals; Maptionnaire, Survey123, cartcipe, mappingforchange en social pinpoint.

Deze enquête is bedoeld om een resultaat in cijfers uit te kunnen drukken over de bereidheid van mensen om hun propaantank locatie op te geven aan het ministerie, gemeente of omgevingsdienst. De enquête is verspreid door middel van de platformen LinkedIn, Facebook en Reddit. Op deze platformen is gebruik gemaakt van het netwerk dat ter beschikking was van Niek Broekhoven en Daniël Poot. Ook is er gebruik gemaakt van de facebook groep “living off the grid NL”. Voor extra respons op de enquête is er ook gebruik gemaakt van het

enquêteren aan huis. Hierbij zijn de huizen bezocht die eerder in het “fysiek inwinnen van propaantanks” methode zijn ingewonnen.

Voor een mapathon is er ook gekeken naar het animo en mogelijke motivatie voor leerlingen op een GIS opleiding om mee te doen met een mapathon. Een mapathon is een samenkomst van vrijwilligers om objecten in kaart te brengen en op deze manier datasets volledig te maken. Hierbij zijn door middel van survey123: (<https://arcg.is/0TrDau0>) vragen opgesteld en doorgestuurd naar de leerlingen van de opleiding applied geo information science in de studie jaren 2 tot en met 4. De vragen hebben met name betrekking tot de bereidheid en eventuele motivatie voor een mapathon. Hierbij wordt gekeken naar eventuele wedstrijd elementen en prijzen om betere en snellere resultaten te behalen.

Voor een mapathon is het een randvoorwaarde dat dit goed wordt georganiseerd aan de hand van een duidelijke uitleg over wat moet worden gedaan aan de deelnemer.

Deze stap en een algeheel stappenplan zijn opgenomen in bijlage 8.9. Mapathon. Hierbij is gebruik gemaakt van het stappenplan van OpenStreetMap, (2017). Ook zijn er een aantal extra randvoorwaarden aan het uitvoeren van een mapathon:

- Aanwezigheid van ervaren mappers (zonder ervaren mappers kan er geen duidelijke uitleg of hulp worden gegeven, dit resulteert in slechtere betrouwbaarheid van de data).
- Aanwezigheid van It'er (zonder It hulp kan een kleine technische fout er al voor zorgen dat niemand aan het werk kan)
- Goede wifi verbinding
- Beschikbaarheid over data die gebruikt wordt (in het geval van de REV casus: satelliet beelden met resolutie < 10cm en eventueel de data die gevalideerd dient te worden, de herkende objecten door een machine learning).
- Locatie voor het evenement

2.7. Data opvragen bij derden

2.7.1. Gasleveranciers

Als alternatief op verschillende data inwin methodieken is het ook een mogelijkheid om de gasleveranciers van propaan zelf te benaderen voor een locatie dataset van hun afnemers. Op deze wijze kunnen alle propaantanks in een dataset geleverd worden door de gasleveranciers en is het ook duidelijk welke propaantanks nog actief in gebruik zijn. Bij deze methodiek zitten wel een aantal obstakels. Zo zal niet alle informatie geleverd mogen worden door de gasleveranciers wegens privacyregels. Ook zullen niet alle gasleveranciers hun data op dezelfde wijze opslaan en dus is het een erg versnipperde dataset met veel leveranciers. Voor concurrentie redenen kan het opvragen van de data ook mogelijk voor problemen zorgen bij de gasleveranciers. Niet elke gasleverancier wil natuurlijk zijn volledige klantenbestand online delen met zijn concurrent.

Naast de gasleveranciers zelf is het ook mogelijk om de overkoepelende vereniging voor gas bedrijven de VVG (vereniging vloeibaar gas) te benaderen. Eventueel beschikt deze vereniging over een overkoepelende dataset Netbeheerders, maar hierop wil de vereniging geen antwoord geven.

Er wordt aangenomen dat propaantanks alleen gebruikt worden in buitengebieden waar geen gasnet aangelegd is. Hierdoor zijn er al veel locaties te filteren wanneer er data beschikbaar is van gasleveranciers hun leidingen en aansluitpunten. Sommige gasleveranciers hebben deze data als open source op hun website staan of als opvraag service zoals Croteq, Enexis en Westland infra.

Op deze wijze kan het onderzoeksgebied verkleind worden van geheel Nederland tot "Nederland zonder gasaansluiting" (Netbeheer Nederland, 2022).

2022 Gas



*Figuur 5 Gasleveranciers
Nederland (Energieleveranciers,
2022)*

2.8. Multi Criteria Analyse

Om de eerder genoemde methoden tegen elkaar af te wegen is een multi criteria analyse (MCA) opgesteld. In deze MCA worden de methoden gescoord op basis van betrouwbaarheid (*in hoeverre is een methode accuraat en precies genoeg*), duur (*hoeveel tijd gaat er over de methode heen*) en kosten (*hoeveel kost uitvoering van de methode*) van het project.

Door het uitzetten van een vragenlijst bij stakeholders van het project (projectleiders, medewerkers van omgevingsdiensten en opdrachtgevers) is de weging van de drie criteria gegenereerd. Hiervoor is een gedetailleerdere uitwerking in Bijlage Multi Criteria analyse beschrijving. De resultaten zijn gevisualiseerd in Tabel 2 Uitkomst vragenlijst criteria.

Na het verzamelen van de antwoorden van de vragenlijst is het MCA model in Tabel 3 Multi Criteria Analyse uitgewerkt aan de hand van de volgende stappen. De criteria zijn aan de linkerkant van de tabel gezet. De methoden behoren aan de bovenkant. De criteria (betrouwbaarheid, duur en kosten) worden gekoppeld aan de factor. In de vakken waar de factoren en methoden kruisen, wordt een methode gescoord op de bijbehorend criterium. Deze score strekt van 1 (slecht scorend op dit criterium) tot en met 5 (uitstekend scorend op dit criterium). Waarop deze scores worden gebaseerd is te vinden in Bijlage Multi Criteria analyse beschrijving

Wanneer voor elke methode een score is gegenereerd na vermenigvuldiging met de factor van het criterium, wordt de som per methode berekend, zodat elke methode een totaalscore krijgt. Deze totaalscore reikt van 1 t/m 10.

De methoden zijn nu gescoord op basis van betrouwbaarheid, duur en kosten. Hierdoor kan worden bepaald welke methode het best passend is op basis van deze criteria. Om meer context te creëren en de realiseerbaarheid mee te meten is in hoofdstuk 3.5. Vergelijkingsmatrix gemaakt.

3. Resultaten

Tijdens het onderzoek zijn de meeste methoden onderzocht aan de hand van pilot uitvoeringen. Uit deze pilot uitvoering zijn resultaten gekomen waarop de drie criteria voor de MCA getoetst kunnen worden. Ook is er gekeken of een methode realiseerbaar is voor geheel Nederland. Om een goed overzicht te houden zijn de scores van de MCA, werkwijze van de methode, realiseerbaarheid van de methode en eventuele opmerkingen toegevoegd in een vergelijkingsmatrix.

3.1. Resultaten fysiek data inwinnen

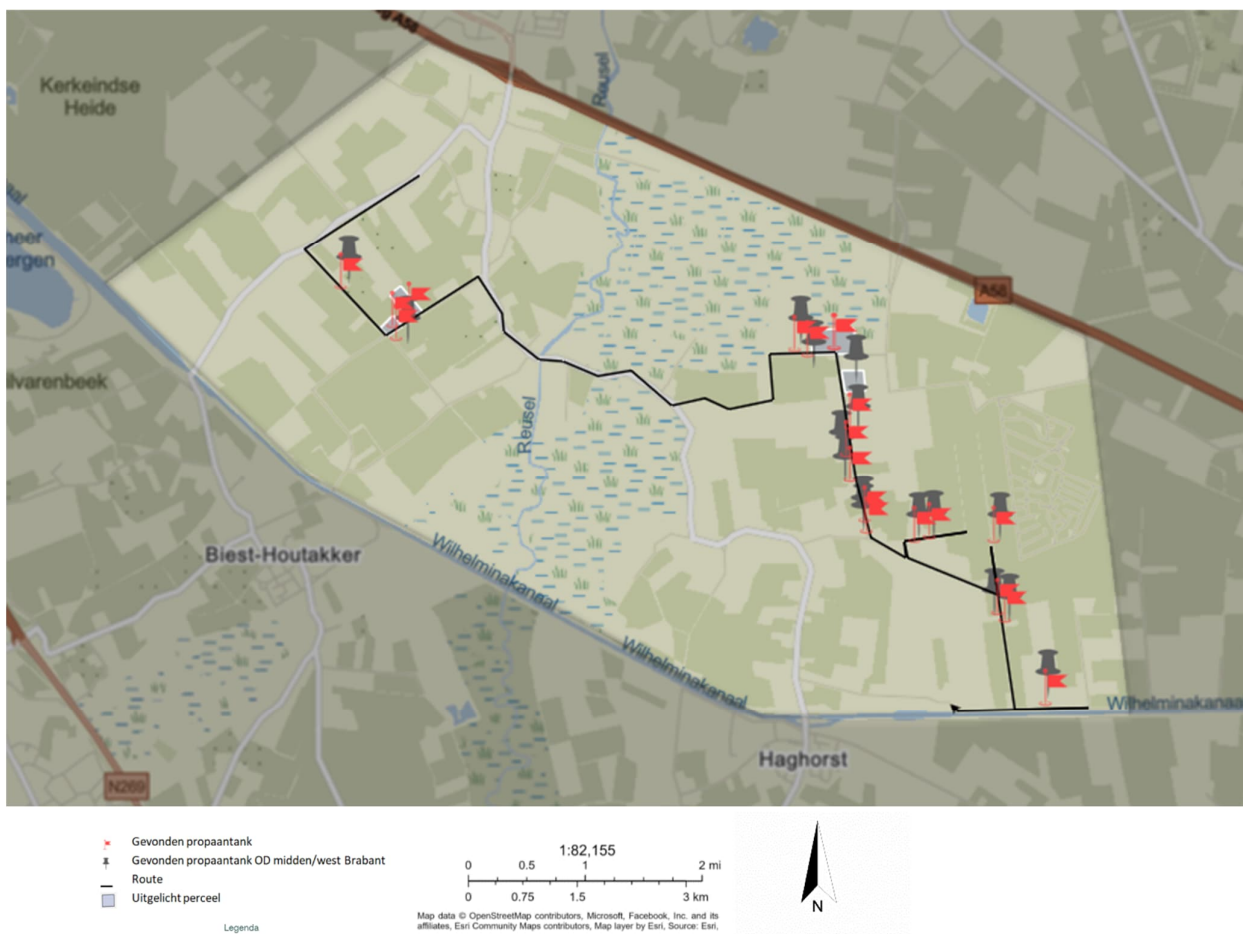
3.1.1. Fysiek data inwinnen met voorbereiding

Tijdens het monitoren zijn er verschillende resultaten geboekt en een 18 tal propaantanks ingewonnen. De accuraatheid van de locatie is hierbij hoog, doordat de propaantanks direct door een werknemer wordt geïnventariseerd. Alle propaantanks op de route die vanaf de weg zichtbaar zijn, zijn opgeslagen. Het is dus mogelijk dat er propaantanks niet zichtbaar waren en dus niet zijn geïnventariseerd.

In Figuur 6 Resultaten Data fysiek inwinnen met planning zijn de 18 propaantanks opgenomen door middel van een rode vlag. Hierbij zijn een aantal percelen uitgelicht door middel van een polygoon met witte rand. Deze percelen geven aan dat er ofwel een propaantank is gevonden die niet bekend was bij omgevingsdienst Midden-West-Brabant of een propaantank die niet tijdens het fysiek data inwinnen is gevonden maar wel bekend is bij omgevingsdienst Midden- West-Brabant. Tijdens deze methode zijn er 3 propaantanks gevonden die niet bekend waren bij de omgevingsdienst en is er 1 propaantank niet gevonden die wel bekend was bij de omgevingsdienst.

Op de geïnventariseerde locaties is tijdens het enquête onderzoek voor de methode crowd sourcing een bezoek gebracht om meer respons op de enquête te krijgen. Dit is in de lente gebeurd waardoor er ook gekeken is hoeveel propaantanks niet meer zichtbaar zijn doordat er mogelijk bladeren voor hangen. Hieruit is gekomen dat 3 van de 18 propaantanks niet zichtbaar waren door bosjes of bomen die in de winter wel zichtbaar waren.

Op basis van de gevonden resultaten en de inschatting van ongeveer 13.000 niet gevonden propaantanks is er een schatting gemaakt op de hoeveelheid uren en kosten die deze methodiek zal hebben, bijlage 8.7. Berekening Fysiek data inwinnen. Deze schatting is gemaakt op basis van een landelijke signaleringskaart (Relevant netwerk externe veiligheid, z.d.) waar 4467 propaantanks staan. Boudewijn de Hoop (omgevingsdienst rivierenland) heeft ingeschat dat dit ongeveer 25% van het totaal is.



Figuur 6 Resultaten Data fysiek inwinnen met planning

3.1.2. Resultaat

Het inwinnen van data door middel van fysiek data inwinnen is een mogelijkheid voor het in kaart brengen van objecten binnen het REV, waarvan in dit geval propaantanks. De accuraatheid van de data en de actualiteit is hoger dan bijvoorbeeld de resulterende data bij machine learning. De locatie en actualiteit van de propaantanks worden namelijk direct geverifieerd wordt door een werknemer. Wel dient er rekening gehouden te worden dat niet alle vaste propaantanks zichtbaar zijn vanaf de weg en het veel tijd en daardoor ook geld scheelt om vooraf juiste routes uit te zetten.

De duur van deze methode is ongeveer 20 werkweken aan de hand van de berekening en de totale kosten zijn geschat op afgerond €27.000,- euro. Er wordt verwacht dat met deze methode in een bladerloos seizoen ongeveer 90% van de propaantanks kunnen worden ingewonnen, de overige 10% zijn niet zichtbare propaantanks vanaf de weg.

3.1.3. Fysiek data inwinnen zonder voorbereiding

Tijdens het fysiek data inwinnen zonder voorbereiding zijn er significant slechtere resultaten geboekt dan tijdens het fysiek data inwinnen met voorbereiding. De accuraatheid en actualiteit van de data zijn hierbij wel hoog, aangezien de data veldwaarnemingen zijn en direct gevalideerd door een werknemer. Alle propaantanks op ons pad die zichtbaar zijn vanaf de weg zijn opgeslagen. Het is dus mogelijk dat er propaan tanks niet

zichtbaar waren en dus niet zijn ingewonnen. Tijdens het inwinnen zijn er 2 propaantanks gevonden, gevisualiseerd in Figuur 7 Resultaten data fysiek inwinnen zonder planning. Vergeleken met de 18 propaantanks die er tijdens het data inwinnen met voorbereiding zijn gevonden is dit een significant groot verschil. Doordat het verschil zo groot is wordt deze methode niet verder behandeld en wordt er geen schatting gemaakt op de kosten en duur van deze methode.



Figuur 7 Resultaten data fysiek inwinnen zonder planning

3.1.4. Resultaat

Het inwinnen van data door middel van fysiek data inwinnen zonder voorbereiding is geen lucratieve wijze van data inwinnen. Het verschil tussen de twee fysiek data inwinnen methoden is significant zo groot dat het inwinnen van data zonder voorbereiding geen verder onderzoek behoeft.

3.2. Machine learning

Voor de machine learning methode zijn twee verschillende methodieken onderzocht. Hierbij onderscheiden we een deep learning model dat classificeert en een deep learning model dat objecten herkent.

3.2.1. Classificatie met deep learning

Het classificeren van Propaantanks door middel van deep learning gebaseerd op de pixelwaardes in de RGB banden van het raster creëert een model waar er veel zogeheten “valse positieven” uit komen. Dit houdt in dat er veel objecten herkend worden die geen propaantank zijn, maar wel door het model herkend worden als propaantank. Het classificatie model is uitgevoerd op 4 luchtfoto's met een bereik van 4 bij 4 kilometer. Het aantal propaantanks in dit gebied waren 12 propaantanks waarvan er 2 propaantanks herkend zijn door het model, naast de 2 propaantanks die herkend zijn waren er 54 valse positieven. Dit is onderzocht door een “Summarize within” analyse waarbij de resultaten van het classificatie model geselecteerd worden wanneer ze overlappen met een van de al bekende propaantank locaties.

De reden voor valse positieven is onderzocht door middel van een spectrale analyse, waarbij in bijlage 8.3 (Deep learning proces) te zien is dat de spectrale analyse aantoont dat er veel overeenkomsten zijn in de pixelwaardes van verschillende objecten zoals; witte daken, busjes en aanhangers. Deze uitkomst is de verklaring voor het resulteren in vooral valse positieven als resultaat van het deep learning model met classificatie.

3.2.2. Object herkenning met deep learning

Het herkennen van propaantanks door middel van object herkenning resulteert wederom in met name valse positieven, maar door het verder trainen van een model zou het model uiteindelijk accuraat genoeg kunnen zijn om propaantanks te herkennen met meer accuraatheid. Uit de resultaten van het uitvoeren van het object herkenning door middel van deep learning model zijn door middel van een “Summarize within” analyse alle gevonden propaantanks gevisualiseerd in Tabel 1: Resultaat objectherkenning - gevonden propaantanks. In Figuur 8 Visualisatie herkenning propaantank. | Gele polygoon is herkende object is de overlap van een van de gevonden propaantanks) gevisualiseerd.

In het onderzoeksgebied van 20 propaantanks op een vlak van 6 bij 6 kilometer zijn 5 propaantanks gevonden. Hierbij zijn ook 58 valse positieven gevonden, maar dit aantal heeft met name te maken met het “confidence” level dat laag gehouden is tijdens het uitvoeren van het model. Het “confidence” level geeft de zekerheid aan van een door het model gevonden propaantank. Het “confidence” level geeft de zekerheid aan van een door het model gevonden propaantank. Het “confidence” level bij deze uitvoering was 0,65 (65 procent zekerheid). 45 valse positieven vallen binnen de 0,65 en 0,80 “confidence” level. Door een verhoging van dit level kunnen er dus meer valse positieven direct gefilterd worden.

Tabel 1: Resultaat objectherkenning - gevonden propaantanks

OBJECTID*	Shape*	Classcode	Classname	Classvalue	RED	GREEN	BLUE	Count	Shape_Length	Shape_Area	Summarized Area in SQUAREKILOMETERS	Count of Polygons
1	25	Polygon ZM	propaantank	1	150	13	184	182	5.032669	1.169075	0.000007	1
2	26	Polygon ZM	propaantank	1	150	13	184	178	5.231244	1.141892	0.000006	1
3	27	Polygon ZM	propaantank	1	150	13	184	475	8.44707	3.041065	0.000003	1
4	30	Polygon ZM	propaantank	1	150	13	184	314	6.64257	2.010659	0.000002	1
5	31	Polygon ZM	propaantank	1	150	13	184	286	6.500041	1.83387	0.000003	1

Click to add new row.



