



Automatische beeldherkenning als instrument voor museumcollecties

Innovatieproject in de Nederlandse
natuurhistorische musea

Eindrapportage
Mondriaan Fonds
Prins Bernhard Cultuurfonds
NLBIF
Juli 2022

Inhoud

Samenvatting — 3

Inleiding — 6

Herkenningsmodellen voor vijf deelcollecties — 9

Processtappen — 9

Deelcollecties selecteren — 9

Data verzamelen — 10

Data opschonen en uniformeren — 10

Modellen trainen — 12

Modellen evalueren en hertrainen — 12

Van collectie naar model — 14

Kegelslakken (Conidae) — 14

Zuidoost-Aziatische dagvlinders (papillotten) — 15

Vogeleieren — 17

Eikapsels roggen en haaien — 19

Maskers — 20

Modellen ontsluiten voor gebruik — 21

Verdiepende soort- en objectinformatie — 22

Crowdsourcing-campagne voor één deelcollectie — 27

Tekeningen Anselmus de Boodt — 27

Communicatie en publiciteit — 29

Terug- en vooruitblik — 31

Wat hebben we geleerd? — 31

Onze modellen in de werkprijktijk — 31

AI en Naturalis — 32

Bijlage — 35

Flyer en visitekaartje — 35

Samenvatting

AI voor registratie en ontsluiting van museumcollecties

Naturalis heeft zich met het project *Automatische beeldherkenning als instrument voor museumcollecties* ten doel gesteld te onderzoeken in welke mate *Artificial Intelligence* (AI), en specifiek automatische beeldherkenning, kan bijdragen aan de registratie en ontsluiting van museumcollecties. Dit gebeurde door in nauwe samenwerking met verschillende projectpartners en betrokkenen een aantal concrete herkenningsmodellen te ontwikkelen voor verschillende typen objecten. Deze herkenningsmodellen zijn toegankelijk gemaakt voor mens en machine via een web-based publieksinterface en een API: <https://museum.identify.biodiversityanalysis.nl/>.

Verdwijnde identificatie-expertise

Objecten (specimens) in natuurhistorische collecties worden van oudsher op naam gebracht door taxonomen, soortspecialisten. Het kunnen benoemen en duiden van soorten is belangrijk voor onderzoek naar, inzicht in en beleving van de natuur en biodiversiteit.

Van populaire soortgroepen is kennis vaak breed verspreid, maar voor andere soortgroepen zijn de specialisten en kennis schaars. Vrijwel alle natuurhistorische musea kampen met enerzijds omvangrijke, en immer groeiende collecties en anderzijds een afnemend aantal beschikbare soortspecialisten. Hierdoor zijn in dit type collecties vaak grote aantallen objecten (specimens) nog niet geïdentificeerd. Automatische beeldherkenning zou kunnen bijdragen aan het op naam brengen van (een deel van) deze specimens, waardoor de specialistische kennis van taxonomen en soortspecialisten gerichter kan worden ingezet voor de lastigere gevallen, bijvoorbeeld waarin geautomatiseerde identificatie op basis van een foto geen uitsluitsel kan geven (bijvoorbeeld omdat microscopische kenmerken relevant zijn) of als er sprake is van een nog niet eerder ontdekte soort.

Wat leerden we uit de pilots?

In dit project onderzochten we welke typen objecten of collecties wel of niet in aanmerking komen voor automatische beeldherkenning, hoe deze beeldherkenning kan worden ingezet in het werkproces van collectieregistratie en ontsluiting, en hoe AI antwoord kan geven op publieksvragen.



The screenshot shows the homepage of the 'Automatische beeldherkenning voor museumcollecties' project. At the top, there is a navigation bar with the Naturalis Biodiversity Center logo and links for Home, Colofon, API, and FAQ. Below the navigation bar, there are four images: a shell, a dark, curved object, a mask, and a butterfly. The main heading is 'Automatische beeldherkenning voor museumcollecties' in green. Below the heading, there is a 'switch to English' link. The main text asks: 'Welke soort vlinder heb ik gedigitaliseerd? Welk roggeneikapsel heb ik op het strand gevonden? Tot welke cultuur behoort dit masker in mijn collectie? Deze webapplicatie, gebaseerd op kunstmatige intelligentie (specifiek *machine learning*), kan je helpen dergelijke identificatievragen geautomatiseerd te beantwoorden.' Below this, it states: 'Bij invoer zal een model *altijd* drie resultaten geven, de drie meest gelijkende klassen dat het model bevat. Het model kan dus niet aangeven dat het iets niet kent.' At the bottom, it says: 'Om de determinaties die de modellen leveren te kunnen verifiëren, is een link naar aanvullende informatie toegevoegd.'