Figuur 8 Visualisatie herkenning propaantank. | Gele polygoon is herkende object

Uit gesprekken met professionals op het gebied van Deep learning (David Yu, Loek van Ooijen en Willem Vlot) is gebleken dat ze vertrouwen hebben in het herkennen van propaantanks met een deep learning model. Wel geven de professionals aan dat dit nog niet eerder ontwikkeld is en de ontwikkelingsfase van een deep learning model met een goede betrouwbaarheid minimaal een half jaar duurt. De kosten hiervan zijn door Loek van Ooijen (NEO) geschat op zo'n 50.000 euro.

3.3. Crowd sourcing

In Bijlage Bereidheid melden propaantank enquête zijn de resultaten opgenomen van het enquête onderzoek. Door een lage respons t.o.v. de verwachting (19 respondenten) op het enquête onderzoek zijn de resultaten niet significant, maar geeft het wel een indicatie over de bereidheid van mensen om wel of niet hun propaantanks op te geven bij een overheidsinstantie of gemeente.

De resultaten tonen aan dat de meeste mensen het bezit van hun propaantank niet op zouden geven (het dubbele van het aantal mensen die dit wel zou doen). De hoofdreden voor respondenten hierbij is dat het geen meerwaarde voor hen heeft of het gebrek aan vertrouwen in de overheid. Respondenten geven aan dat ze de beweegredenen van de overheid hierin niet vertrouwen, gezien de mogelijkheid dat hier in de toekomst consequenties aan verbonden kunnen worden.

Het relatief lage aantal mensen dat mee zou willen werken heeft gezorgd voor een nieuwe denkwijze, waar andere manieren van crowd sourcing zijn onderzocht. In Bijlage animo meting mapathon is te zien dat onder een peiling bij studenten AGIS een groter draagvlak is om mee te doen aan een vorm van een mapathon. In een mapathon komen meerdere mensen samen met het doel objecten te definiëren op basis van satellietbeelden. Wanneer hier voor een student een beloning tegenover staat, zoals in de resultaten te zien, zijn ze eerder bereid om werk voor dit doel te verrichten. De kosten van een dergelijke methode zijn aanzienlijk lager dan een opdracht met werknemers van bedrijven, want voor studenten voldoet een kans op een voucher o.i.d. al. Ook is het idee van een applicatie in samenwerking met een organisatie (bijvoorbeeld de ANWB, vanwege de vele wandelaars) tot stand gekomen. Zo kan er een puntensysteem in een applicatie ontwikkeld worden voor korting op spullen in de ANWB winkels (motivatie voor mensen). Door deze groep wandelaars te benaderen door middel van een applicatie met puntensysteem wordt er direct een doelgroep gemotiveerd die op vele buitengebieden in Nederland komen. Er wordt verwacht dat deze doelgroep bereid is verschillende locaties te bezoeken voor het gebruik van de applicatie.

3.4. Multi Criteria Analyse

Voor de MCA is ter voorbereiding een vragenlijst opgesteld met beoordeling van criteria, waarin de vraag wordt gesteld welk criterium belangrijker is op een schaal van 1 t/m 5. In Bijlage Multi Criteria analyse beschrijving is een weergave van de vragen en opzet van de vragenlijst te vinden. De vragen zijn gesteld aan de projectleiders, opdrachtgevers en stakeholders van de omgevingsdiensten.

De uitkomsten van de vragenlijst zijn gevisualiseerd in Tabel 2 Uitkomst vragenlijst criteria. Deze tabel geeft weer welke score is gegeven bij de vraag 'Wat vindt u belangrijker?'. Hoe dichterbij de score de 1 komt, hoe belangrijker het eerste criterium is. Andersom geldt dat een score dichterbij de 5 een zwaardere weging voor het tweede criterium bij de vraag geeft.

Tabel 2 Uitkomst vragenlijst criteria

Wat vindt u belangrijker?	Wat vindt u belangrijker?	Wat vindt u belangrijker?
Betrouwbaarheid (1) – duur (5)	Kosten (1) – betrouwbaarheid (5)	Duur (1) – kosten (5)
2	4	4
2	4	3
5	5	2
2	3	4
2	2	4
3	3	3
2,7	3,5	3,3

Na het ontvangen van deze resultaten is het gemiddelde van de scores berekend, zoals dikgedrukt te zien is in de tabel. De afwijking van de mediaan (3) bepaalt de factor waarin een criterium meespeelt. Een verdere uitleg van de berekeningen is te vinden in Bijlage Multi Criteria analyse beschrijving.

Na het ontvangen van de respons is een van elk van de drie vragen berekend hoe ver een criterium vanaf 3 (het gemiddelde) afwijkt in positieve en negatieve getallen. Deze drie afwijkingen bij elkaar opgeteld geven een getal van -1 tot en met 1, waar een factor aan verbonden kan worden. Deze factoren zijn te zien Tabel 3 Multi Criteria Analyse. Hoe tot deze uitkomsten is gekomen, is te vinden in Bijlage Multi Criteria analyse beschrijving. Op basis van deze factoren is een waarde van 1 t/m 5 gegeven. Deze waarden zijn in de tabel al verrekend met de formule: ' $\text{CEL FACTOR} * \text{waarde}$ ' in Excel. De waarden zijn gegeven op basis van de volgende aannames:

Betrouwbaarheid:

Gebaseerd op eigen onderzoek

Duur van het project:

- Half jaar geeft score 5
- 1 jaar geeft score 3
- 2 jaar of meer geeft score 1

Kosten van het project:

- 50.000 euro en minder geeft score 5
- 100.000 euro geeft score 3
- 200.000 euro en meer geeft score 1

Deze scores samen geven een cijfer aan elke bovenstaande methode op een schaal van 1 tot en met 10. Deze MCA geeft echter geen inzicht op de realiseerbaarheid. Hiervoor wordt naar

Tabel 4 methodieken vergelijking matrix verwezen.

Tabel 3 Multi Criteria Analyse

Factor	Machine learning classificatie	Machine learning objectherkenning	Fysiek inwinnen met voorbereiding	Fysiek inwinnen zonder voorbereiding	Mapathon	Enquete bezitters propaantanks
Betrouwbaarheid	1	2	4	4	4	4
Duur van het project	0,3	1,2	1,2	0,3	1,2	0,9
Kosten van het project	0,7	2,8	3,5	0,7	3,5	2,8
Totaal		5	6	8,7	5	8,7

3.5. Vergelijkingsmatrix

In Tabel 4 'methodieken vergelijking matrix' wordt een tabel weergegeven die bij elke methode een breder beeld geeft van de realiseerbaarheid, dekking van het gebied, de workflow en de opmerkingen in combinatie met de uitkomsten van de MCA. De kolommen 'opmerkingen' en 'workflow' bieden een bredere context aan de bijbehorende cijfers.

De tabel dient tevens als ondersteuning voor de discussie, conclusie en het advies.

Tabel 4 methodieken vergelijking matrix

Methodiek	Realiseerbaar	Betrouwbaarheid	Kosten	Duur	Vlakdekkend	Workflow	Opmerkingen	MCA score (totaal)
Classificatie (machine learning)	Uit onderzoek: Nee Misschien mogelijk met meer spectrale banden dataset.	Laag (zie opmerking)	Onder €50.000 BRON: NEO (Schatting voor een minimum betrouwbaarheid ongeveer 70%)	< half jaar (Schatting Neo expert voor 75% betrouwbaarheid)	Geheel Nederland	Machine learning trainen -> beelden segmenteren -> input data filteren -> input data segmenteren -> uitvoeren getraint model -> uitkomsten evalueren	Veel ruis door de pixelwaardes van propaantanks die overeenkomen met andere objecten. Deze bevindingen zorgen voor weinig resultaten uit dit onderzoek. Mogelijk een goede methode wanneer multi spectrale data beschikbaar is op een juiste resolutie. Momenteel is dit beschikbaar in een niet toereikende resolutie van 25cm.	5

Methodiek	Realiseerbaar	Betrouwbaarheid	Kosten	Duur	Vlakdekkend	Workflow	Opmerkingen	MCA score (totaal)
Object herkenning (machine learning)	Ja	Laag (zie opmerking)	Onder €50.000 Bron: NEO	< half jaar	Geheel Nederland	Machine learning trainen -> input data filteren -> uitvoeren getraint model -> uitkomsten evalueren	<p>Deep learning model trainen en uitvoeren kost meer rekenkracht dan beschikbaar. Dit zorgt voor weinig meetbaar resultaat uit dit onderzoek.</p> <p>Uit gesprekken met professionals zou het mogelijk moeten zijn propaantanks in te winnen door middel van herkenningstechniek met deep learning.</p>	6

Methodiek	Realiseerbaar	Betrouwbaarheid	Kosten	Duur	Vlakdekkend	Workflow	Opmerkingen	MCA score (totaal)
Via leveranciers (VVG)	Nee (zie opmerking)	Hoog	Laag (Kosten betreffen werkuren, geen materiaal of factuur kosten)	< half jaar Aanvragen kan langer duren Mail opstellen (kost een dag) Samenvoegen van alle datasets (plus minus week wanneer alles ruimtelijke data is)	Geheel Nederland (Bij het meewerken van leveranciers of de VVG)	Email opgesteld en verzonden naar zowel VVG als betrokken partijen. Gebeld met verschillende partijen.	Leveranciers en VVG geven aan niet bereid te zijn data af te staan. Verschillende redenen waaronder; privacy voor de klant, gevoeligheid klantenbestand, niet toegestaan door VVG.	n.v.t.
Crowd sourcing (particulieren zelf laten melden)	Nee (zie opmerking)	Hoog	Laag (kosten voor het rondsturen van directe brieven)	Onbekend Schatting: 0,5 tot 1 jaar (afhankelijk van de snelheid dat mensen hun propaantank zouden aangeven)	Methode is uit de resultaten van de enquête geen goede optie. Mensen geven aan hier wegens privé redenen niet aan bij te willen dragen.	Brief opstellen / Email opstellen -> insturen naar mogelijke propaantank eigenaren.	Afhankelijk van de bereidbaarheid van particulieren zelf. Uit onderzoek blijkt het grootste gedeelte van de mensen niet bereid hun propaantank te melden.	7,7

Methodiek	Realiseerbaar	Betrouwbaarheid	Kosten	Duur	Vlakdekkend	Workflow	Opmerkingen	MCA score (totaal)
Crowd sourcing (mapathon)	Ja	Hoog (studenten met GIS kennis zorgt over het algemeen voor goede data)	Laag (bij een wedstrijd gerichte mapathon kan er gebruik gemaakt worden van kleine prijzen, bijvoorbeeld max. 250 euro tegoedbonnen en linturen)	Onbekend (afhankelijk van het aantal deelnemers en mapathon sessies) Schatting tussen half jaar en een jaar	Is een rendabele manier om propaantanks in te winnen met studenten die verstand hebben van intekenen op geografische data.	Samenwerking met GIS opleidingen zoeken -> Organiseren mapathon -> Duidelijke informatie verschaffen -> inwinnen en controle ronden	Rekening gehouden te worden met de actualiteit van de satellietdata.	8,7
Fysiek inwinnen (voorbereid met mogelijke gebieden)	Ja	Hoog (accurate locatie en actueel)	± 36.000 Een volledige berekening is te vinden in Berekening Fysiek data inwinnen	1 uur = ± 18 tanks Een volledige berekening is te vinden in Berekening Fysiek data inwinnen	Geheel Nederland wanneer mogelijk	Rondrijden met voorbedacht gebieden waar mogelijk propaantanks aanwezig zijn. (buitengebied met name)	Hoge kwaliteit. Propaantanks worden ingewonnen op de juiste locatie wanneer zichtbaar (niet alle woningen hebben de propaantanks zichtbaar staan).	8,7

Methodiek	Realiseerbaar	Betrouwbaarheid	Kosten	Duur	Vlakdekkend	Workflow	Opmerkingen	MCA score (totaal)
Fysiek inwinnen (zonder voorbereiding)	Ja (niet lucratief)	Hoog (accurate locatie en actueel)	Hoog (rondrijden zonder opgesteld plan zorgt voor een slechte werkwijze)	1 uur = ± 2 tanks schatting manuren: 13.000 tanks / 2 = 6.500 uur (niet lucratief vergeleken met fysiek inwinnen met voorbereiding)	Geheel Nederland wanneer mogelijk	Rondrijden op gebieden waarvan mogelijk propaantanks staan (keuze voor buitengebied zonder voorbereiding van route)	Hoge kwaliteit. Propaantanks worden ingewonnen op de juiste locatie wanneer zichtbaar (niet alle woningen hebben de propaantanks zichtbaar staan en het kost veel manuren vooral door de slechte voorbereiding. Kost meer manuren dan fysiek inwinnen met voorbereiding).	5

3.6. Beantwoorden van deelvragen

In dit hoofdstuk worden de opgestelde deelvragen op basis van de methoden en resultaten beantwoord.

“Wat zijn kenmerken van propaantanks die vastgelegd dienen te worden in de dataset van propaantanks?”

Er zijn verschillende kenmerken die vastgelegd dienen te worden in de REV dataset van propaantanks. Deze kenmerken zijn vastgelegd in een standaardisatie door de stichting Geonovum. Deze standaardisatie heet de IMEV (Geonovum, 2022). Tijdens het project is de prioriteit gelegd op het inwinnen van propaantank locaties.

“Welke methoden zijn er voor het slim inwinnen van data?”

Tijdens het project zijn er drie verschillende methoden onderzocht voor het inwinnen van data. Crowd sourcing, Machine learning en fysiek data inwinnen. Alle drie de methoden zijn mogelijkheden om slim data in te winnen. Het onderscheid zit in de duur, betrouwbaarheid en kosten van de methoden.

“Welke methode is passend voor het inwinnen van data voor de propaantank dataset?”

Uit de resultaten van het onderzoek 3.4. Multi Criteria Analyse is gebleken dat fysiek data inwinnen met voorbereiding en het inwinnen door middel van een mapathon de beste methoden zijn om uit te voeren gebaseerd op de 3 criteria die gesteld zijn. Wel dient hierbij rekening gehouden te worden dat een mapathon mogelijk niet Nederland vlak dekkend is. Dit is afhankelijk van de opkomst voor een mapathon en hoe vaak er een mapathon gehouden wordt.

Ook de machine learning methode zou een mogelijke methode kunnen zijn voor het inwinnen van data voor de propaantank dataset. Hierbij zijn de huidige resultaten uit dit onderzoek alleen nog niet toereikend genoeg om daar een uitspraak op te doen.

“Welke methode is haalbaar voor het slim inwinnen van propaantank data voor heel Nederland?”

De methoden fysiek data inwinnen met voorbereiding en een machine learning model zijn mogelijkheden om propaantank data in te winnen voor heel Nederland. De machine learning methode dient hiervoor de juiste data beschikbaar te hebben en een pc met veel rekenkracht.

De methode crowd sourcing zou ook een mogelijkheid zijn om data in te winnen voor heel Nederland. Dit is alleen afhankelijk van de opkomst bij bijvoorbeeld een mapathon of de hoeveelheid dat een collector app gebruikt wordt.

“Welke methodes vallen onder de mogelijkheid om slim data in te winnen van propaantanks, waarbij de methodes nuttig zijn voor het register externe veiligheidsrisico's?”

Crowd sourcing -> Het inzetten van een mapathon voor studenten met een GIS gerelateerde opleiding.

Fysiek data inwinnen -> Rondrijden met een voorbedachte route door de gedefinieerde buitengebieden.

Machine learning -> Naarmate deze beter geoptimaliseerd is na een eerste inwinning op basis van fysiek inwinnen en een mapathon. De machine learning kan dan worden gebruikt om veranderingen waar te nemen.

“Hoe worden de datasets gevalideerd?”

Tijdens een mapathon kan een extra validatie ronde worden toegevoegd, waardoor er binnen deze werkwijze een validatie plaatsvindt. Voor het fysiek inwinnen kan een werknemer van het ministerie of een uitbesteed bedrijf de taak op zich nemen om de ingewonnen gegevens te valideren, door middel van satellietbeelden of steekproeven.

“Hoe kunnen de datasets worden gevisualiseerd?”

De datasets zullen gevisualiseerd worden door middel van een online interactieve kaartviewer. Net als: de signaleringskaart van Relevant netwerk externe veiligheid (z.d.).

“Welke methode(n) is/zijn efficiënt, betrouwbaar en realiseerbaar als mogelijkheid om slim data in te winnen van datasets die nuttig zijn voor het Register Externe Veiligheidsrisico's?”

Alle drie de methoden afhankelijk van de uitwerking -> Combinaties van zijn ook mogelijk (Bijv. Crowd sourcing (applicatie of mapathon) om een ML beter te trainen, Fysiek data inwinnen om een ML beter te trainen)
De voorkeur zit in combinaties (Training samples kunnen ingewonnen worden dmv crowd sourcing of fysiek data inwinnen.

4. Discussie

Voor dit onderzoek zijn verschillende soorten methodieken onderzocht om propaantanks te inventariseren. Na het ondervinden en onderzoeken is uitgezocht welke van deze methodieken of combinaties hiervan het best implementeerbaar zijn voor het Ministerie van Infrastructuur en Waterstaat.

4.1. Methoden

4.1.1. Fysiek data inwinnen

Fysiek data inwinnen is een methode waarbij de actualiteit van de propaantank locatie hoger is ten opzichte van de andere methoden. Omdat bij bijvoorbeeld machine learning gewerkt wordt met satelliet beelden die maximaal een jaar oud kunnen zijn. Ook hoeft de data niet gevalideerd te worden bij fysiek data inwinnen aangezien het direct door twee werknemers geïnventariseerd wordt. Een machine learning model of studenten die een mapathon uitvoeren kunnen fouten maken waardoor een validatie proces benodigd is.

Wel dient er rekening gehouden te worden met propaantanks die niet zichtbaar zijn vanaf de weg of propaantanks die achter bosjes staan. Om dit op te lossen dient er fysiek data ingewonnen te worden in het bladerloze seizoen. Uit ervaring tijdens het fysiek inwinnen van data waarbij zowel in de zomer als de winter langs dezelfde propaantanks is gereden waren er in de zomer 3 van de 18 niet zichtbaar. Dit is een verschil van afgerond 17%.

4.1.2. Remote sensing met machine learning (classificatie)

Uit de resultaten is gebleken dat het gebruik van deze methode zorgt voor veel zogeheten "valse positieven". Bij deze methode wordt dan ook verwacht dat het niet gaat lukken om propaantanks op een efficiënte wijze in te winnen. Wel is het een mogelijkheid deze methode verder te onderzoeken wanneer er multi-spectrale luchtfoto's komen met meer dan 3 banden en een resolutie van 10cm of kleiner. De huidige multi-spectrale beelden zijn 25cm. Dit is niet toereikend genoeg om propaantanks goed te herkennen eveneens als de huidige 3 banden luchtfoto's van 8cm. Tijdens het onderzoek is de kans laten liggen om de lagere resolutie Multi spectrale luchtfoto's te combineren met de hogere resolutie RGB banden (Ranchin, 1996). Deze methode heeft wel een kanttekening, namelijk dat de "grovere resolutie" de output is (Esri, 2016). Dus bij een resolutie van 25cm en een resolutie van 8cm word de output resolutie 25cm. Hierdoor is de volle potentie van deze methode niet benut.

4.1.3. Herkenningstechniek met machine learning (object herkenning)

Bij het machine learning model op object herkenning waren er betere resultaten verwacht. Dit komt met name door de rekenkracht van de computer waardoor het model niet vaak genoeg is getraind en de relatief weinig training samples. Er zijn verschillende mogelijkheden om dit op te lossen. Zo zouden er meer training samples gemaakt kunnen worden mochten er meer bekende propaantank locaties zijn bij bijvoorbeeld een andere omgevingsdienst of uit het RRGs. Door meer training samples kan het model beter getraind worden. Het machine learning model heeft ook een groot voordeel op de andere methoden. Zo kan een machine learning model wanneer het getraind is en kwalitatief in orde is, constant uitgevoerd worden. Nieuwe propaantanks die geplaatst worden kunnen hierdoor direct opnieuw herkend worden waar andere methoden zoals een mapathon en fysiek data inwinnen telkens opnieuw uitgevoerd zou moeten worden.

Tijdens het onderzoek is een Nvidia grafische kaart gebruikt die compatibel is met CUDA software (NVIDIA, 2022). Deze specificaties blijken te weinig rekenkracht te hebben voor het draaien van een dergelijk model.

Deze grafische kaart dient minimaal 4GB VRAM te hebben. Windows versie 10 of 11 en een ondersteunde versie van visual studio code (NVIDIA, z.d.).

4.1.4. Crowd sourcing

Uit de resultaten van het onderzoek naar de bereidheid van particulieren om zelf hun propaantank aan te geven is gebleken dat de meerderheid van de ondervraagde hier niet aan bij wil dragen. Dit komt met name door een gebrek aan betrokkenheidsgevoel en wantrouwen van de eigenaar.

Door het afschrijven van deze methodiek is het zaak om binnen crowd sourcing de best mogelijke overige methodiek te zoeken. Door de positieve resultaten van het animo op een mapathon, waarin AGIS studenten op basis van satellietbeelden propaantanks inventariseren, wordt deze methodiek ook meegenomen. Op basis van eerder onderzoek van Abdellaoui, (2022) is gebleken dat het arbeidsintensief is om satellietbeelden af te struinen op zoek naar propaantanks. Een voordeel van de mapathon is dat meerdere handen licht werk maken. Een nadeel is dat een mapathon maar een beperkte tijd duurt. Tijdens het onderzoek had nog meer achtergrondinformatie over mapathons kunnen worden gezocht om zo een mapathon te organiseren en met concrete cijfers te komen, die een indicatie geven van een betrouwbaarheidspercentage.

4.1.5. Data opvragen bij derden

Het data opvragen bij gasleveranciers is na het meermaals geprobeerd te hebben niet succesvol gebleken. De gasleveranciers laten vrijwel niet weten of ze over de locatiedata beschikken van de aanwezige propaantanks. Een voordeel aan deze methodiek zou zijn dat de data van hoge betrouwbaarheid is, maar hier staat tegenover dat het arbeidsintensief is om met alle bedrijven een individuele band te onderhouden en hen zo ver te krijgen de data te leveren. Tijdens het onderzoek had nog onderzocht kunnen worden hoe de partijen reageren op een beloning tegenover het leveren van data. Hiervoor was het nodig geweest onderzoek te doen naar de AVG om te weten te komen welke informatie wel en niet mag worden gedeeld.

4.2. Combinaties discussie

Tijdens het project is het aan het licht gekomen dat er ook meerdere combinaties mogelijk zijn tussen de verschillende methoden. In dit hoofdstuk worden een aantal combinaties ter discussie gesteld.

4.2.1. Machine learning en mapathon

Op basis van het onderzoek van Abdellaoui (2022) is gebleken dat het inwinnen van propaantanks door middel van satellietbeelden arbeidsintensief is. Op basis van deze bevindingen en het gegeven dat een machine learning model op object herkenning vatbaar is voor valse positieven, is het een mogelijkheid deze twee methoden te combineren.

Om deze valse positieven en propaantanks te valideren kan er gebruik gemaakt worden van een mapathon. Door studenten de gevonden propaantanks te laten valideren ontstaat er een goedkope en gerichte vorm van validatie. Waarbij studenten alleen de, door het machine learning gevonden objecten hoeven te controleren. De nieuwe gevalideerde propaantanks kunnen vervolgens weer gebruikt worden als training samples voor meer input van het machine learning model. Om de tekortkomingen van beide methoden te voorkomen, zou deze combinatie een uitkomst kunnen bieden.

4.2.2. Machine learning met fysiek inwinnen

Omdat uit het onderzoek is gebleken dat het machine learning model o.b.v. objectherkenning beter getraind moet worden en meer training samples daar een van de onderdelen van is, is het een mogelijkheid om eerst meer training samples in te winnen door middel van fysiek data inwinnen. Door een werknemer van een omgevingsdienst of uitbesteding kan er een minimum aantal propaantanks ingewonnen worden die dienen als nieuwe training samples (eventueel is er al meer locatie data beschikbaar bij andere omgevingsdiensten). Door het inwinnen van deze training samples kan het machine learning model geoptimaliseerd worden en zal het beter presteren. Om de machine learning te kunnen laten draaien en de nadelen hiervan teniet te doen, zou het fysiek inwinnen dus een oplossing kunnen bieden. Dit zou dan afhankelijk zijn van aanwezigheid van data bij omgevingsdiensten.

5. Conclusie

In het onderzoek is een antwoord gezocht op de hoofdvraag: *'Welke methode is efficiënt, betrouwbaar en realiseerbaar als mogelijkheid om slim data in te winnen van datasets die nuttig zijn voor het Register Externe Veiligheidsrisico's?'*

Na het onderzoeken is uit de resultaten gebleken dat fysiek data inwinnen met voorbereiding een van de beste methoden is om data in te winnen. De methode scoort hoog op alle drie de onderdelen in de multicriteria analyse en is realiseerbaar voor geheel Nederland.

Machine learning door middel van classificatie is geen goede methode gebleken om propaantanks in te winnen uit dit onderzoek.

Machine learning door middel van object herkenning is op zichzelf geen goede methode gebleken uit dit onderzoek om data in te winnen. Dit komt door een gebrek aan samples en rekenkracht, deze samples kunnen wel worden gecreëerd uit andere methoden.

Het benaderen van propaantank eigenaren is geen goede methode gebleken voor het inwinnen van data. Tijdens het onderzoek is uit de resultaten van de enquête gebleken dat het grootste gedeelte van de mensen hun gegevens niet vrijwillig zouden delen.

Uit het onderzoek naar een mapathon is gebleken dat dit een goede manier is om eventueel data in te evalueren, maar niet om zelf in te winnen. Door middel van een enquête zijn resultaten geboekt over leerlingen die beschikken over GIS kennis die bereid zijn tijdens een mapathon propaantanks te valideren. Door de lage kosten en betrouwbaarheid ratio's die deze methode heeft scoort de methode hoog op de Multi criteria analyse.

Data opvragen bij derden is tijdens het onderzoek bevonden als geen goede methode om data in te winnen voor het REV.

Na het onderzoek is gebleken dat de beste resultaten en beste methoden om slim data in te winnen gedaan kan worden door de methoden "Mapathon" en "fysiek data inwinnen met voorbereiding". Dit is gebaseerd op de drie gekozen criteria die gebruikt zijn voor de multi-criteria analyse en de realiseerbaarheid van de methode. Maar uit discussie is gebleken dat ondanks de mindere resultaten, een aantal grote voordelen zijn voor machine learning met object herkenning. Zo kan een machine learning uitgevoerd worden op elk moment wanneer dit model goed getraind is. Waar mapathon en fysiek inwinnen methoden zijn waarbij er telkens opnieuw actie ondernomen moet worden wanneer er nieuwe propaantanks geplaatst zijn.

Bij een verder onderzoek naar de methode is het mogelijk om door middel van betere rekenkracht en meer training samples een beter model te maken om propaantanks te herkennen. Hier zouden betere resolutie luchtfoto's ook een verschil kunnen maken. Naast de methoden die onderzocht zijn is er ook nagedacht over combinaties tussen verschillende methoden. Hierbij is het belangrijk dat sommige methoden elkaar kunnen aanvullen, zoals de combinatie van machine learning object herkenning met een mapathon. Hierbij kunnen tijdens een mapathon de door het machine learning model herkende objecten gevalideerd worden. Dit zorgt voor een efficiënte werkwijze tijdens de mapathon waarbij studenten de gevonden objecten valideren in plaats van satelliet beelden doorzoeken naar propaantanks. Deze combinatie verhelpt de pijnpunten van beide methoden, de valse posities voor machine learning, de herhalende uitvoering voor een mapathon en de arbeidsintensiviteit van een mapathon en is daarom een efficiënte, betrouwbare en realiseerbare mogelijkheid tot het inwinnen van data die nuttig zijn voor het REV.

6. Advies en aanbevelingen

Hier wordt het advies kenbaar gemaakt, waarna de aanbevelingen voor een vervolg worden gegeven.

6.1. Advies

Uit het onderzoek en de resultaten van de Multi-criteria analyse is gebleken dat de twee methode “mapathon” en “fysiek inwinnen” het hoogste scoren op de Multi-criteria analyse. Maar uit het onderzoek is ook gebleken dat geen enkele methode geheel foutloos is, alle onderzochte methoden hebben zijn voor en nadelen. Uit dit gegeven is het daarom belangrijk om de negatieve kanten van een methode aan te vullen met de positieve kanten van een andere methode.

Gezien de toegevoegde waarde van het machine learning model, waarbij na het opstellen van een model deze methode elke dag zonder moeite opnieuw uitgevoerd kan worden. Wordt er geacht dat dit de best passende methode is voor een organisatie als het Ministerie van Infrastructuur en Waterstaat, omdat er verwacht wordt dat er de komende jaren nog steeds nieuwe propaantanks geplaatst worden. Om het probleem van valse positieven op te lossen is de best passende combinatie het gebruik van een machine learning met een mapathon als validatie van de gevonden objecten.

Voor het opstellen van een machine learning model dient er beschikking te zijn over training samples om het machine learning model te trainen. Er wordt aangeraden aan het ministerie om naast de in dit project ingewonnen 125 training samples, aanvraag te doen binnen de organisatie en aangesloten omgevingsdiensten naar locatie data van propaantanks. Alle gevalideerde geïnventariseerde propaantanks helpen hierbij. Er word een minimum van 150 training samples gesteld, maar idealiter zijn dit 250 training samples (David Yu, ESRI AI specialist, Online communicatie). Wanneer dit niet haalbaar is met de beschikbare data, wordt het ministerie aangeraden om veldwerknemers van omgevingsdiensten aan te sturen om tijdens het werk ook propaantank locaties te inventariseren.

Het ministerie van Infrastructuur en Waterstaat wordt aangeraden de ontwikkeling van en machine learning model uit te besteden, gezien het feit dat het ontwikkelen van een machine learning model specifieke vakkennis benodigd. vakkennis die beschikbaar is bij partijen zoals bijvoorbeeld NEO en Burokarto. Wanneer het ministerie er voor kiest deze machine learning modellen intern op te bouwen of uit te voeren dient er rekening gehouden te worden met een aantal randvoorwaarden :

- Gebruik van een pc met een grafische kaart die minimaal 4GB VRAM heeft, er word aangeraden te werken met 8GB VRAM
- Een minimum van 150 training samples
- CUDA software voor het gebruik van de grafische kaart (Esri, z.d.-d)
- Grafische kaart die compatibel is met CUDA (NVIDIA, 2022).
- Visual studio code (NVIDIA, 2022a)
- Windows versie compatibel met CUDA (NVIDIA, 2022a) (Windows 10 of 11)

De kosten van een machine learning model zijn ongeveer €50.000,- euro. Het duurt een half jaar om het model op te stellen. Hierbij wordt de betrouwbaarheid op ongeveer 70% geschat (Loek van Oijen, AI specialist NEO, Persoonlijke communicatie). Er wordt aangeraden dit op te schalen naar een betrouwbaarheid die passend geacht wordt door het Ministerie.

Bij het uitvoeren van het machine learning model zullen er altijd valse positieven worden gevonden. Dit zijn gevonden objecten die niet het object betreffen waar het model voor getraind is. Hierbij komt de mapathon van pas. Door gebruik te maken van vrijwilligers kunnen de gevonden objecten gevalideerd worden tot het object dat ingewonnen dient te worden. Voor de uitvoering van een mapathon kan het stappenplan gevolgd worden in bijlage 8.9. Mapathon.

Voor een mapathon dient minimaal aan de volgende randvoorwaarden voldaan te worden (OpenStreetMap, 2017):

- Aanwezigheid van ervaren mappers (zonder ervaren mappers kan er geen duidelijke uitleg of hulp worden gegeven, dit resulteert in slechtere betrouwbaarheid van de data).
- Aanwezigheid van It'er (zonder It hulp kan een kleine technische fout er al voor zorgen dat niemand aan het werk kan).
- Goede wifi verbinding
- Beschikbaarheid over data die gebruikt wordt (in het geval van de REV casus: satelliet beelden met resolutie < 10cm en eventueel de data die gevalideerd dient te worden, de herkende objecten door een machine learning).
- Locatie voor het evenement

De kosten voor een mapathon zijn met name de kosten die gemaakt worden voor de evenementlocatie en de versnaperingen die aangeboden worden tijdens de mapathon. Hierbij is het belangrijk dat het ministerie mogelijk beschikt over een locatie waar de mapathon georganiseerd kan worden of op een locatie zoals bij de HAS Hogeschool. In dat geval dienen de kosten voor een locatie niet meegenomen te worden. Bij de volgende schatting wordt uitgegaan van een mapathon met ongeveer 50 tot 100 vrijwilligers. Hierbij is een schatting gemaakt van €45,- tot €75,- per persoon voor een arrangement van 8 uur inclusief lunch (Kruiver, 2022).

6.2. Aanbevelingen

Uit de bevindingen van dit onderzoek wat afgebakend is op de propaantanks, worden er ook aanbevelingen geboden op het inwinnen van windmolens en gas meet- en regel- stations.

6.2.1. Windmolens

De windmolens die dienen geïnventariseerd te worden voor het REV zijn over het algemeen groter dan propaantanks en goed herkenbaar op satellietbeelden. Ook is de ervaring tijdens dit onderzoek opgedaan tijdens de workshop van David Yu (David Yu, ESRI AI specialist, online communicatie) dat windmolens in te winnen zijn door middel van een machine learning. Dit waren namelijk de objecten die tijdens de workshop werden ingewonnen. Uit deze resultaten en kennis die opgedaan is wordt daarom de aanbeveling gedaan om windmolens te inventariseren met hetzelfde geadviseerde stappenplan als voor de propaantanks beschreven is.

6.2.2. Gas meet- en regelstations

Voor gas meet- en regelstations zal bij nader onderzoek moeten worden onderzocht of deze op basis van satellietfoto's herkend zouden kunnen worden. Wanneer dit mogelijk blijkt, wordt geadviseerd om hetzelfde stappenplan te volgen als in dit rapportage.

7. Literatuur

- Abdellaoui, R. (2022, januari). *OMGEVINGSDIENST MIDDEN EN WEST-BRABANT*. Omgevingsdienst Midden West Brabant.
- BAG/BAGv3 (*FeatureServer*). (z.d.). BAG. Geraadpleegd op 24 juni 2022, van <https://basisregistraties.arcgisonline.nl/arcgis/rest/services/BAG/BAGv3/FeatureServer>
- Bennett, L. (2018, 22 augustus). *Machine Learning in ArcGIS*. YouTube. Geraadpleegd op 23 juni 2022, van https://www.youtube.com/watch?v=duZ7jor_YrU
- Brouwer, T., & Vlot, W. (2020, 15 oktober). *Introductie tot GeoAI*. ArcGIS.Com. Geraadpleegd op 24 februari 2022, van <https://experience.arcgis.com/experience/7c07d00e2346429e90e0ae5a234dfcbf>
- Centraal Bureau voor de Statistiek. (2022, 9 juni). *Pompprijzen motorbrandstoffen; locatie tankstation, brandstofsoort*. Geraadpleegd op 24 juni 2022, van <https://www.cbs.nl/nl-nl/cijfers/detail/81567NED>
- Cotfas, L., & Diosteanu, A. (2010, november). *Evaluating accessibility in crowdsourcing GIS*. Journal of Applied Collaborative Systems. Geraadpleegd op 10 maart 2022, van <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.192.977&rep=rep1&type=pdf>
- Deindorfer, H. (2019, 13 februari). *Hands-on experience with the image classification wizard (ArcGIS pro 1.3)*. ArcGIS Blog. <https://www.esri.com/arcgis-blog/products/imagery/imagery/hands-on-experience-with-the-image-classification-wizard-arcgis-pro-1-3/>
- Detect Objects Using Deep Learning (Image Analyst)—ArcGIS Pro | Documentation*. (z.d.). ArcGISPro. Geraadpleegd op 7 juni 2022, van <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/detect-objects-using-deep-learning.htm>
- Energieleveranciers. (2022). *Regio-indeling netbeheerders gas in nederland*. Geraadpleegd op 23 juni 2022, van <https://www.energieleveranciers.nl/netbeheerders/gas>
- Esri. (z.d.-a). *Classify objects using deep learning (image analyst)—ArcGIS pro | documentation*. Geraadpleegd op 24 juni 2022, van <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/classify-objects-using-deep-learning.htm>
- Esri. (z.d.-b). *Deep learning frequently asked questions—ArcGIS pro | documentation*. <https://pro.arcgis.com/en/pro-app/2.8/help/analysis/deep-learning/deep-learning-faq.htm>
- Esri. (z.d.-c). *Deep learning software en tools installeren en opzetten binnen ArcGIS pro en ArcGIS API for python*. ArcGIS Pro. Geraadpleegd op 23 juni 2022, van <https://www.esri.nl/nl-nl/support/technische-artikelen/arcgis-pro/deep-learning-419>
- Esri. (z.d.-d). *Deep learning software en tools installeren en opzetten binnen ArcGIS pro en ArcGIS API for python*. <https://www.esri.nl/nl-nl/support/technische-artikelen/arcgis-pro/deep-learning-419>
- Esri. (z.d.-e). *Deep learning using the ArcGIS image analyst extension—ArcGIS pro | documentation*. <https://pro.arcgis.com/fr/pro-app/latest/help/analysis/image-analyst/deep-learning-in-arcgis-pro.htm>
- Esri. (z.d.-f). *Detect objects using deep learning (image analyst)—ArcGIS pro | documentation*. ArcGIS Pro. Geraadpleegd op 23 juni 2022, van <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/detect-objects-using-deep-learning.htm>
- Esri. (z.d.-g). *Detect objects using deep learning (image analyst)—ArcGIS pro | documentation*. Geraadpleegd op 23 juni 2022, van <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/detect-objects-using-deep-learning.htm>
- Esri. (z.d.-h). *Detect objects using deep learning (image analyst)—ArcGIS pro | documentation*. ArcGIS. Geraadpleegd op 23 juni 2022, van <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/detect-objects-using-deep-learning.htm>
- Esri. (z.d.-i). *Overview of image classification—ArcGIS pro | documentation*. <https://pro.arcgis.com/en/pro-app/2.8/help/analysis/image-analyst/overview-of-image-classification.htm>