Homepage 'Automatische beeldherkenning voor museumcollecties'.

In het project zijn voor uiteenlopende typen objecten afkomstig uit de collecties van Naturalis, de projectpartners Museon-Omniversum en Het Natuurhistorisch en andere organisaties diverse modellen opgeleverd waarmee typen objecten of specifieke soorten kunnen worden gedetermineerd of geïdentificeerd. Het project leverde een aantal leerpunten op voor de museale sector over het ontwikkelen en inzetten van beeldherkenning en de specifieke projectfasen in de aanloop daarnaartoe. Belangrijke punten zijn bijvoorbeeld de eisen waaraan de beelden en bijbehorende data die benodigd zijn om modellen te trainen moeten voldoen, de benodigde hoeveelheid trainingsdata voor verschillende typen objecten en hoe het beste onderscheidbare klassen in je model te definiëren en op welk (taxonomisch) niveau te classificeren.

Publieksinterface en verdiepende informatie

De in dit project gelanceerde publieksinterface levert de gebruiker bij de determinaties die de verschillende herkenningsmodellen genereren aanvullende informatie over de soort of het type object. Dit gebeurt zowel ingebed in de publieksinterface zelf als via een link naar bestaande aanvullende online bronnen. Doel hierbij is dat gebruikers de AI-identificaties (ook wel determinaties) kunnen verifiëren aan de hand van gevalideerde kennis over het specifieke type object of de soort. Vanuit het project is hiertoe inhoud - bondige brokjes informatie in tekstuele vorm,

al dan niet voorzien van extra afbeeldingen - ontwikkeld door Naturalis en projectpartners, met inzet van onder andere projectmedewerkers. Ook is door samenwerking met partners handig gebruik gemaakt van inhoud die in het kader van andere, lopende projecten is gecreëerd. De inhoud is duurzaam geborgd in de informatie-infrastructuur van Naturalis en is geschikt voor hergebruik in andere context.

Symposium voor en met de erfgoedsector

We organiseerden op 16 mei 2022 een afsluitend symposium *Museumcollecties & AI* in Naturalis. Hier deelden we de projectresultaten en lessen met collega's uit musea, bibliotheken en archieven, met biodiversiteitsonderzoekers en met andere geïnteresseerden in het onderwerp AI. Tijdens deze middag is ook de opgeleverde publieksinterface gelanceerd: een website waarop de ontwikkelde herkenningsmodellen te gebruiken zijn door zowel publiek als erfgoedinstanties. Ook kwamen enkele andere AI-initiatieven in de erfgoedsector aan bod.

We zijn het Mondriaan Fonds, NLBIF en het Prins Bernhard Cultuurfonds zeer erkentelijk voor hun steun. Dankzij deze financiële bijdragen heeft dit project tot stand kunnen komen. In de voorliggende inhoudelijke en financiële eindverantwoording gaan we dieper in op de projectaanpak, de resultaten, de leerpunten en de toekomst.



Symposium 'Museumcollecties & AI' in Naturalis.