- Esri. (z.d.-j). *Segmentation—ArcGIS Pro | Documentation*. <https://pro.arcgis.com/en/pro-app/2.8/help/analysis/image-analyst/segmentation.htm>
- Esri. (z.d.-k). *Segmentation—ArcGIS pro | documentation*. <https://pro.arcgis.com/en/pro-app/2.8/help/analysis/image-analyst/segmentation.htm>
- Esri. (z.d.-l). *Segmentation—ArcGIS pro | documentation*. <https://pro.arcgis.com/en/pro-app/2.8/help/analysis/image-analyst/segmentation.htm>
- Esri. (z.d.-m). *Spectral profile—ArcGIS pro | documentation*. Geraadpleegd op 23 juni 2022, van <https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/spectral-profile-chart.htm>
- Esri. (z.d.-n). *The image classification Wizard—ArcGIS pro | documentation*. Geraadpleegd op 23 juni 2022, van <https://pro.arcgis.com/fr/pro-app/latest/help/analysis/image-analyst/the-image-classification-wizard.htm#:~:text=The%20Classification%20Wizard%20is%20found,working%20in%20a%202D%20map.>
- Esri. (z.d.-o). *Train deep learning model (image analyst)—ArcGIS pro | documentation*. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/image-analyst/train-deep-learning-model.htm>
- Esri. (2016, 23 maart). *Combining rasters at different resolutions*. Esri Community. <https://community.esri.com/t5/arcgis-spatial-analyst-questions/combining-rasters-at-different-resolutions/td-p/455211>
- Esri. (2020, december). *Deep learning FAQ*. http://downloads.esri.com/resources/ImageAnalyst/DeepLearning_FAQ.pdf
- Esri. (2021, november). *ArcGIS Pro 2.9 is beschikbaar*. ArcGIS Pro. Geraadpleegd op 23 juni 2022, van <https://www.esri.nl/nl-nl/nieuws/2021/arcgis-pro-2-9-beschikbaar>
- GeoPython Conference. (2021, 10 maart). *GeoPython 2020: Detecting and analyzing solar panels in switzerland, adrian meyer* [Video]. GeoPython Conference. <https://www.youtube.com/watch?v=D4JU13snGnU>
- Greefhorst, D. (2017, 25 april). *Gegevenskwaliteit in de omgevingswet*. RIVM. https://www.rivm.nl/sites/default/files/2018-12/2017-0098_bijlage5.pdf
- Kruiver, R. (2022, 12 januari). *Wat kost een congres*. Evenementenbureau 24–7 Events. <https://247events.nl/kennisdossiers/wat-kost-een-congres/>
- Motivations of crowdsourcing contributors* (Nr. 2515–8961). (2018, april). <https://www.emerald.com/insight/content/doi/10.1108/INMR-02-2018-004/full/html>
- Nationaal georegister. (2021). *Luchtfoto 2021 7,5cm RGB ortho (voorjaarsvlucht)*. <https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/5c753916-a9fc-4f6e-8ca1-ab23d01ab61e?tab=general>
- Netbeheer Nederland. (2022). *Wie is mijn netbeheerder - Netbeheer Nederland*. NetbeheerNederland.nl. Geraadpleegd op 1 april 2022, van <https://www.netbehernederland.nl/consumenteninformatie/wie-is-mijn-netbeheerder>
- NVIDIA. (z.d.). *Installation guide windows :: CUDA toolkit documentation*. (C) Copyright 2005. <https://docs.nvidia.com/cuda/cuda-installation-guide-microsoft-windows/index.html#system-requirements>
- NVIDIA. (2022a, mei). *Installation guide windows :: CUDA toolkit documentation*. (C) Copyright 2005. <https://docs.nvidia.com/cuda/cuda-installation-guide-microsoft-windows/index.html#system-requirements>
- NVIDIA. (2022b, mei 31). *NVIDIA CUDA GPUs - compute capability*. NVIDIA Developer. <https://developer.nvidia.com/cuda-gpus#compute>
- OpenStreetMap. (2017, 30 januari). *LearnOSM*. https://learnosm.org/nl_NL/coordination/mapathon/#begin-de-planning-van-uw-mapathon-ten-minste-een-maand-v%C3%B3%C3%B3rdat-de-gebeurtenis-georganiseerd-en-voorbereid-wordt
- Ranchin, T. (1996, 6 februari). *The ARSIS method: a general solution for improving spatial resolution of images by the means of sensor fusion*. HAL-SHS - Sciences de l'Homme et de la Soci t . <https://halshs.archives-ouvertes.fr/hal-00464720/>

- Relevant netwerk externe veiligheid. (z.d.). *Signaleringskaart EV - relevant - netwerk externe veiligheid*. Relevant. Geraadpleegd op 23 juni 2022, van <https://relevant.nl/themas/dataservice-ev/signaleringskaart-ev>
- Rijksoverheid. (z.d.). *Over het register externe veiligheid drisico's (REV)*. Register externe veiligheid. Geraadpleegd op 16 februari 2022, van <https://registerexterneveiligheid.nl/overrev>
- Singh, R. (2020, 25 oktober). *Geospatial deep learning with ArcGIS - rohit singh, director of esri r&d center*. YouTube. Geraadpleegd op 23 juni 2022, van <https://www.youtube.com/watch?v=zTGqUeSr5tA>
- Stakeholdersanalyse*. (z.d.). Projectmanagementsite. Geraadpleegd op 14 maart 2022, van <https://projectmanagementsite.nl/stakeholdersanalyse/>
- Van Es, J. (2022). *Informatiemodel externe veiligheid (IMEV)*. Geonovum. Geraadpleegd op 23 juni 2022, van <https://www.geonovum.nl/geo-standaarden/informatiemodel-externe-veiligheid/informatiemodel-externe-veiligheid-imev>
- Van Til, M. (2020, 11 maart). *Waarom liever luchtfoto's uit vliegtuigen dan uit drones of satellieten?* READAR. Geraadpleegd op 16 februari 2022, van <https://readar.com/waarom-wij-voor-luchtfotos-uit-vliegtuigen-kiezen-in-plaats-van-drones-of-satellieten/>
- Watkosteenauto.nl. (2016, 12 januari). *Wat kost een auto op benzine?* <https://watkosteenauto.nl/wat-kost-een-auto-op-benzine/#:%7E:text=Een%20gemiddelde%20benzineauto%20rijdt%20ongeveer,zo'n%201%20op%2011>
- Werkzoeken.nl. (2022). *Salaris enquêteur - wat verdient een enquêteur?* Geraadpleegd op 24 juni 2022, van <https://www.werkzoeken.nl/salaris/enqueteur/>

8. Bijlagen

8.1. Bijlage datacollectie

Materiaal:

Luchtfoto HR 2021 PDOK services (Nationaal georegister, 2021)

Resolutie 8cm bij 8cm Hoge resolutie

Datum inwinning: 18-02-2021 en 11-06-2021

Inwinnende partij: Samenwerkingsverband Beeldmateriaal

Aanbieder: PDOK-services

service: WMS API

Vorm: Imagery layer

Coördinaten systeem: Amersfoort EPSG:4289

Coördinaten projectie: RD_NEW EPSG:28992

Propaantank inventarisatie OD Midden West-Brabant (Abdellaoui, 2022)

Resolutie:

Datum inwinning: 30-08-2021 en 28-01-2022

Inwinnende partij: OD West-Brabant

Aanbieder: OD West-Brabant (intern)

Service: Shapefiles

Vorm: Polygoon bestand

Coördinaten systeem: Amersfoort EPSG:4289

Coördinaten projectie: RD_NEW EPSG:28992

Tabel 5 Tabeldefinities Propaantank inventarisatie

Attributen:	Definitie:	Mogelijke invulling	Betekenis
FID	Uniek ID (PK)	Gesorteerd nummer (uniek)	nvt
Shape	Vector vorm	Polygoon / punt	nvt
ID	Classificatie	0 – 1 – 2 - 9	0 = Kruis (geen Propaantank) 1 = Vlag (Propaantank zichtbaar) 2 = Vraagteken (mogelijk propaantank niet zichtbaar)

			9 = foutieve invulling
Oppervlak	Oppervlakte van object	0 (niet berekend)	nvt

BAG V2.0 (BAG/BAGv3 (FeatureServer), z.d.)

Schaalniveau: 1:2.500

Resolutie:

Datum plaatsing: 20 sep. 2021 (laatste update: 17 mrt. 2022)

Inwinnende partij: Kadaster

Aanbieder: ESRI Nederland

Service: WFS

Vorm: Polygon geometry

Coördinaten systeem: Amersfoort EPSG:4289

Coördinaten projectie: RD_NEW EPSG:28992

Bronpagina: <https://basisregistraties.arcgisonline.nl/arcgis/rest/services/BAG/BAGv3/FeatureServer>

8.2. Bijlage datakwaliteit

Alle datasets die gebruikt worden, worden getest op data kwaliteit kenmerken die van toegevoegde waarde zijn bij het gebruik van de dataset. Tabel 6 Kwaliteitcontrole Propaantank inventarisatie OD Midden West-Brabant en Tabel 7 kwaliteitscontrole Luchtfoto's beeldmateriaal.nl geven het kenmerk weer, de gebruikte functie om te testen, de records voor en na de functie, de gebruikte attributen bij het testen, eventuele opmerkingen en een eindoordeel van de keuring wanneer nodig. De datakwaliteit kenmerken zijn gekozen aan de hand van de literatuur van het Rivm (Greefhorst, 2017)

Propaantank inventarisatie OD Midden West-Brabant

Voor de dataset kwaliteit check is gebruik gemaakt van Excel. De reden voor de keuze van het handmatig bewerken voor een datakwaliteit check is dat de dataset een shapefile is. Deze bestanden zijn moeilijker te checken en te filteren door het programma FME, omdat hierdoor de volgorde van de index (een geclassificeerd cijfer per rij dat een koppeling vormt tussen de shape file en geo database file) verandert. Dit zou ervoor kunnen zorgen dat gegevens uit de geo database verkeerd gekoppeld worden aan de shapefile.

Tabel 6 Kwaliteitcontrole Propaantank inventarisatie OD Midden West-Brabant

Data kwaliteit kenmerk	Gebruikte functie	Records voor	Records na	Gebruikte attributen	Opmerking	Keuring
Tijdigheid	Geen	6.153	6.153	Geen	De data is handmatig ingewonnen tussen: 30-08-2021 en 28-01-2022 Hierbij is ortho HR luchtfoto's gebruikt van: 18-02-2021 en 11-06-2021	Dataset is samengevoegd met de meest recente luchtfoto's de tijdigheid van de dataset is daarom van goede kwaliteit en relevant.
Volledigheid	Select layer by attribute (*Esri)	6.153	6.153	ID (enige relevante attribuut)	Gecheckt op de aanwezigheid van de ID waardes met 0, 1, 2 of 9.	Alle rijen hebben een ID waarde toegekend van 0, 1, 2 of 9. De data is volledig op ID.
uniekheid	Delete Identical (*Esri)	6.153	6.126	Geometry (Geo database)	De data is verwijderd bij identieke overlap van de features.	Na het verwijderen van de duplicaten zijn de gegevens volledig uniek.

					27 duplicaten zijn aangetroffen en verwijderd.	
validiteit	Select layer by attribute (*Esri)	6.126	6.123	ID (enige relevante attribuut voor de bewerking)	Voor het attribuut ID zijn 3 verschillen vaste waarde die ingevuld kunnen worden zoals in de metadata uitgelegd. De rijen zijn gecheckt op het hebben van een waarde 0, 1 of 2. Bij dit proces zijn 3 records verwijderd.	Na het verwijderen van de afwijkende waarde is de data goed gekeurd op validiteit.

Tabel 7 kwaliteitscontrole Luchtfoto's beeldmateriaal.nl

Data kwaliteit kenmerk	Gebruikte functie	Records voor	Records na	Gebruikte attributen	Opmerking	Keuring
Tijdigheid	Geen	n.v.t	n.v.t	geen	De data is ingewonnen op 18/02/2021 en 11/06/2021	De data is de meest recente raster data/luchtfoto's van Nederland en dus is de tijdigheid goed. De data is ingewonnen in een bladloos seizoen en daardoor geschikt voor de analyse
Volledigheid	Geen	n.v.t	n.v.t	geen	Nederland vlak dekkende data	De data is Nederland vlak dekkend en daarom geschikt en volledig voor een dergelijk analyse.

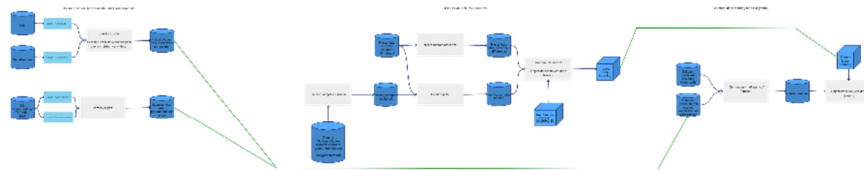
8.3. Bijlage Deep learning proces

Figuur 9 Totaal proces Deep learning - Filteren gebied -> Deep learning training -> Deep learning uitvoeren is een workflow voor het deep learning proces, omdat de afbeelding in zijn geheel niet goed te lezen is, is de afbeelding in drie delen opgedeeld te zien in Figuur 10 Deep learning proces - Filtering gebied tot en met Figuur 12 Deep learning proces - Deep learning uitvoeren

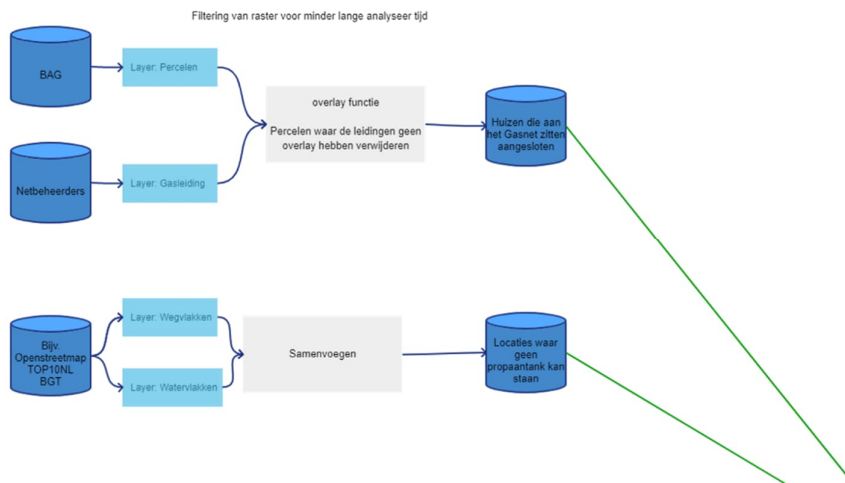
Figuur 10 Deep learning proces - Filtering gebiedvisualiseert het proces van het filteren van de input rasters. Dit zijn de rasters die dienen geanalyseerd te worden. Het filteren van deze data is benodigd om de analyse tijd van het deep learning proces te verkleinen.

In Figuur 11 Deep learning proces - Deep learning training is het proces opgenomen van het maken en trainen van een deep learning model.

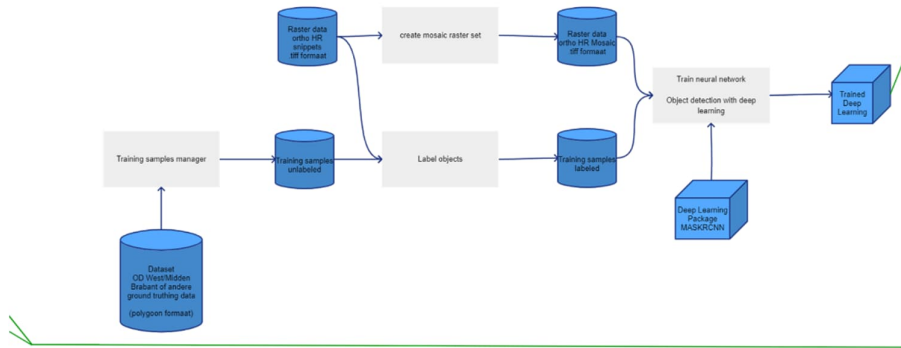
In Figuur 12 Deep learning proces - Deep learning uitvoeren is het uitvoeren van het deep learning proces opgenomen.



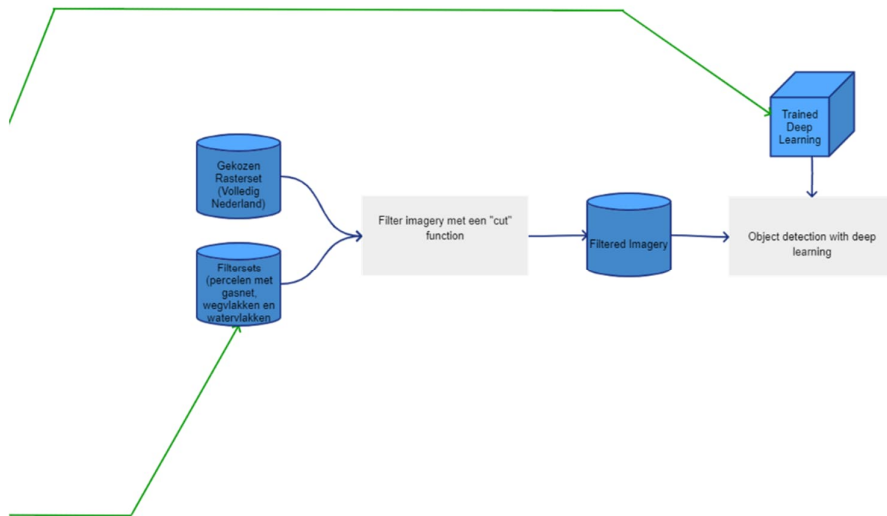
Figuur 9 Totaal proces Deep learning - Filteren gebied -> Deep learning training -> Deep learning uitvoeren



Figuur 10 Deep learning proces - Filtering gebied



Figuur 11 Deep learning proces - Deep learning training



Figuur 12 Deep learning proces - Deep learning uitvoeren

8.4. Bijlage Deep learning analyse beschrijving

In de analyse beschrijving worden een aantal afkorting gebruikt die hieronder kort uitgelegd worden:

GPU = Grafische kaart (graphics processing unit)

CPU = Processor (Central processing unit)

De gebruikte software bij deze analyse zijn:

ArcGIS Pro 2.9

Deep learning package essentials 2.8

In de analyse beschrijving wordt gebruik gemaakt van nummering en "*" opsommingstekens. De nummering geeft hierbij een commando aan en de opsommingstekens geven de parametrisering van een functie aan.

In de analyse beschrijving worden drie verschillende analyses beschreven. Hieronder staan de analyses vormen kort uitgelegd:

Classificatie door middel van deep learning = Het classificeren van objecten door middel van een deep learning model, gebaseerd op de beschikbare spectrale banden gevisualiseerd per pixel.

Spectrale analyse = Het analyseren van gekozen objecten en de waardes per pixel die deze objecten hebben. Dit resulteert in een grafiek met de waardes per object, overeenkomsten tussen objecten en hun spectrale banden worden hierin gevisualiseerd.

Object herkenning door middel van deep learning = Het herkennen van objecten door middel van een deep learning model, gebaseerd op gelabelde afbeeldingen wordt een model getraind om objecten te herkennen en lokaliseren.

Tijdens het uitvoeren van de methode is er gebruik gemaakt van een externe harde schijf: "Toshiba 2tb SSD". Op deze harde schijf zijn alle deep learning modellen, gebruikte rasters, getrainde samples en training schema's opgeslagen. Tijdens de analyse beschrijving wordt er af en toe gerefereerd naar deze harde schijf en de opgeslagen bestanden die op deze harde schijf staan. Ook word er aangegeven wat te doen als er geen beschikking is over deze harde schijf.

8.4.1. Stappen voor het gebruik van deep learning

Voor het gebruik van machine learning of deep learning binnen ArcGIS Pro is het benodigd om de deep learning of machine learning libraries te downloaden. Deze libraries bieden de benodigde code en bestanden om deep learning in ArcGIS Pro mogelijk te maken. Ook is het benodigd om de CUDA software en visual studio te downloaden zodat de grafische kaart (GPU) gebruikt kan worden voor het uitvoeren van de deep learning modellen. De CUDA software is een software architectuur die het mogelijk maakt om algoritmes uit te voeren op de grafische kaart. Het gebruik van de GPU zorgt voor een betere performance dan met het gebruik van de CPU. Om deze software te downloaden kunnen voor beide (CUDA en Deep learning libraries) de stappen op de volgende tutorial pagina gevolgd worden:

<https://www.esri.nl/nl-nl/support/technische-artikelen/arcgis-pro/deep-learning-419>

Voor de deep learning classificatie is het mogelijk het raster/luchtfoto die gebruikt wordt te segmenteren voordat de analyse gestart wordt. Het segmenteren van een raster bestand kan gedaan worden door:

1. *Selecteer het betreffende raster/luchtfoto*
2. *Klik op de "imagery tab"*
3. *Klik op de "classification tools"*
4. *Klik op de "segmentation" functie*

Voor de parametrisering van deze functie wordt verwezen naar de beschrijving van Figuur 15 Image classificatie wizard - Segmentatie

8.4.2. Deep learning classificatie analyse beschrijving

Voor een deep learning model dienen er een aantal stappen uitgevoerd te worden. Alle stappen voor een deep learning classification model kunnen 1 vóór 1 uitgevoerd worden door middel van de "classification wizard" van ArcGIS Pro.

De workflow van dit proces bestaat uit 3 hoofd stappen:



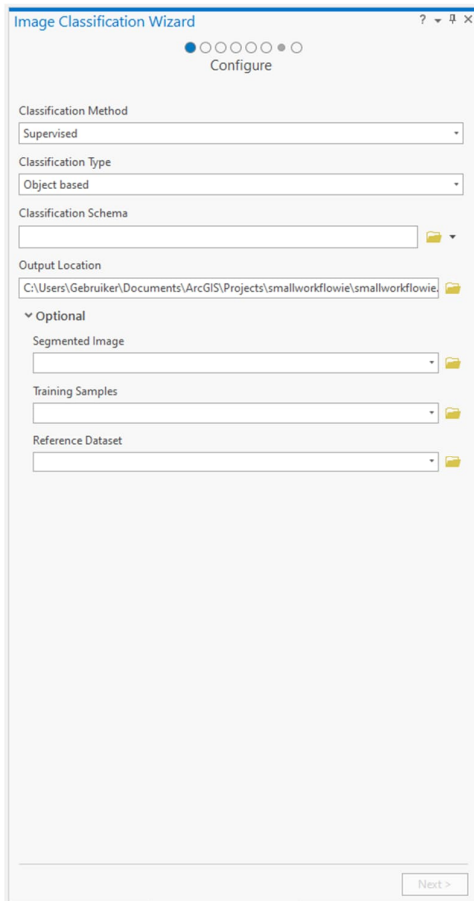
Figuur 13 Workflow machine learning in ArcGIS Pro (Esri, z.d.-b)

Stap 1 Training data tekenen

Stap 1 bestaat uit het verzamelen van training data en het labelen van training data. Dit kan op verschillende manieren door al ingewonnen data van propaantanks te gebruiken of door zelf training data in te tekenen. Om training data te maken wordt er gebruik gemaakt van de "training sample manager".

1. *Selecteer de imagery layer in de "content pane" die gebruikt wordt om de training samples te maken.*
2. *Klik op de "imagery" tab boven in het scherm.*
3. *Klik op de "image classification wizard" in de "image classification" groep.*

Een scherm opent dat het menu voor de classification wizard is, Figuur 14 Image classificatie wizard - Configuratie weergeeft dit menu.



Figuur 14 Image classificatie wizard - Configuratie

- Als **classification method** kan de supervised methode blijven staan. De reden hiervoor is dat we zelf willen bepalen welke klassen geclassificeerd worden in luchtfoto's
- Als **classification type** kan de Object based type blijven staan. Deze methode geeft een meer realistische weergave van de objecten in een segmentatie beeld (Esri, z.d.-f)
- Als **classification schema** kan het schema toegevoegd worden dat op de harde schijf en de bijbehorende map staat (Toshiba externe harde schijf -> trainingmodel -> trainingsamples_unlabeled -> trainingschema.ecs).
Zonder de bijbehorende externe harde schijf kan het classificatie schema in latere stappen zelf gemaakt worden door de stappen en intro te volgen bij Figuur 16 Image Classificatie wizard - Training sample manager.
- Als **output locatie** kies een logische map en naam voor de classificatie.

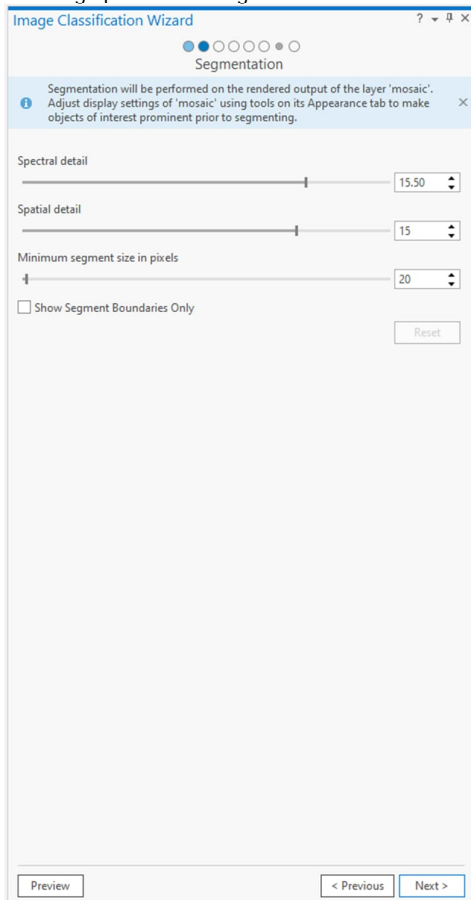
Optionele parameters (bij gebruik van de toshiba externe harde schijf, het gebruik van een eerder gesegmenteerd beeld of het gebruik van een ground-truthing dataset)

- Als **segmented image** kan eventueel een eerder gesegmenteerd beeld ingevoerd worden van de luchtfoto die als input gebruikt wordt. Anders word dit beeld later in de image classification wizard gesegmenteerd (er wordt aangeraden van te voren de input luchtfoto's te segmenteren door middel van de "segmentation" functie).

- Als **trainingsamples** kunnen de trainingsamples die ingewonnen zijn gebruikt worden (Toshiba harde schijf -> trainingmodel -> trainingsamples_unlabeled -> [trainingsampledata.shp](#)). Verder in de classification wizard kunnen meerdere samples ingetekend worden.
- Als **reference dataset** kan er een dataset ingevuld worden met referentie data van een bekende locatie van de propaantanks die geïdentificeerd zijn (ground-truthing).

Segmentatie van de beelden

Als tweede stap van de classification wizard word er een segmentatie functie gegeven in Figuur 15 Image classificatie wizard - Segmentatie. Deze stap is ook apart van te voren uit te voeren, dit wordt aangeraden wegens het lange proces dat segmentatie kan duren.



Figuur 15 Image classificatie wizard - Segmentatie

- Als **spectral detail** wordt er aangeraden een hogere waarde te selecteren. De default hierbij is een goede waarde aangezien de schaal van 1 tot 20 loopt. De reden voor een hogere waarde is dat er verschillende objecten qua pixelwaarde lijken op dat van de propaantanks, om hier beter onderscheidt in te kunnen maken is het gebruikelijk een hogere spectral detail waarde toe te kennen (Esri, z.d.-g).
- Als **spatial detail** waarde wordt er aangeraden een hogere waarde te selecteren. De default waarde is hierbij een goede waarde aangezien de schaal van 1 tot 20 loopt. De reden voor een hogere waarde is dat de propaantanks een relatief klein object zijn en er vaak verschillende objecten omheen staan zoals bomen, bosjes of huizen. Om goed onderscheidt te kunnen maken tussen dicht bij elkaar liggende objecten (Esri, z.d.-h).
- De **minimum segment size in pixels** is de hoeveelheid pixels er per segmentatie stuk gegroepeerd worden, omdat propaantanks relatief kleine objecten (Steragas, 2022) zijn en in een 8 bij 8 centimeter

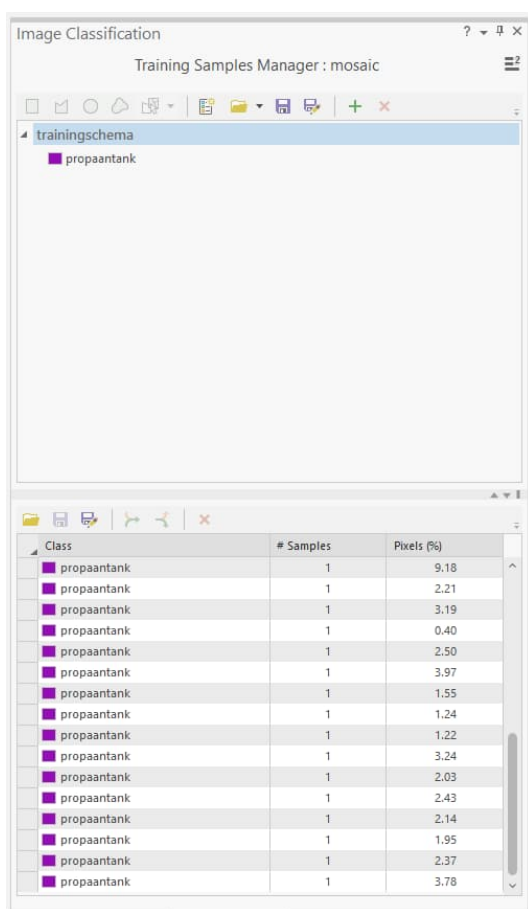
raster weinig pixels bevatten kan de waarde omlaag gebracht worden. Een waarde van 10 pixels per segment is een goede waarde voor een relatief klein object (Esri, z.d.-i).

Intekenen van de trainingsamples

Op Figuur 16 Image Classificatie wizard - Training sample manager is de training sample manager afgebeeld. De sample manager opent automatisch met een "model schema" waarin verschillende klassen zijn opgenomen. Deze klassen kunnen 1 voor 1 verwijderd worden door een klasse te selecteren en op het kruisje te klikken. Vervolgens kunnen er 2 klassen toegevoegd worden, namelijk "niet propaantank" en "propaantank". De reden voor twee klassen is dat het model anders geen onderscheid kan maken in de verschillende pixel waardes. Met één klasse zouden alle pixels geclassificeerd worden als propaantank.

Om een object in te tekenen dient er eerst de juiste klasse geselecteerd te worden en vervolgens de manier van intekenen; polygoon, vierkant, cirkel etc.

1. Zorg voor een juist schema met de benodigde klassen (zoals hierboven beschreven).
2. Klik op de klasse waar je een sample van wilt intekenen.
3. Klik op de tool die je wilt gebruiken om de training sample in te tekenen.



Figuur 16 Image Classificatie wizard - Training sample manager

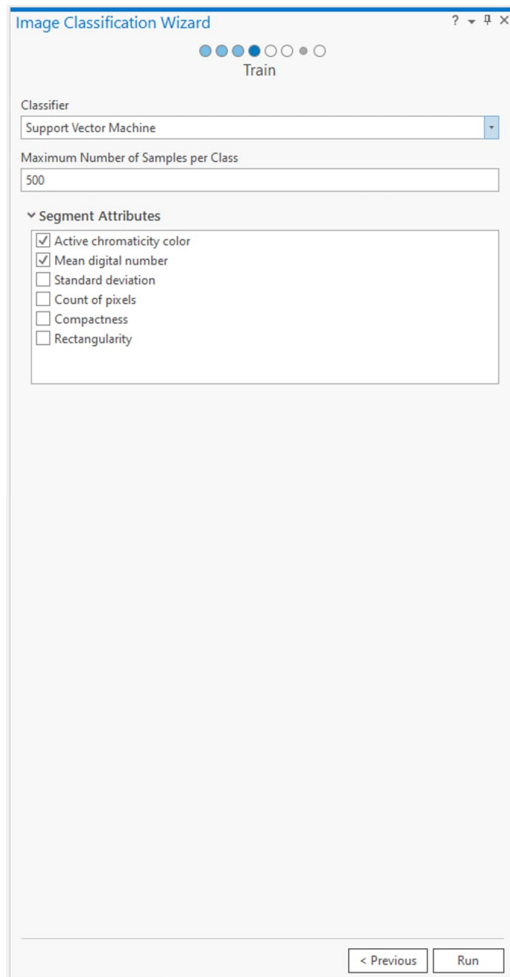
Wanneer er genoeg training samples ingetekend zijn (afhankelijk van de betrouwbaarheid en kwaliteit van je model) kan het bestand opgeslagen worden als trainingsample bestand. Dit is een shapefile met de betreffende ingetekende training samples.

Het minimaal aantal training samples voor een model is bij elk object verschillend en afhankelijk van de betrouwbaarheid graad en kwaliteit van je model waarnaar gestreefd wordt. Over het algemeen wordt een

minimum van 150 samples genomen (Lee, 2021), maar omdat samples dubbel gebruikt kunnen worden door het toepassen van een halvering van de "stride size" ten opzichte van de "tile size" is een minimum van 75 samples genoeg (de werking van deze dubbele samples en de "stride en tile size" wordt duidelijk uitgelegd bij Figuur 23 Geoprocessing - Export Training Data For Deep Learning).

Stap 2 Train het model

Na het intekenen van de training samples kan er op "next" gedrukt worden en opent zich het "train" menu
 Figuur 17 Image classification wizard - Train.



Figuur 17 Image classification wizard - Train

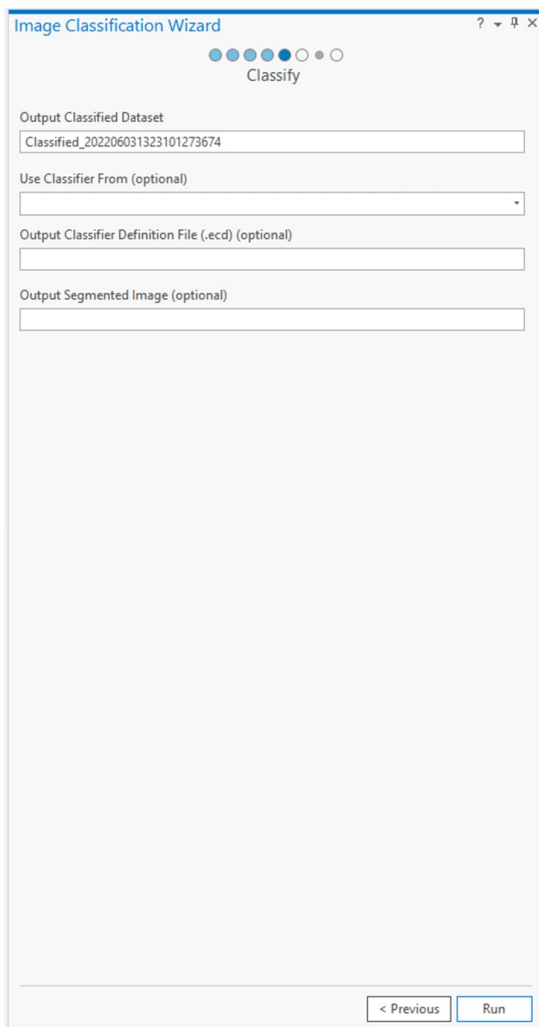
- Als **Classifier** mag de parameter op "support vector machine" blijven staan. Dit is een relatief nieuwere classifier en wordt door vele onderzoekers gebruikt (Deindorfer, 2019). Ook is deze classifier minder vatbaar voor ruis, oneven aantal training samples of gecorrleerde banden.
- De **maximum number of samples per class** mag op default blijven. Tenzij er meer samples ingetekend zijn. Dit zijn 500 samples.
- De **segment attributes** Zijn de attribuut waardes waar het model op getraind wordt om te herkennen in de training samples attribuut waardes. De belangrijkste hierbij zijn de "chromaticity color" en "mean digital number", omdat deze waarde over het algemeen hetzelfde blijven voor een propaantank. De grootte van de propaantank, vorm en aantal pixels kan per zichtbare propaantank veranderen door bijvoorbeeld een overkapping, boom of andere objecten die een gedeelte van de propaantank verbergen.

Als laatste stap dient het gesegmenteerde beeld gecontroleerd te worden. Dit kan gedaan worden door het gebruikte input raster en het gesegmenteerde resultaat te analyseren door middel van de "L" knop. Door het indrukken van de "L" knop wordt het gesegmenteerde beeld transparant en kunnen objecten van de luchtfoto vergeleken worden met het segmentatie beeld. Wanneer objecten zoals de propaantank duidelijk visueel zijn in

de segmentatie beelden en er niet te veel vertekeningen zijn kan het segmentatie beeld gebruikt worden in de volgende functie.

Classify

In Figuur 18 Image classification wizard – Classify Kan het model uitgevoerd worden om te gaan classificeren.



Figuur 18 Image classification wizard – Classify

- Als **Output classified dataset** Kies een duidelijke naam en omgeving om de geclassificeerde data laag op te slaan.
- Als **Use classifier From (optional)** dient de classificatie uit de vorige stap: “train het model” ingevoerd te worden.
- Als **Output classifier definition file (.ecd) (optional)** kies een duidelijke naam en omgeving om het geconverteerde deep learning model op te slaan. Dit wordt het door jou gemaakt deep learning classificatie model.

- Als **Output segmented image (optional)** kies een duidelijke naam en omgeving om de geclassificeerde data laag op te slaan (deze stap kan overgeslagen worden als je voor de bewerkingen al een gesegmenteerde data laag hebt gemaakt).

Merge classes

De merge classes functie kan overgeslagen worden aangezien er geen klasse zijn die gegroepeerd dienen te worden.

Reclassifier

De reclassifier functie kan gebruikt worden om eventuele fout geclassificeerde punten in de output dataset te her classificeren. Hierdoor kan het model gecorrigeerd worden.

8.4.3. Analyseren met een spectraal profiel (analyse beschrijving proces)

Om een raster te analyseren op de pixelwaardes (spectrale profiel) van de banden die het raster bezit, dient er een grafiek opgesteld te worden. De reden voor het uitvoeren van een spectrale analyse zijn de valse positieven die uit het classificatie met deep learning model zijn gekomen. Door een spectrale analyse kan onderzocht worden of er veel objecten zijn met een zelfde pixel waarde.