33
Pitangus sulphuratus
Aantal 2 - legel. 124
Dat: 18-5-1900
Loc:
Coll: Penard

32
Pitangus sulphuratus
Aantal 2 - legel. 154
Dat: 18-5-1900
Loc:
Coll: Penard

31
Pitangus sulphuratus
Aantal 2 - legel. 155
Dat: 18-5-1900
Loc:
Coll: Penard

30
Pitangus sulphuratus
Aantal 2 - legel. 164
Dat: 18-5-1900
Loc:
Coll: Penard

34
Pitangus sulphuratus
Aantal 2 - legel. 134
Dat: 18-5-1900
Loc:
Coll: Penard

34
Pitangus sulphuratus
Aantal 2 - legel. 144
Dat: 18-5-1900
Loc:
Coll: Penard

33
Pitangus sulphuratus
Aantal 2 - legel. 155
Dat: 18-5-1900
Loc:
Coll: Penard

33
Pitangus sulphuratus
Aantal 2 - legel. 165
Dat: 18-5-1900
Loc:
Coll: Penard

35
Pitangus sulphuratus
Aantal 4 - legel. 145
Dat: 18-5-1900
Loc:
Coll: Penard

35
Pitangus sulphuratus
Aantal 2 - legel. 57
Dat: 23-2-1901
Loc:
Coll: Penard

35
Pitangus sulphuratus
Aantal 2 - legel. 65
Dat: 19-4-1901
Loc:
Coll: Penard

36
Pitangus sulphuratus
Aantal 2 - legel. 47
Dat: 26-10-1900
Loc:
Coll: Penard

35
Pitangus sulphuratus
Aantal 2 - legel. 59
Dat: 23-2-1901
Loc:
Coll: Penard

35
Pitangus sulphuratus
Aantal 2 - legel. 57
Dat: 19-4-1901
Loc:
Coll: Penard

36
Pitangus sulphuratus
Aantal 1 - legel. 150
Dat: 31-12-1900
Loc:
Coll: Penard

36
Pitangus sulphuratus
Aantal 2 - legel. 59
Dat: 31-3-1901
Loc: Cutcherts N. of Paramaribo
Coll: Penard

36
Pitangus sulphuratus
Aantal 3 - legel. 68
Dat: 19-4-1901
Loc: Cutcherts N. of Paramaribo
Coll: Penard

37
Pitangus sulphuratus
Aantal 1 - legel. 150
Dat: 31-12-1900
Loc:
Coll: Penard

37
Pitangus sulphuratus
Aantal 5 - legel. 160
Dat: 31-3-1901
Loc:
Coll: Penard

36
Pitangus sulphuratus
Aantal 2 - legel. 69
Dat: 27-4-1901
Loc:
Coll: Penard

37
Pitangus sulphuratus
Aantal 5 - legel. 160
Dat: 31-3-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 2 - legel. 170
Dat: 6-5-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 4 - legel. 64
Dat: 6-5-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 3 - legel. 171
Dat: 9-5-1901
Loc: Cutcherts N. of Paramaribo
Coll: Penard

38
Pitangus sulphuratus
Aantal 3 - legel. 170
Dat: 27-6-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 4 - legel. 64
Dat: 6-5-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 3 - legel. 171
Dat: 9-5-1901
Loc: Cutcherts N. of Paramaribo
Coll: Penard

38
Pitangus sulphuratus
Aantal 7 - legel. 170
Dat: 1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 4 - legel. 64
Dat: 6-5-1901
Loc:
Coll: Penard

38
Pitangus sulphuratus
Aantal 3 - legel. 171
Dat: 9-5-1901
Loc: Cutcherts N. of Paramaribo
Coll: Penard

38
Pitangus sulphuratus
Aantal 7 - legel. 170
Dat: 1901
Loc:
Coll: Penard

Inleiding

Automatische beeldherkenning wordt al op diverse manieren toegepast in de museale en culturele sector. Bekende voorbeelden zijn tekst- en handschriftherkenning, het herkennen van merktekens of de herkenning van personen en objecten in afbeeldingen. Ook in biodiversiteitsonderzoek speelt AI in toenemende mate een rol, bijvoorbeeld in de herkenning en verspreiding van soorten planten, dieren en schimmels. Deze natuurwaarnemingen, zoals verzameld en door natuurliefhebbers en kenners ingevoerd in de app ObsIdentify of op het platform Waarneming.nl, maar ook van wildcamera's, vormen grote datasets.¹ De waarnemingen worden al dan niet met behulp van AI-identificaties op naam gebracht, en waar nodig of wenselijk gevalideerd door specialisten.

Samenwerkingsproject

In dit project, een samenwerking van Naturalis met projectpartners Museon-Omniversum, Het Natuurhistorisch, Natuurmuseum Brabant en het Rijksmuseum, leggen we de focus op de *museale* collecties en de herkenning van soorten van bijvoorbeeld ongeregistreerde of nieuwe specimens. Het is een opmaat naar de inzet van beeldherkenning voor zowel het Nederlandse museale veld als Europese natuurhistorische (en cultuurhistorische) instellingen voor het bespoedigen van de collectieregistratie en -ontsluiting.

Hoe werkt automatische beeldherkenning van soorten?