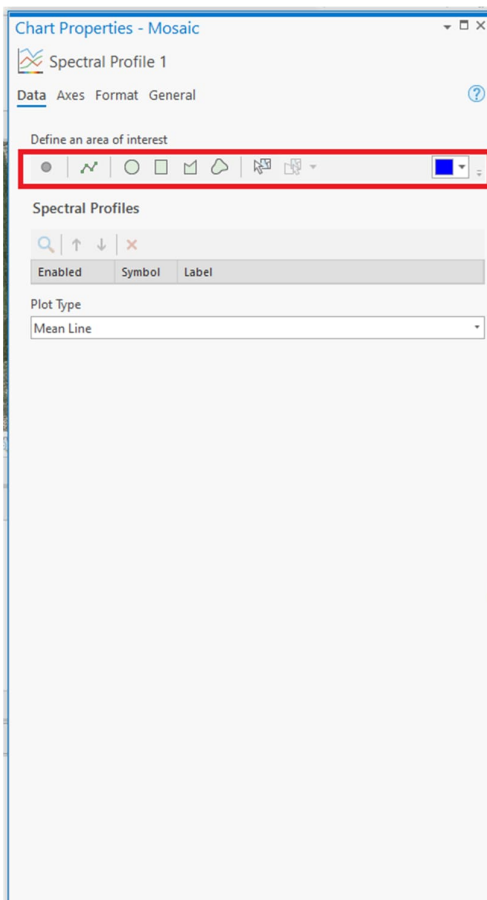
Stap 1.

rechtermuis knop op de te analyseren rasterdata (luchtfoto's) en kies de optie "spectral profile" onder "create chart"

Er zal zich een "chart eigenschappen paneel" openen *Figuur 19 Chart eigenschappen - Mosaic*.

Stap 2.

Om vervolgens de juiste objecten te analyseren gebruik een van de "select area of interest tool". Dit kan een polygoon, lijn of vector punt zijn. Zoals in *Figuur 19 Chart eigenschappen - Mosaic* eigenschappen.



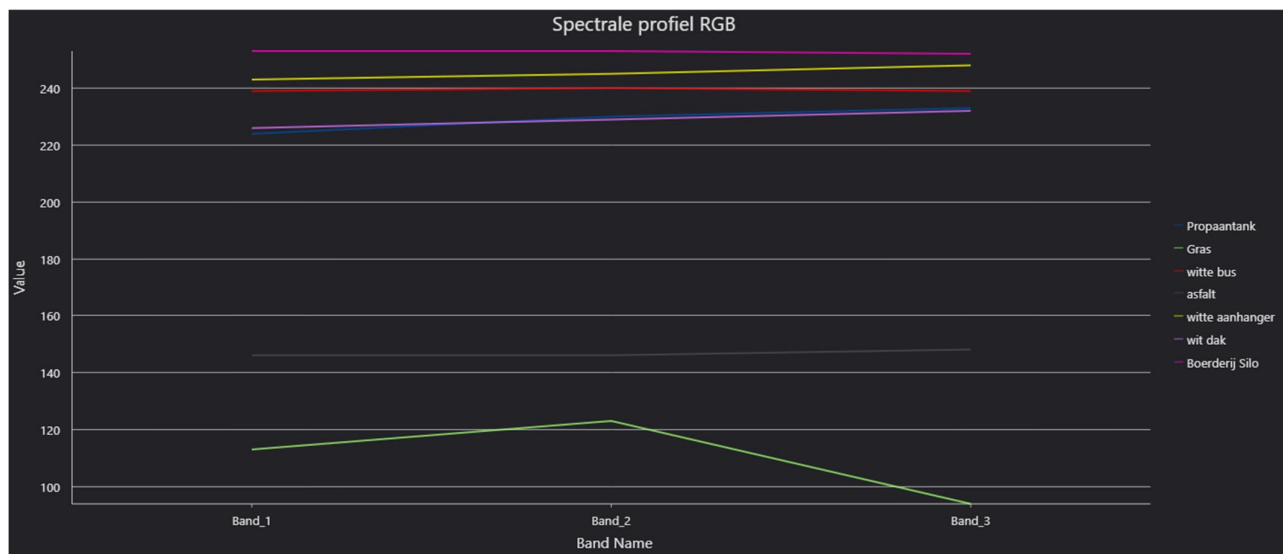
Figuur 19 Chart eigenschappen - Mosaic

Stap 3

Gebruik de area of interest tools en klik op de objecten die je wilt analyseren (in dit geval objecten die binnen de input data vallen en vergelijkbaar zijn met propaantanks. Geef elk object een passende naam voor referentie in de grafiek.

Het resultaat is een spectraal profiel Figuur 20 Spectraal profiel RGB.

In deze grafiek kan per "band" gezien worden welke objecten veel overeenkomsten hebben in pixelwaardes. Dit zijn de kleurwaardes die per band gereflecteerd worden.



Figuur 20 Spectraal profiel RGB

8.4.4. Deep learning object detection analyse beschrijving

Voor een deep learning model dienen er een aantal stappen uitgevoerd te worden. Voor een object herkennings model zijn deze stappen anders dan dat van een classificatie model.

De workflow van dit proces bestaat uit 3 hoofd stappen die afgebeeld zijn in Figuur 13 Workflow machine learning in ArcGIS Pro (Esri, z.d.-b).

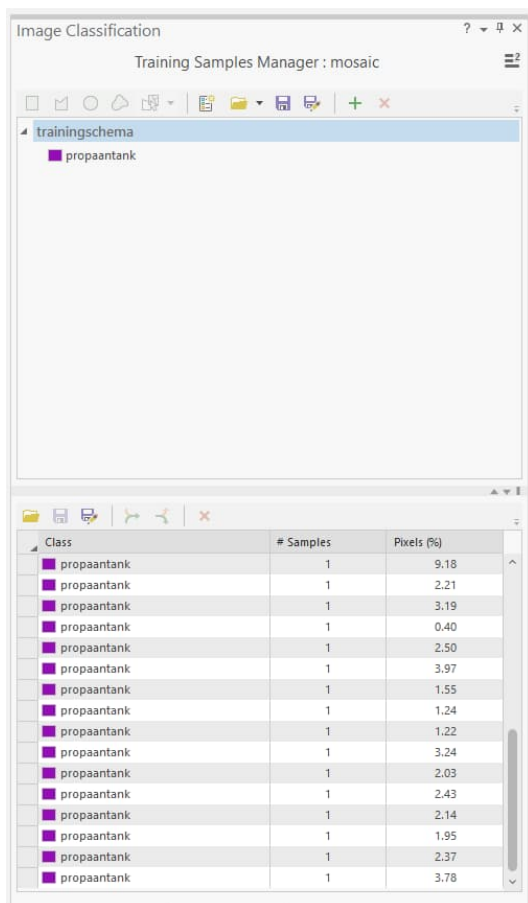
Stap 1 Training data tekenen en labelen

Stap 1 bestaat uit het verzamelen van training data en het labelen van training data. Dit kan op verschillende manieren door al ingewonnen data van propaantanks te gebruiken of door zelf training data in te tekenen. Om training data te maken wordt er gebruik gemaakt van de "training sample manager". Deze tool is te vinden onder de "imagery" tab op ArcGIS Pro na het selecteren van een imagery layer in de "content pane".

1. *Selecteer de imagery layer in de "content pane" die gebruikt wordt om de training samples te maken.*
2. *Klik op de "imagery" tab boven in het scherm.*
3. *Klik op de "classification tools" in de "image classification" groep.*
4. *Klik op de "training sample manager"*

Op Figuur 21 Image classificatie - Training sample manager is de sample manager afgebeeld. De sample manager opent automatisch met een "model schema" waarin verschillende klassen zijn opgenomen. Deze klassen kunnen 1 voor 1 verwijderd worden door een klasse te selecteren en op het kruisje te klikken. Vervolgens kan er 1 klasse toegevoegd worden namelijk "propaantank". Om een object in te tekenen dient er eerst de juiste klasse geselecteerd te worden en vervolgens de manier van intekenen; polygoon, vierkant, cirkel etc.

1. Zorg voor een juist schema met de benodigde klasse "propaantank" zoals hierboven beschreven.
2. Klik op de klasse waar je een sample van wilt intekenen.
3. Klik op de tool die je wilt gebruiken om de training sample in te tekenen.



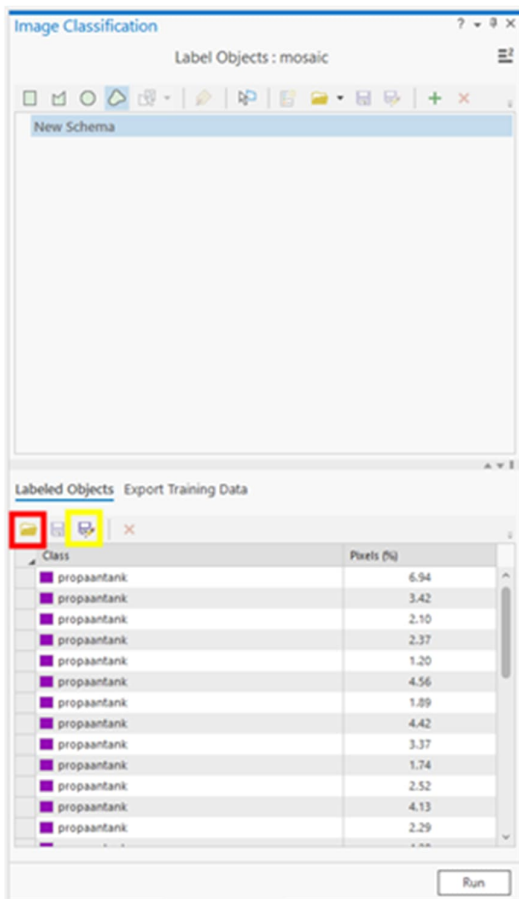
Figuur 21 Image classificatie - Training sample manager

Wanneer er genoeg training samples ingetekend zijn (afhankelijk van de betrouwbaarheid en kwaliteit van je model) kan het bestand opgeslagen worden als trainingsample bestand. Dit is een shapefile met de betreffende ingetekende training samples.

training samples labelen

Na het intekenen van de training samples kan het opgeslagen bestand ingevoerd worden bij de functie "label objects for deep learning". Bij het labelen van de training samples worden "bounding boxes" gemaakt van de ingetekende training samples en de gebruikte luchtfoto's. Ook word er een map opgeslagen met de foto's die binnen deze bounding boxes vallen.

Open de "label objects for deeplearning" functie, een scherm opent zich te zien op Figuur 22 Image classificatie - Label objects.



Figuur 22 Image classificatie - Label objects

Klik op de open folder knop (aangegeven in met het rode vierkant) en laad de eerder opgeslagen shapefile in uit de voorgaande stappen van het training samples maken.

Klik op de disk knop (aangegeven met het gele vierkant) om een locatie toe te wijzen waar de gelabelde objecten kunnen worden opgeslagen. De output van dit bestand is een map met meerdere bestanden en mappen.

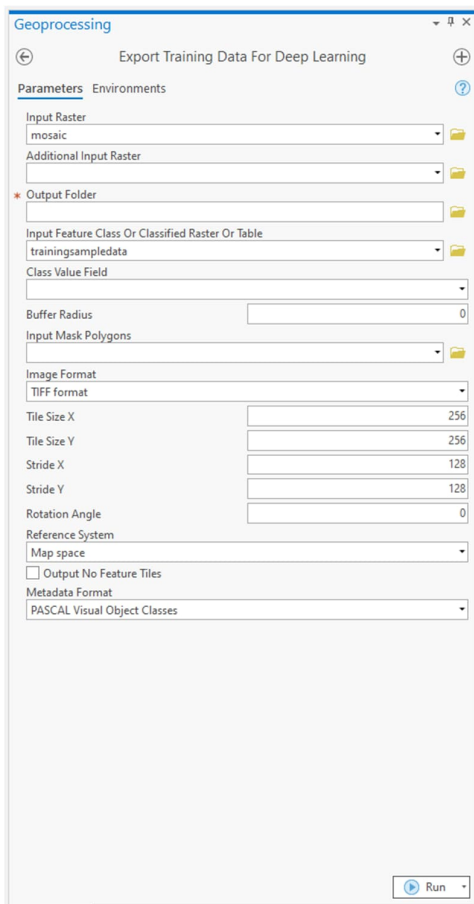
Klik op "run" om de objecten te labelen.

Stap 2 deep learning model trainen

Het trainen van een deep learning model bestaat uit twee stappen: het exporteren van de training data en het trainen van een model. Keuzes van bepaalde parameters zijn met name gebaseerd op de expertise van de heer "David Yu", data analyst bij Esri en deep learning expert.

Export training data for deep learning model

Door middel van de export training data for deep learning model functie wordt de gelabelde data klaar gemaakt om tot deep learning model omgevormd te worden.



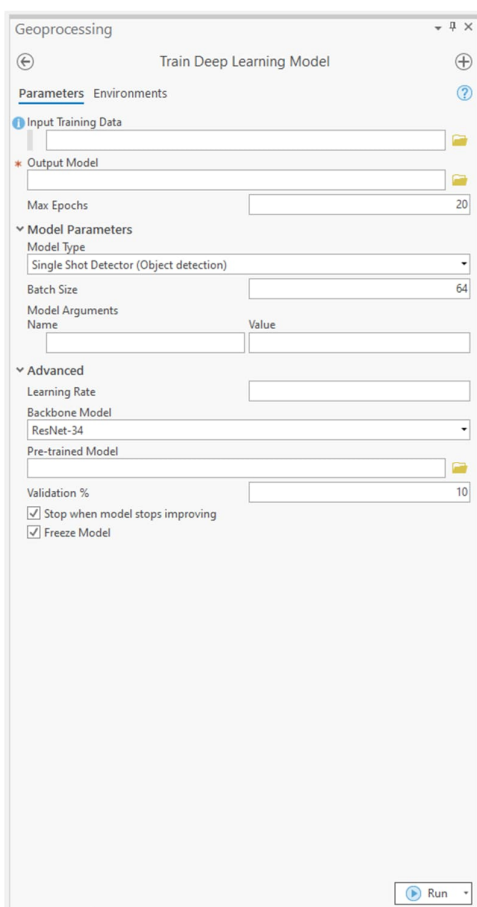
Figuur 23 Geoprocessing - Export Training Data For Deep Learning

Figuur 23 Geoprocessing - Export Training Data For Deep Learning laat het menu scherm van de functie zien.

- Het **input raster** is het raster dat gebruikt wordt om de training samples op te baseren. Kies hiervoor het raster dat gebruikt is om de training samples in te tekenen.
- Geef het bestand een juiste output folder.
- De **input feature class** zijn de gelabelde training samples.
- De **class value field** wordt niet gebruikt.
- De **tile size** is de grootte van de image tiles die gemaakt worden tijdens het exporteren van het input raster. Deze grootte kan op default blijven.
- De **stride size** is de afstand die vanuit het begin punt van "tiling" genomen wordt om de volgende tile te maken. Deze grootte kan ook op default blijven. Hierdoor komt er overlap in de tegels (stride size is kleiner dan de tile size). Door de stride size kleiner te houden dan de tile size worden er meer training samples in kaart gebracht (dubbele samples) en dit zal zorgen voor een beter model.
- Het **metadata format** wordt aangeraden om gebruik te maken van het PASCAL VOC format bij object herkenning (DAVID YU, Esri deep learning specialist, online interactieve workshop).

Train deep learning model

Na het trainen en labelen van de samples dient er een deep learning model opgezet te worden. Dit wordt gedaan met de “train deep learning model” functie.



Figuur 24 Geoprocessing - Train Deep Learning Model

Figuur 24 Geoprocessing - Train Deep Learning Model weergeeft het functie menu voor de functie “train deep learning model”.

- Voor de **input training data** word de geëxporteerde training data gebruikt uit de vorige functie.
- als **output model** dient er een passende naam en omgeving gemaakt te worden voor het deep learning model.
- De **max epochs** parameter is de waarde hoe vaak de volledige dataset door het neurale netwerk word gedaan. 1 epoch is dus de gehele dataset 1 maal door het neurale netwerk laten gaan. Deze waarde mag op 100 gezet worden. Door de “stop when model stops improving” parameter zal het model stoppen wanneer het niet meer verbeterd wordt. Door een hoog aantal epochs te gebruiken in combinatie met de stop model parameter zoekt de functie het juiste aantal epochs (DAVID YU, Esri deep learning specialist, online interactieve workshop)..
- Als **model type** dient er de single shot detector gebruikt te worden (DAVID YU, Esri deep learning specialist, online interactieve workshop)., omdat de uitvoering van dit model het lichtste is van de verschillende “object detection” modellen. Voor betere prestatie van het model kan er gekozen worden

voor de FasterRCNN en de MaskRCNN modellen, maar hierdoor zal het model wel langer duren tijdens het trainen en uitvoeren (DAVID YU, Esri deep learning specialist, online interactieve workshop). Daarom wordt dit alleen aangeraden wanneer een pc met genoeg rekenkracht gebruikt wordt.

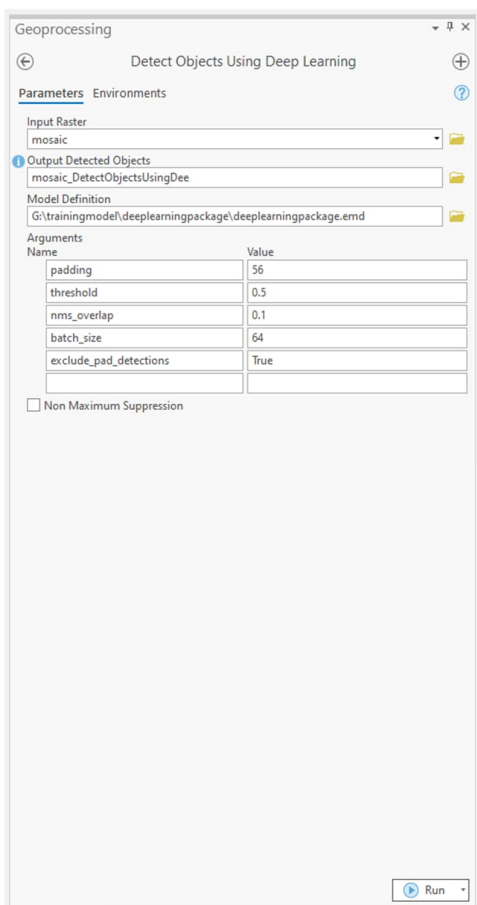
- De **batchsize** is de waarde in hoeveel sets de gehele dataset wordt verdeeld om het model te trainen. Deze waarde kan op default blijven zolang er genoeg rekenkracht is. Wanneer er niet genoeg rekenkracht is kan de waarde verhoogd worden. Dit zorgt ervoor dat er meerdere sets gemaakt worden en de betreffende sets daardoor kleiner zijn. De GPU of CPU hoeft hierdoor minder rekenkracht te gebruiken om een set of batch uit te voeren.
- De **advanced parameters** kunnen op default blijven, de functie zoekt de juiste waardes automatisch bij de ingevoerde data (DAVID YU, Esri deep learning specialist, online interactieve workshop). Ook wordt de juiste learning rate gezocht door de eerder benoemde max epoch waarde en de "stop when model stops improving"
Alleen de "freeze model" parameter mag "unchecked" gezet worden. Dit zorgt ervoor dat de backbone model lagen niet "bevroren" worden. Dit zorgt voor een langere proces duur, maar wel voor een beter model (Esri, z.d.-l).