Geautomatiseerde beeldherkenning wordt gedaan op basis van *deep learning*, een tak van *machine learning* waarin gebruikt wordt gemaakt van neurale netwerken. Met deze technieken kan een

computer op basis van gevalideerde foto's van exemplaren automatisch een herkenningsmodel maken, waarmee de betreffende soorten op nieuwe foto's ook herkend worden. Deze manier van soortherkenning is een aanvulling op het zelf - als mens - vormen van indelingen of categorieën op basis van uiterlijke kenmerken. Voor zo'n herkenningsmodel zijn, afhankelijk van het type object en van variatie binnen soorten en in aanzichten, relatief grote betrouwbare datasets nodig. Een samenwerking met partners is dan ook onontbeerlijk om voldoende data te verzamelen waarmee een herkenningsmodel getraind kan worden om heel specifieke soorten te herkennen.

Projectdoelen

De in ons oorspronkelijke projectplan geformuleerde projectdoelen zijn:

- het delen van kennis van de door Naturalis vergaarde kennis op het gebied van beeldherkenning met de andere natuurhistorische musea (en museumsector), en voor Naturalis specifiek het verkrijgen van feedback en ervaringen van de musea;
- procesoptimalisatie (data)beheer binnen natuurhistorisch collecties, zoals tijdsbesparing bij digitalisering en systematisch opslaan van collectie in de depots;
- ondersteunen van collectiebeheerders in hun taken en het vergroten van hun kennis;
- inzet van geautomatiseerde beeldherkenning bij het determineren van natuurvondsten en natuurwaarnemingen van museumbezoekers;
- relatief eenvoudige determinaties te automatiseren en te versnellen waardoor specialisten zich kunnen focussen op de moeilijke of twijfelgevallen, en op het vastleggen en overdragen van hun kennis;
- de kwaliteit en kwantiteit van waarnemingen te verhogen, zodat uiteindelijk een vollediger beeld ontstaat van verspreidingen van soorten in tijd en ruimte;
- zo indirect bijdragen aan behoud van de (Nederlandse) biodiversiteit.

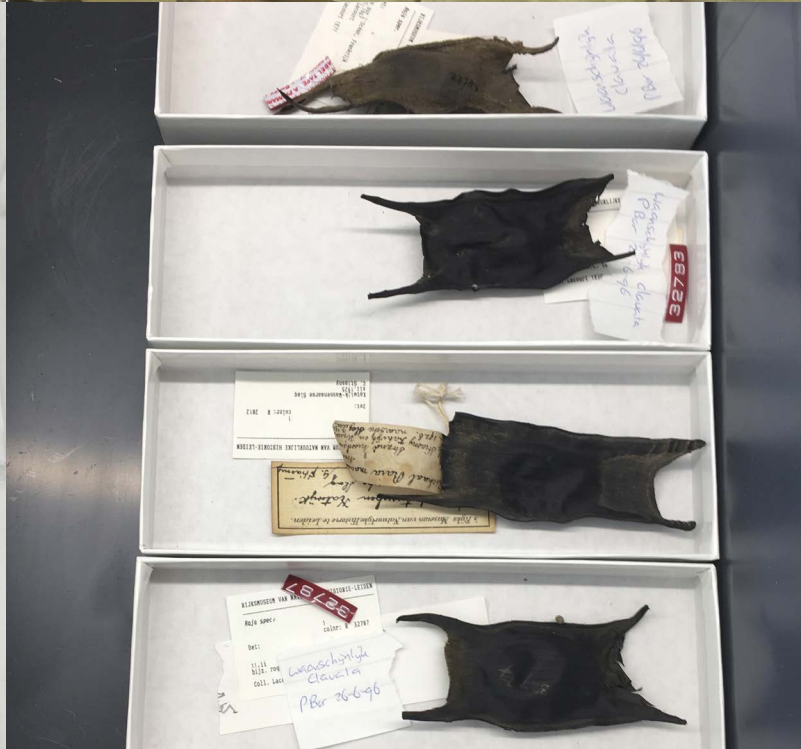
¹ Website <https://waarneming.nl/> en app ObsIdentify: <https://waarneming.nl/apps/obsidentify/>. Beide zijn initiatieven van Stichting Observation International, een partner van Naturalis. ObsIdentify is mede gefinancierd door Prins Bernhard Cultuurfonds.



Afgevaardigden van de projectpartners tijdens de startbijeenkomst op 25 februari 2020 in Naturalis.

In de hierop volgende hoofdstukken gaan we in op de activiteiten in en uitkomsten van ons project. Concrete resultaten die vormgeven aan de gestelde doelen zijn:

- Herkenningsmodellen voor vijf deelcollecties;
- Crowdsourcing-campagne voor één deelcollectie;
- Een *user interface* en API om de modellen te gebruiken;
- Verdiepende informatie over soorten / objecttypen in deze modellen;
- Geleerde lessen uit deze pilots;
- Communicatie en publiciteit (waaronder een eindsymposium).



Herkenningmodellen voor vijf deelcollecties

Processtappen

Om modellen te trainen ten behoeve van identificatie van soorten of van typen objecten volgden we een aantal vaste processtappen:

1. Deelcollecties selecteren
2. Data verzamelen
3. Data opschonen en uniformeren
4. Modellen trainen
5. Modellen evalueren en hertrainen

1. Deelcollecties selecteren

De projectpartners onderzochten gezamenlijk welke collecties in aanmerking konden komen voor het ontwikkelen van een model om automatische beeldherkenning toe te passen. Hiervoor organiseerde het projectteam op 25 februari 2020 een startbijeenkomst in Naturalis waarbij afgevaardigden van de projectpartners aanwezig waren. Tijdens deze bijeenkomst is gekeken naar welke vragen er leefden bij de partners en welke gedigitaliseerde collecties zij konden bijdragen.

Bij het selecteren van collecties voor beeldherkenning waren de eigenschappen van de objecten zelf, de toepasbaarheid in het werkproces of een bestaande publieksvraag - ook buiten de natuurhistorische collecties - leidend. Daarnaast was de beschikbaarheid van voldoende gedetermineerd en gedigitaliseerd materiaal een belangrijk criterium om de modellen binnen de looptijd van het project te kunnen realiseren.

Ook hebben we breder binnen Naturalis en onder de aangesloten projectpartners via een enquête onderzocht wat collectiebeheerders verwachten van automatische beeldherkenning als instrument voor museumregistratie en welke wensen men heeft ten aanzien van een *interface* (applicatie) die in dergelijk gebruik voorziet.

We hebben met deze input een inventarisatielijst van kansrijke collecties samengesteld. Uit deze

inventarisatielijst selecteerden we vervolgens een aantal collecties om in dit project concreet mee aan de slag te gaan:

1. **Kegelslakken (Conidae)**
2. **Zuidoost-Aziatische dagvlinders** (op basis van de collectie papillotten Naturalis)
3. **Vogeleieren** van in Nederland voorkomende soorten
4. **Eikapsels roggen- en haaien** van aan de Nederlandse kust voorkomende soorten
5. **Maskers** (etnografische collectie maskers)
6. **Tekeningen** albums Anselmus de Boodt

Voor collecties 1 t/m 5 hebben we herkenningmodellen ontwikkeld. Voor de door het Rijksmuseum ingebrachte collectie van tekeningen met planten, dieren en paddenstoelen van De Boodt is in het project voor een andere aanpak gekozen. Voor deze tekeningen is een crowdsourcing-campagne op het platform van Wikimedia Commons opgezet zodat het publiek de afgebeelde soorten kon *taggen* met gestructureerde data. Dit deelproject wordt nader toegelicht in de paragraaf *Tekeningen Anselmus de Boodt*.

Tijdens het project zijn enkele collecties uit de inventarisatielijst waarmee we aan de slag gingen voor het maken van een herkenningmodel vervolgens weer afgevallen. Een belangrijke reden hiervoor was dat er niet voldoende gedetermineerd en gedigitaliseerd materiaal voorhanden was om de enorme variatie te dekken (bijvoorbeeld mineralen², fossielen en botten). Een andere reden was dat er andere projecten op hetzelfde gebied al elders plaatsvonden - binnen of buiten Naturalis - en we

2 Mineralen zijn zeer variabel qua vorm en vereisen daarom veel trainingsmateriaal. Ook zijn ze als type object complex voor beeldherkenning: sommige mineralen zijn alleen op basis van niet-visuele kenmerken te onderscheiden of alleen op basis van bijvoorbeeld hardheid of lichtbreking. Bovendien zijn veel specimens in deze collectie samengestelde objecten, dat wil zeggen bestaande uit meer dan één mineraal.