Stap 3 Gebruik het deep learning model

De laatste stap in het proces is het uitvoeren van het deep learning model.

Na het uitvoeren van het deep learning model kan het model geëvalueerd worden op de prestatie van het model.

Bijvoorbeeld door middel van een nieuwe luchtfoto met een data laag van de aanwezige propaantanks.

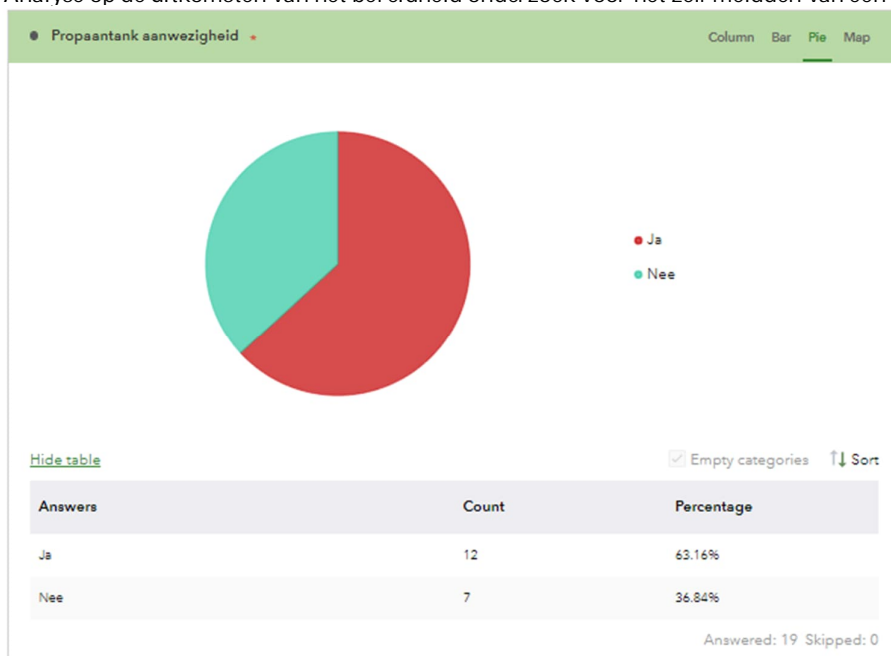


Figuur 25 Geoprocessing - Detect objects using deep learning

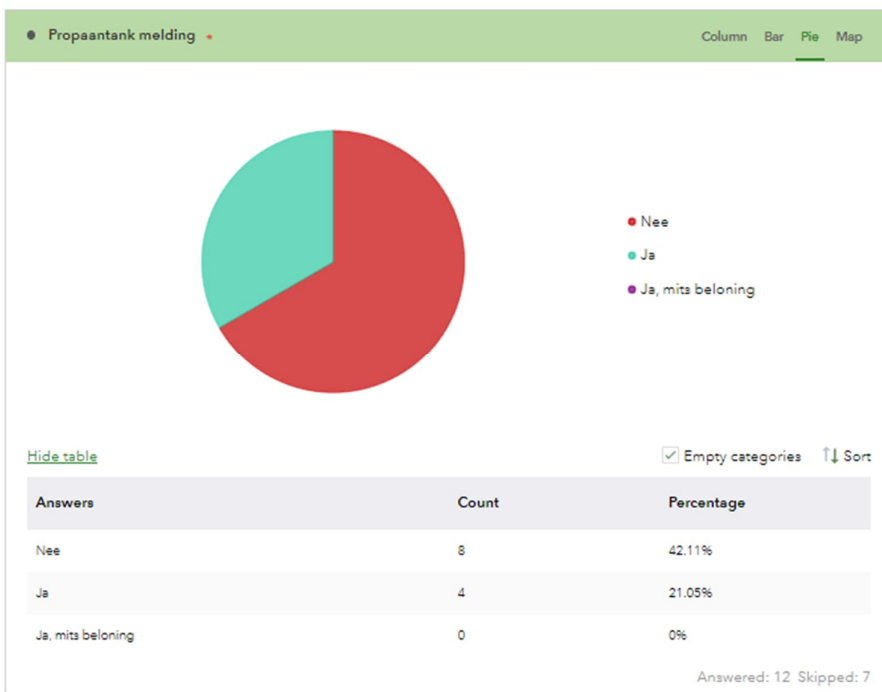
- Voor het **input raster** voeg het raster (luchtfoto) in dat geanalyseerd dient te worden.
- Voor de **output** geef een logische naam en locatie.
- Voor de **model definition** geef het deep learning model op dat in de vorige stap is gemaakt.
- Klik de **non maximum suppression** knop aan, deze functie zorgt ervoor dat duplicaten, in dit geval dubbele propaantanks, niet in de output worden meegenomen.

8.5. Bijlage Bereidheid melden propaantank enquête

Analyse op de uitkomsten van het bereidheid onderzoek voor het zelf melden van een propaantank:



Figuur 26 Bereidheid melden propaantank enquête, vraag: "Bent u bewoner van een perceel met een vaste propaantank?"



Figuur 27 Bereidheid melden propaantank enquête, vraag: "Zou u bereid zijn uw propaantank locatie te melden bij uw gemeente?"



Figuur 28 Bereidheid melden propaantank enquête, vraag: "Wat is de voornaamste reden voor het niet melden van uw propaantank?"

Anders namelijk... - Motivatie niet melden propaantank

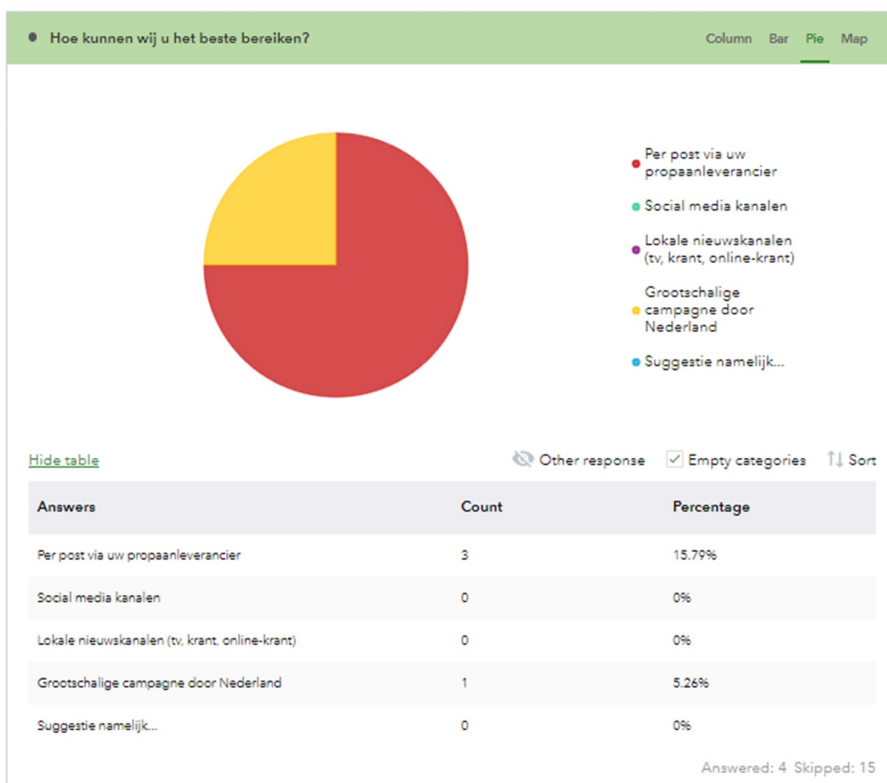
Geen urgentie niet verplicht

Geen zin in de consequenties

Als alle tanks in kaart zijn gebracht, komt er extra regelgeving, extra keuringen, of zelf een verbod op. De overheid is niet te vertrouwen met deze informatie.

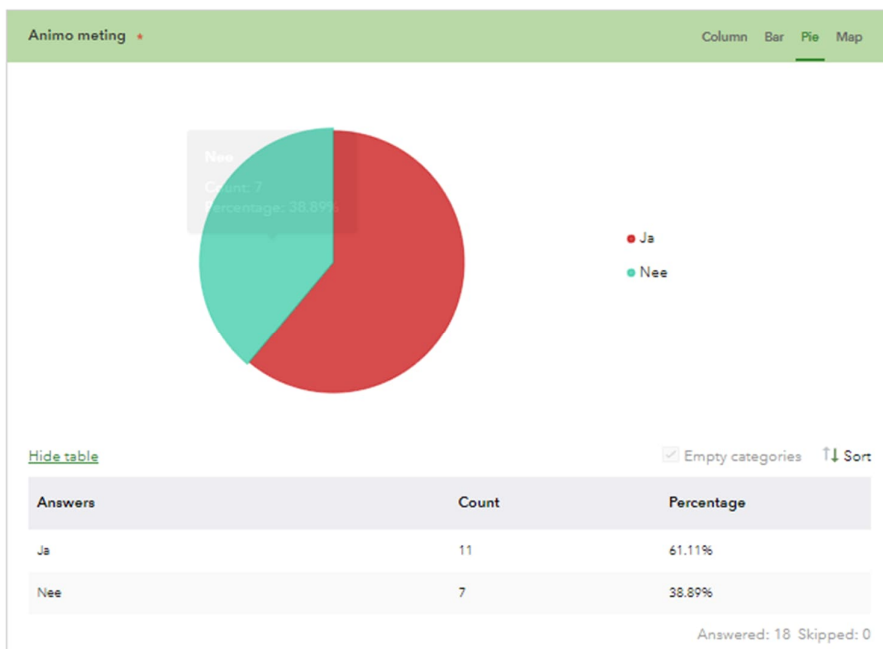
Gemeenten melden andersom ook weinig tot niets

Figuur 29 Bereidheid melden propaantank enquête, vraag: "Wat is de voornaamste reden voor het niet melden van uw propaantank? - resultaat op anders, namelijk..."

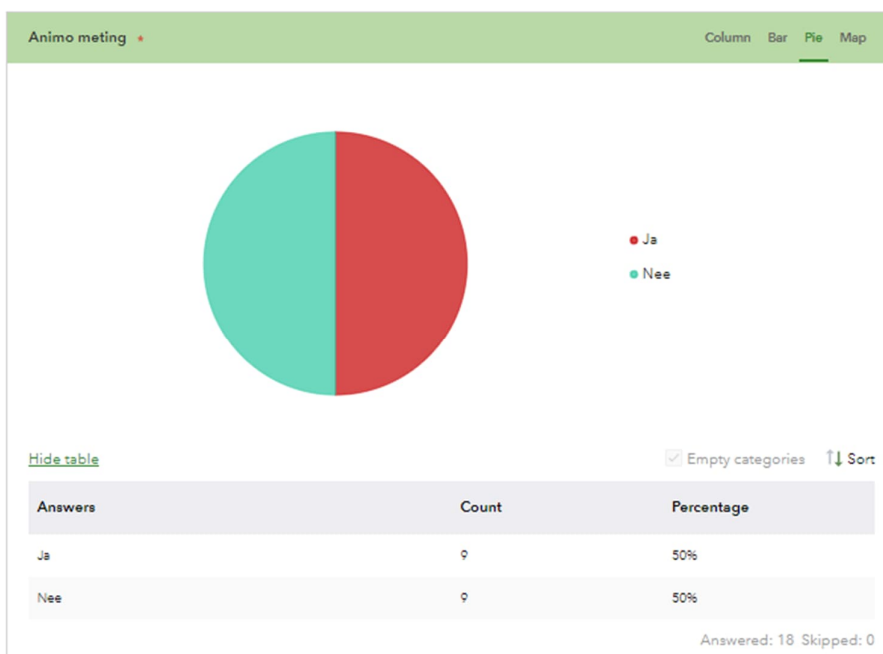


Figuur 30 Bereidheid melden propaantank enquête, vraag: "Hoe kunnen wij u het beste bereiken?"

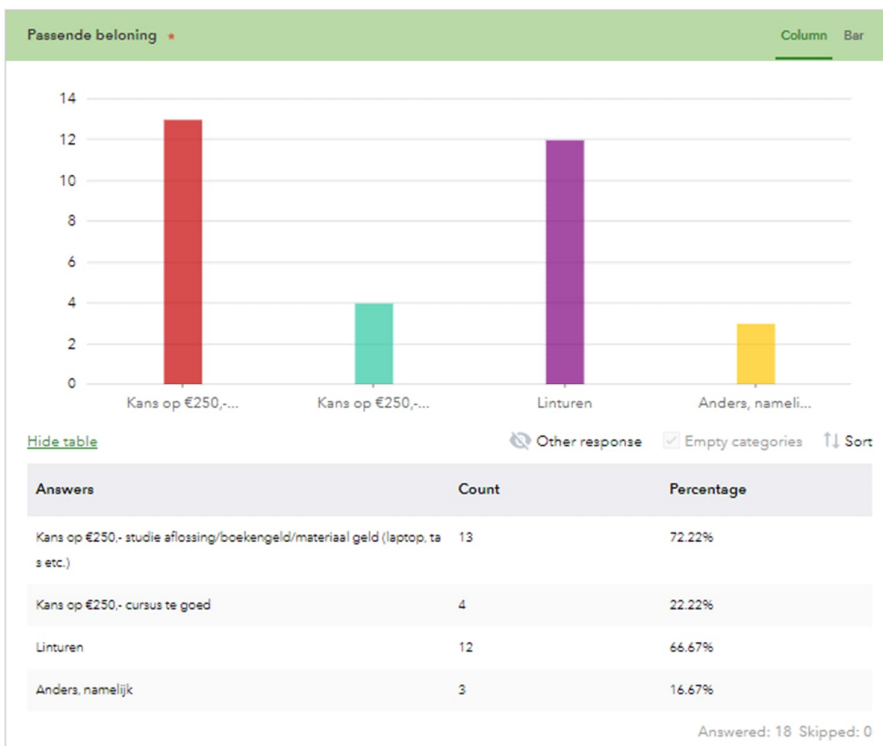
8.6. Bijlage animo meting mapathon



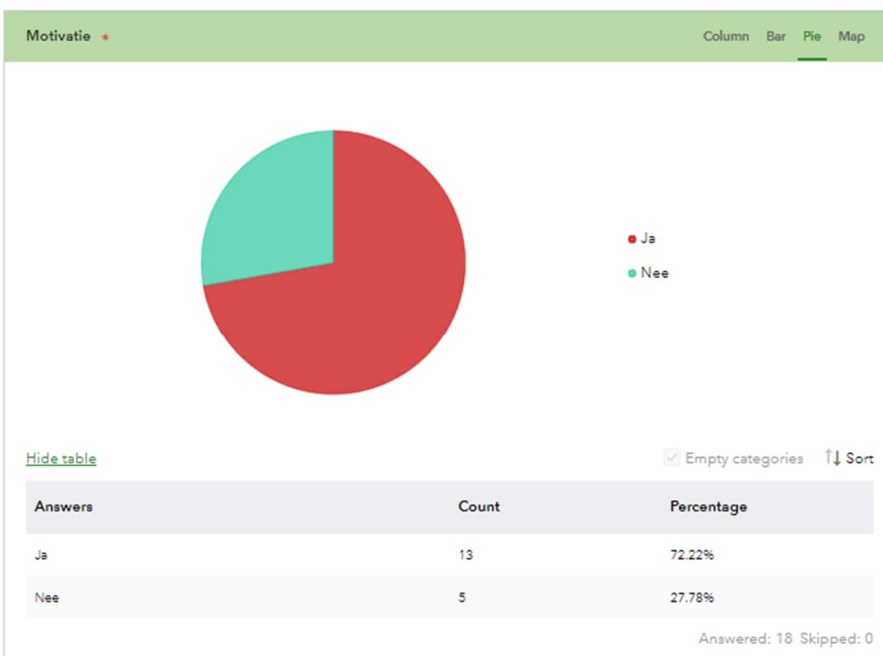
Figuur 31 Animo meting mapathon resultaten, vraag: "Zou je bereid zijn om mee te werken aan een mapathon en een aantal objecten in kaart te brengen?"



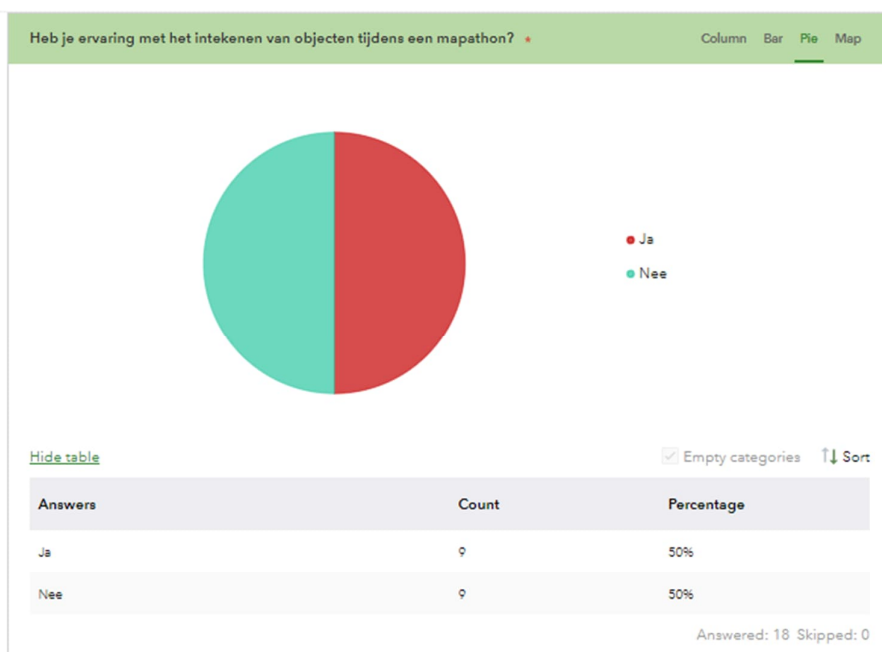
Figuur 32 Animo meting mapathon resultaten, vraag: "Zou je bereid zijn mee te werken aan een mapathon als wedstrijd? (met beloning voor de winnaars)"



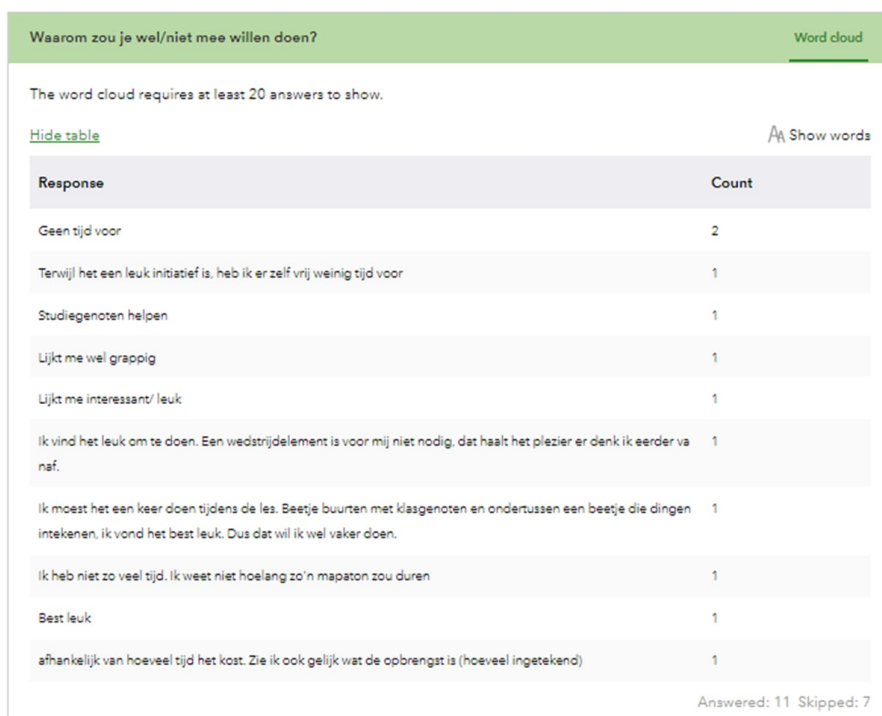
Figuur 33 Animo meting mapathon resultaten, vraag: " Wat zou een passende beloning zijn voor de 5 tot 10 mensen die de meeste objecten in kaart hebben gebracht? (meerdere antwoorden mogelijk) "



Figuur 34 Animo meting mapathon resultaten, vraag: " Zou je beter gemotiveerd zijn meer objecten in kaart te brengen als er een leaderboard is? "



Figuur 35 Animo meting mapathon resultaten, vraag: "Heb je ervaring met het intekenen van objecten tijdens een mapathon?"