Objecten uit de zes deelcollecties.

met onze middelen geen werk dubbelop wilden doen (bijvoorbeeld het geval bij de herbarium-collectie³).

2. Data verzamelen

Om een model te trainen moet deze gevoed worden met data. Hieronder verstaan we een soort- of objecttype-aanduiding, die een specifieke klasse in het model vertegenwoordigt, en een bepaald aantal afbeeldingen per soort of objecttype. Zo'n soort of type object wordt in een herkenningmodel aangeduid met de term klasse. Het model zal na determinatie een antwoord geven in de vorm van een label: een klasse-aanduiding. In natuurhistorische collecties kunnen bijvoorbeeld specifieke soorten, genera of families staan voor afzonderlijke klassen in het model.

Bij andersoortige collecties kan een ander kenmerk, zoals stijl of herkomst, doel of verschijningsvorm van een object een klasse typeren. Een voorbeeld van deze classificatie is hoe het maskermodel is ingericht, met klassen als een West-Afrikaans gelede masker of een Balinees wajangmasker.

Hoeveel afbeeldingen heb je nodig?

Het aantal afbeeldingen dat per klasse benodigd is, is van meerdere factoren afhankelijk, zoals de variëteit binnen een soort of objecttype, variatie in aanzichten en variatie in randinformatie zoals achtergrondkleur of artefacten in het beeld. Hierdoor kan dat minimum aantal vereiste afbeeldingen (examples) per klasse sterk variëren van model tot model. Voor elk in het project getraind model werden data verzameld uit de collecties van Naturalis en van de projectpartners. Om de werking (prestaties) van de modellen te optimaliseren én om de modellen te testen, werden ook data verzameld van andere instituten, in en buiten Nederland.

Omdat de beschikbare hoeveelheid trainingsdata per klasse niet altijd vooraf bekend was, bleken er na training van een model vaak scheve verhoudingen in de hoeveelheden trainingsdata per klasse te zijn. Sommige klassen hadden maar één of twee afbeeldingen, anderen veel meer. Dit heeft invloed op de werking van het model. Minimaal twee afbeeldingen per klasse is een vereiste, maar minimaal tien of vijftig (of zelfs meer) is meestal beter voor de werking van het model. Om die reden konden achteraf klassen uit het model vallen als er onvoldoende afbeeldingen van die specifieke soort of dat type object beschikbaar waren. Bij een model dat werkt op minimaal tien afbeeldingen (examples) per klasse gebeurt dat - logischerwijs - eerder dan bij een minimum van twee afbeeldingen per klasse.

Extra trainingsdata genereren

In dit project hebben we gebruik gemaakt van al beschikbare afbeeldingen van objecten uit de collecties van de betrokken partners om modellen te trainen. Ter aanvulling zijn ook specifiek voor het project gedetermineerde specimina uit verschillende collecties gedigitaliseerd. Daartoe werden speciale instructies opgesteld voor het digitaliseringsteam van Naturalis, om te beantwoorden aan de specifieke toepassing in het trainen van de herkenningmodellen. Het voor het project gegenereerde beeldmateriaal komt aan de orde bij de ontwikkeling van de betreffende modellen.

3. Data opschonen en uniformeren

Een volgende belangrijke processtap in het project was het uniformeren - of standaardiseren - van de beschikbare collectiedata. Deze data bepaalt de klassen van een model. Wanneer er variatie bestaat in notatie van één en dezelfde soort, bijvoorbeeld door verschillen in de gehanteerde schrijfwijzen, maar ook door spelfouten, verouderde wetenschappelijke soortnamen of door variabel gebruik van haakjes in de wetenschappelijke naam of verschillen in auteursnaam of jaartal van beschrijving bij dezelfde soort, worden er onterecht verschillende klassen in het model gesuggereerd (zie ook kadertekst). Echte schrijffouten proberen we te corrigeren in de onderliggende collectiedatabases (data-opschoning).

³ Voor herbariumvellen is een grootschalige internationale beeldherkennings-pilot op de Kaggle-community voor *machine learning* georganiseerd, waarbij Naturalis ook collectie aanleverde (Kaggle Herbarium 2021 - Half-Earth Challenge): <https://www.kaggle.com/competitions/herbarium2021-fgvc8/overview>. Resultaten zijn verwerkt in het artikel *The Herbarium 2021 Half-Earth Challenge Dataset and Machine Learning Competition* <https://www.frontiersin.org/articles/10.3389/fpls.2021.787127/full>

Pimpelmees heeft data-opooschoning

In de gebruikte collectiedatabases troffen we tijdens dit project bijvoorbeeld een zestal verschillende aanduidingen aan voor de klasse "pimpelmees", ook wel *Cyanistes caeruleus* (Linnaeus, 1758). Deze data moest geüniformeerd worden tot één klasse om te voorkomen dat het herkenningmodel de bijbehorende trainingsdata toekende aan zes verschillende klassen. Zoek de verschillen:

- "pimpelmees" ≠
"Cyanistes caeruleus (Linnaeus, 1758)"
- "Cyanistes caeruleus" ≠
"Cyanistes caeruleus (Linnaeus, 1758)"
- "Parus caeruleus Linnaeus, 1758" ≠
"Cyanistes caeruleus (Linnaeus, 1758)"
- "Cyanistes caeruleus (Linnaeus 1758)" ≠
"Cyanistes caeruleus (Linnaeus, 1758)"
- "Cyanistes caeruleus (Linnaeus, 1758)" ≠
"cyanistes caeruleus (linnaeus, 1758)"
- "Cyanistes caeruleus (Linnaeus, 1758)" ≠
"Cyanistes caeruleus (Linnaeus, 1758)"

Van oude namen naar huidige namen

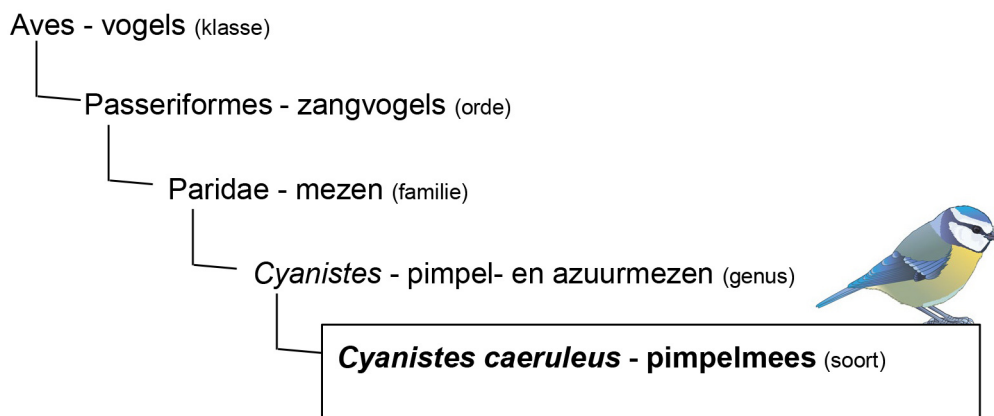
Om een model te realiseren dat niet dergelijke oneigenlijke klassen bevat en dat daarnaast ook geschikt is voor gebruik door andere musea, werden de soortnamen - de labels - eerst gestandaardiseerd en gematcht met behulp van biodiversiteitsdata-infrastructuren als Global Biodiversity Information

Facility (GBIF) en World Register of Marine Species (WoRMS).⁴ Dergelijke infrastructuren bevatten gestructureerde informatie over de geschiedenis van (wetenschappelijke) naamgeving van soorten. Bovendien bieden ze webservices aan voor het herleiden van oude wetenschappelijke namen en andere synoniemen tot huidige geaccepteerde wetenschappelijke namen.

Met dergelijke webservices kun je geautomatiseerd bijvoorbeeld de oude wetenschappelijke naam van de pimpelmees - *Parus caeruleus* Linnaeus, 1758 - herleiden tot de huidige naam *Cyanistes caeruleus* (Linnaeus, 1758). Dit noemen we *name resolving*. In ons project heeft de betrokken software-ontwikkelaar met behulp van deze infrastructuur een *name resolver* ingebouwd in de *pipeline* voor het trainen van een model. Hiermee konden we verschillende schrijfwijzen samenvoegen tot één naam per klasse.