Figuur 36 Animo meting mapathon resultaten, vraag: "Waarom zou je wel/niet mee willen doen?"

8.7. Berekening Fysiek data inwinnen

Kosten berekening manuren met planning

1 uur = ± 18 tanks

13.000 tanks / 18 = 722 uur

Om de schatting wat ruimer te maken wordt het ingeschatte aantal uur opgeschaald naar 800. Hierbij worden de uren berekend voor het filteren van locaties en het vooraf opstellen van de routes.

Een werkweek bestaat uit 40 uur, bij deze berekening wordt er van uitgegaan dat de twee werknemers 40 uur per week full time bezig zijn met de betreffende klus.

800 uur / 40 = 20 werkweken

Het gemiddeld salaris per uur inclusief omkosten voor het bedrijf zijn geschat op 15 euro per uur per werknemer.

(Werkzoeken.nl, 2022)

800 uur x 2 werknemers = 1600 uur

1600 uur x 15 euro = 24.000 euro

Kostenberekening reistijd met planning

Om een kostenberekening te maken gebaseerd op de route die ingepland en gereden is dient er ook een aanname gedaan te worden voor een standaard reistijd per dag dat werknemers op pad gaan. Voor deze standaard reistijd per dag wordt 20km gerekend, 10km heen en 10km terug.

1083 uur (totale geschatte tijd voor het project) met ongeveer 7 uur actieve werktijd op een dag:

1083 / 7 = 154,7 dagen

154,7 dagen x 20km = 3.094 kilometer

De geschatte kilometers voor het in kaart brengen van de propaantanks gebaseerd op Figuur 37 Fysiek data inwinnen route berekening is 11,27 kilometer. Deze berekening is alleen niet in elk gebied hetzelfde en daarom wordt er een ruime schatting gemaakt van 14 kilometer per 18 propaantanks of per uur (resultaten uit manuren berekening) inwinnen. Dit resulteert in reiskosten van:

14 kilometer x 1083 uren = 15.126 kilometer

3.094 + 15.126 = 18.220 kilometer totaal

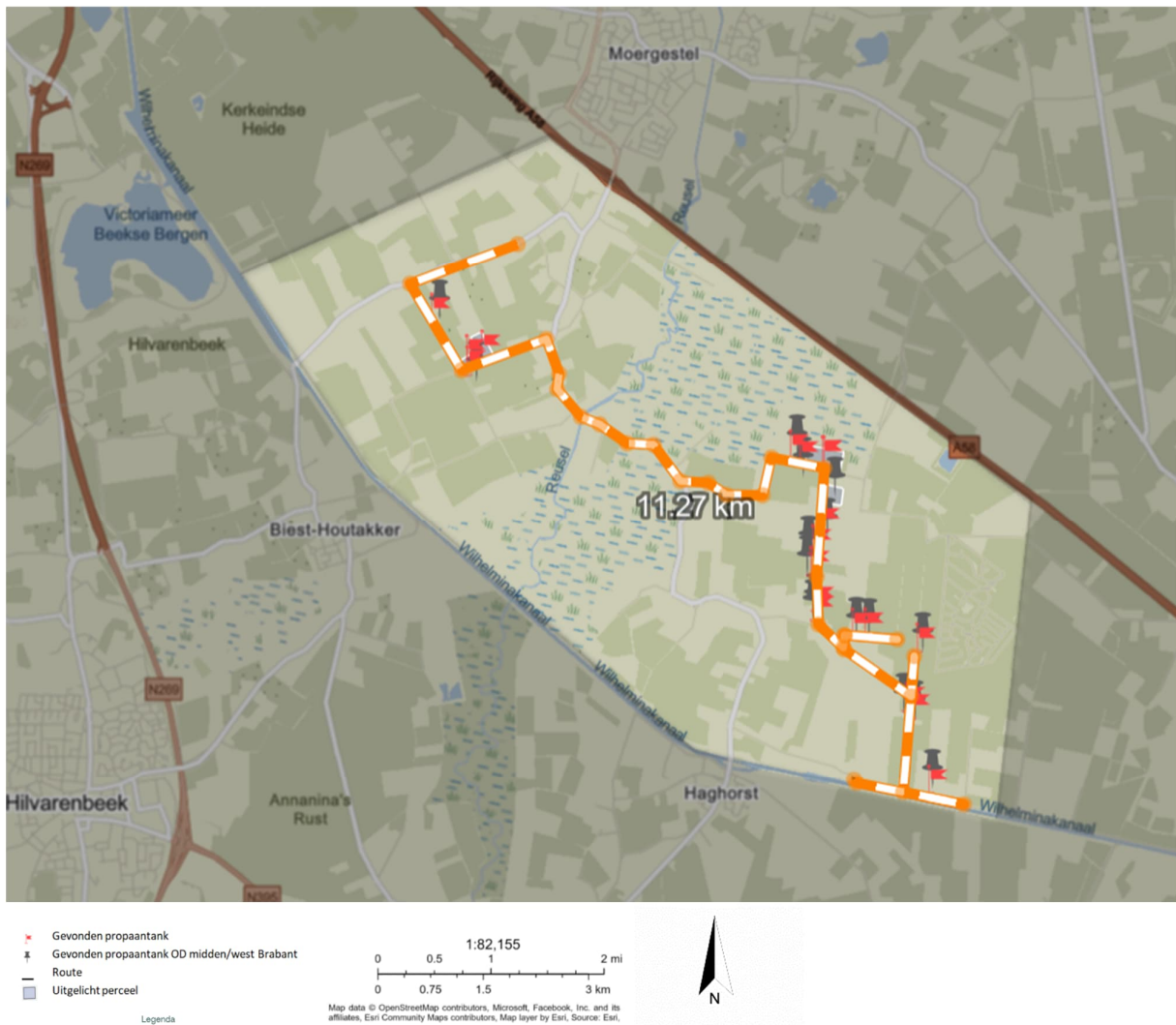
Een gemiddelde benzine auto in Nederland heeft een verbruik van 1 liter benzine op 14 kilometer

(Watkosteenauto.nl, 2016).

18.220 kilometer / 14 kilometer = 1.301,42 liter

De gemiddelde benzine prijs in Nederland wordt het meest recente " maand gemiddelde" gebruikt, dit is de benzineprijs van 2022 april (Centraal Bureau voor de Statistiek, 2022):

1.301,42 liter x 2,088 literprijs = 2.717,36 euro



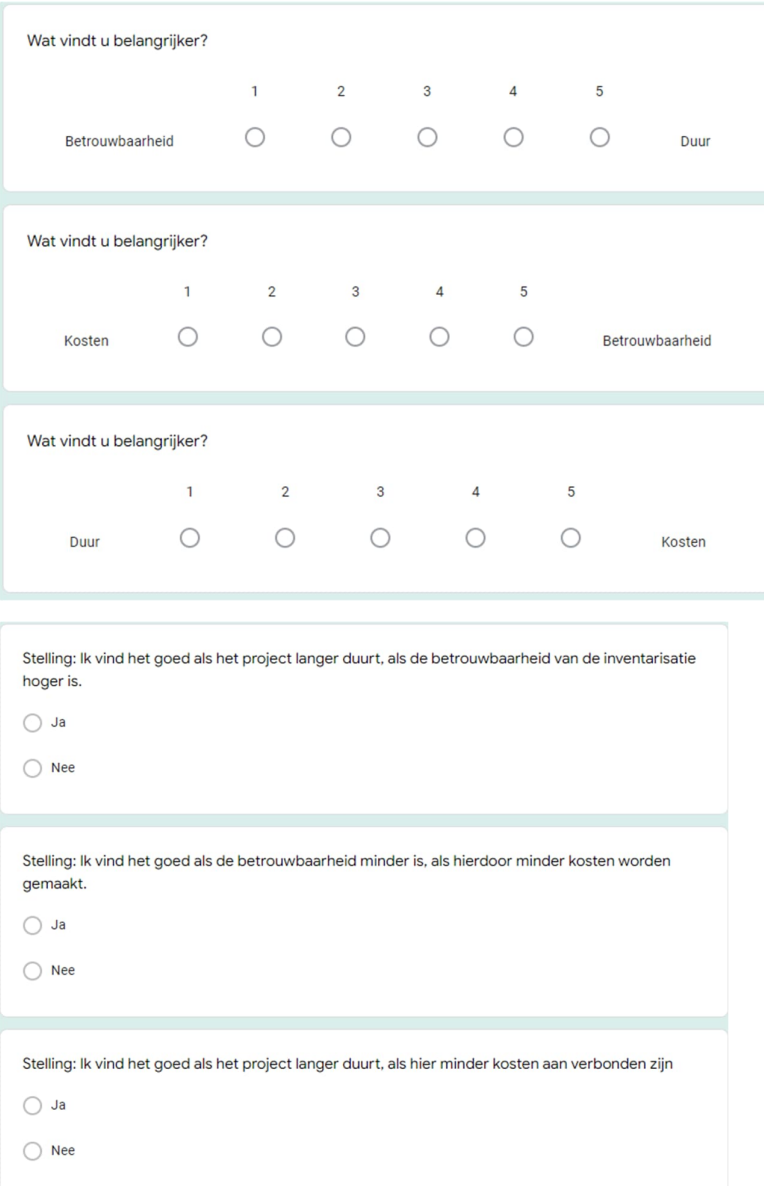
Figuur 37 Fysiek data inwinnen route berekening

8.8. Bijlage Multi Criteria analyse beschrijving

Stap 1: De vragenlijst

De vragenlijst is opgedeeld in de vraag welk criterium belangrijker is en een stelling ter bevestiging of de eerste drie vragen naar waarheid zijn ingevuld.

- Maak in google forms een vragenlijst met de vragen, zoals beschreven in Figuur 38: Vragenlijst voor criteria
- Klik na het ontvangen van de antwoorden van de gewenste respondenten in het bovenstaande menu op 'Antwoorden'
- Klik op 'Antwoorden downloaden (.csv)'
- Open het gedownloadde bestand in de verkenner om het in excel te openen.



Wat vindt u belangrijker?

	1	2	3	4	5	
Betrouwbaarheid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Duur

Wat vindt u belangrijker?

	1	2	3	4	5	
Kosten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Betrouwbaarheid

Wat vindt u belangrijker?

	1	2	3	4	5	
Duur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kosten

Stelling: Ik vind het goed als het project langer duurt, als de betrouwbaarheid van de inventarisatie hoger is.

Ja

Nee

Stelling: Ik vind het goed als de betrouwbaarheid minder is, als hierdoor minder kosten worden gemaakt.

Ja

Nee

Stelling: Ik vind het goed als het project langer duurt, als hier minder kosten aan verbonden zijn

Ja

Nee

Figuur 38: Vragenlijst voor criteria

Stap 2: Bewerkingen in Excel

Om de resultaten beter te kunnen interpreteren worden de resultaten in excel weergegeven.

- Klik in excel op 'Gegevens'
- Klik vervolgens op 'Uit tekstbestand/CSV'
- Klik in de pop-up van de verkenner op het gedownload csv bestand.
- Klik in de wizard van excel op 'scheiding door'.
- Klik in dit menu op 'komma's'

Stap 3: Invulling van de tabel

Hierin worden de resultaten van de vragenlijst omgerekend naar een factor, zoals te zien in Tabel 3 Multi Criteria Analyse.

- Tel bij elke vraag de uitkomstgetallen op van alle respondenten
- Bereken bij elke vraag het gemiddelde door de som van alle respondenten te delen door het aantal respondenten.
- Bereken van deze drie gemiddeldes de afstand tot 3 (de mediaan van de keuze 1 t/m 5)
- Noteer bij elk van de drie criteria hoeveel deze gemiddeldes van elkaar afwijken. (Wanneer (betrouwbaarheid 1-5 duur) een gemiddelde van 2,66 heeft krijgt betrouwbaarheid een +0,34 en duur -0,34.)
- Elk van de drie criteria krijgt 2 van deze afwijkingen van het gemiddelde. Bereken hier de som van.
- Wanneer een negatief getal uitkomt. Neem de positieve variant en trek dit getal van 1 af om de factor te bepalen. De factoren zijn bepaald.
- Maak een soortgelijke tabel, zoals in Tabel 3 Multi Criteria Analyse en voer de factoren in onder de kolom 'factor'.

Stap 4: Invulling van de MCA tabel

Tabel 8: Scorelijst voor input MCA

Score	1	3	5
Betrouwbaarheid			
Duur	> 2 jaar	± 1 jaar	± 6 maanden
Kosten	> 2 ton	± 1 ton	< 0,5 ton

- Gebruik in elk van de grijze vlakken van deze tabel de volgende formule in excel: '= \$ CEL FACTOR * score'. Hier staat CEL FACTOR voor een van de drie factoren, in dit geval 1; 0,7 en 0,3. Score staat voor de score van 1 t/m 5 die in het volgende punt wordt toegelicht.
- Scoor elk van de methoden in Tabel 3 Multi Criteria Analyse op de criteria met een score van 1 (slecht) t/m 5 (goed). Doe dit op basis van Tabel 8: Scorelijst voor input MCA. Voor betrouwbaarheid kan de score worden bepaald op basis van resultaten uit je methoden.
- Voer in de rij 'Totaal' van Tabel 3 Multi Criteria Analyse '=SOM(BOVENSTE GRIJZE CEL : ONDERSTE GRIJZE CEL)' in. Doe dit voor elke kolom. Hieruit volgt een totaalscore van 1 t/m 10.

8.9. Mapathon

Zoals in de methode beschreven is een mapathon een samenkomst van vrijwilligers waarbij objecten in kaart gebracht worden of gevalideerd kunnen worden. Mapathons worden al regelmatig gebruikt bij bijvoorbeeld het Rode Kruis die Missing Maps dagen organiseert. Om een mapathon goed uit te voeren zijn er een aantal randvoorwaarden die onder andere opgesteld zijn door de organisatie van OpenStreetMap (OpenStreetMap, 2017). Deze handleiding biedt handvatten om een goede mapathon te organiseren die hieronder kort beschreven worden in een stappenplan, voor meer context of voorbeelden per stap wordt er verwezen naar de handleiding (OpenStreetMap, 2017). In het stappenplan wordt soms een onderstreepte opmerking geplaatst die extra context biedt voor deze specifieke casus.

1. De organisatie geeft hierbij uit ervaring aan om minimaal een maand van tevoren te beginnen met plannen van een mapathon.

2. De indeling van uw mapathon voorbereiden (grootte van locatie en de beschikbare faciliteiten).

- Is het een centrale locatie en gemakkelijk te bereiken?
- Kan de wifi 100 verbindingen tegelijk aan?
- Kunnen er genoeg personen binnen, toereikend aan het aantal personen dat je verwacht?
- Zijn er voldoende stoelen en tafels?
- Voldoende elektrische aansluitpunten?
- Groot scherm aanwezig voor uitleg en een introductie presentatie?
- Heeft het kleinere ruimtes voor onderverdeling van groepen?
- Zijn er voldoende ervaren helpers aanwezig om de mapathon uit te voeren?

3. De mapathon plannen (datum bepaling en boeken van locatie). (in het geval van het ministerie gecombineerd met een school waar GIS studenten werken is het waarschijnlijk mogelijk de school een locatie te laten faciliteren)

- Bepaal datum en de tijd voor de mapathon (Wees voorzichtig door het niet samen te plannen met enig ander belangrijk evenement dat het aantrekken van vrijwilligers en bezoekers voor uw gebeurtenis zou kunnen beïnvloeden). Reserveer minimaal 2,5 tot 3 uur voor uw evenement (een half uur voor speeches en training, een half uur voor pauze en eten, twee uur om in kaart te brengen of te valideren).
- Boek de locatie.
- Verdeel uitnodigingen aan mappers via sociale media om mensen te trekken.
- Biedt snack en verfrissingen aan op het evenement
- Stel een project in voor het evenement. Bij hulp met het opstellen van projecten en ook voor materiaal voor de mapathon voor de tafels. De Tasking Manager is een goede bron voor nuttige projecten.
- Verzamel helpers die u nodig hebt voor het evenement (It experts, ervaren mapper, invaller)
- Voeg uw evenement toe aan de kalender van mapathons.
- Zorg eventueel dat er iemand beschikbaar is om foto's te nemen en verslag te doen van het evenement.
- Stel iemand aan om de presentatie te filmen en een live koppeling te maken voor eventuele online mappers wanneer u dit wilt toepassen.

4. De week voor de mapathon.

- Bezoek de locatie en controleer alle voorbereidingen
- Stel een lijst samen met alle vrijwilligers om hen te mailen.
- Verstuur een bericht om vrijwilligers te herinneren hun laptop en muis mee te nemen. (Wanneer gebruik gemaakt wordt van openstreetmap, herinner mensen hun account in te stellen.)
- Maak een lay-out voor de inrichting van het evenement met inachtneming van beschikbare elektrische punten en zicht op het presentatiescherm.
- Controleer of u de huidige projecten om in kaart te brengen voor het evenement hebt.
- Controleer of er genoeg op kas is voor verfrissingen en versnaperingen.
- Regel Materiaal voor de mapathon (In het geval van het ministerie de ingewonnen propaantanks die gevalideerd dienen te worden of satelliet data waar propaantanks op herkend moeten worden)
- Controleer of er verlengsnoeren etc. nodig zijn en of deze geregeld zijn.
- Laat uw helpers weten een half uur van te voren minimaal aanwezig te zijn om het evenement op te zetten.
- Controleer of de wifi toegankelijk is voor iedereen

5. Het evenement begint

- Vraag om stilte en begin met een introductie presentatie.
- Introduceer de ervaren mappers en de vrijwilligers die helpen bij het evenement.
- Wees gedurende het evenement aanwezig om te helpen bij eventuele vragen of fouten die gemaakt worden door mappers.

6. De controle rondes - Verwerk in het proces dat er eventueel groepen zijn die ingetekende objecten valideren of dat elk ingetekend object in een controle ronde door andere mappers gecontroleerd worden.