Een tweede aanleiding om een *name resolver* in te zetten is het uniformeren van data afkomstig uit meerdere instituten. Ook wanneer de data per instituut wél goed gestructureerd is, kan het alsnog nodig zijn om een *name resolver* in te zetten om tot eenduidige klassen te komen. Elke collectie hanteert een bepaalde taxonomie (gebaseerd op een bepaald wetenschappelijk standaardwerk of database) en bijbehorende wijze van naamgeving in het collectieregistratiesysteem. Hierdoor was het bijvoorbeeld bij het ontwikkelen van een model voor de kegelslakken (Conidae) noodzakelijk om de in de vijf collecties verschillend gehanteerde naamgevingen te matchen middels een *name resolver*, waarvoor zowel data van GBIF als WoRMS werden gebruikt.

⁴ GBIF: <https://www.gbif.org> en WoRMS: <https://www.marinespecies.org>.



De taxonomische classificatie van de pimpelmees.



Het fotograferen van vogeleieren.

4. Modellen trainen

Beelden als trainingsdata voor model

Afhankelijk van het type object is een bepaald minimum aantal beelden per klasse nodig voor het trainen van een model. Daarbij is het absolute minimum altijd twee, één afbeelding om te trainen en één om het getrainde model mee te toetsen. De gebruikte software 'kijkt' naar de hele afbeelding, inclusief achtergrond, etiketten, kleurkaarten etc. en is niet in staat om onderscheid te maken tussen belangrijke en minder belangrijke onderdelen, zoals een specialist dat wel kan. Hoe meer beelden per klasse hoe beter het model presteert, tot maximaal een paar duizend afbeeldingen per klasse. Van belang is wel dat deze beelden zorgen voor een betere dekking in de variatie binnen de klasse; extra beelden die bijna identiek zijn hebben weinig toegevoegde waarde. Daarnaast is ook de balans in een model belangrijk; in de ideale situatie is de trainingsdataset van iedere klasse ongeveer even groot.

Naturalis aan de slag met de techniek

Voor het trainen en ontsluiten van herkenning-modellen voor dit project heeft Naturalis een software-ontwikkelaar uit eigen ICT-gelederen ter beschikking gesteld die in nauwe samenwerking met het projectteam aan de slag is gegaan. Naturalis heeft als doel gesteld om meer in huis kennis op te bouwen over AI bij de ICT-afdeling voor ontwikkeling en beheer van applicaties, waarmee externe inhuur op dit specialisme niet voor de hand lag. De ontwikkelaar op dit project heeft gedurende het project gespard met een aan Naturalis gedetacheerde AI-specialist van Intel, die zelf niet aan ons project meewerkte maar wel beschikbaar was voor kennisoverdracht. Zo garandeerden we

kennisborging in de organisatie, wat van belang is voor duurzaam onderhoud van de *deliverables* uit dit project en opstart van eventuele nieuwe AI-projecten na dit gesubsidieerde project.

Gebruikte software en hardware

Het trainen van modellen in dit project is gedaan met behulp van TensorFlow, het platform van Google voor *machine learning*.⁵ De code is geschreven in Python 3, gebruik makend van Keras, een open source-pakket dat fungeert als interface voor TensorFlow.

Alle *preprocessing* van de data (downloaden, bestanden lezen, etc.) is door de bij dit project betrokken software-ontwikkelaar geschreven in Python of in Bash-scripts, gebruik makend van standaard Linux-tools. Dit geldt ook voor analyse en presentatie van de resultaten.

Het experimenteel geautomatiseerd bijnijden van afbeeldingen vond plaats met ImageMagick. Alle modellen maken gebruik van de InceptionV3-architectuur. Er zijn tests gedraaid met andere architecturen, zoals VGG16, ResNet50 en Xception. Alle getrainde modellen zijn opgeslagen in een S3-bucket bij Amazon in de cloud.

5. Modellen evalueren en hertrainen

Bij elk model werd een percentage van de beschikbare data uit de trainingsset apart gehouden om na het trainen te kunnen gebruiken voor het testen van het model. Bij een aantal modellen is daarnaast nog

⁵ De versies van TensorFlow die zijn gebruikt zijn 2.0.1, en later 2.7.0. Zie ook <https://www.tensorflow.org/>.

aanvullend materiaal verzameld om het model mee te testen, bijvoorbeeld vergelijkbaar materiaal uit collecties van andere musea. Ook werden bij twee modellen de determinaties die door het model waren gegenereerd voorgelegd aan soortspecialisten die de determinaties van het model beoordeelden. Bij het model voor Zuidoost-Aziatische dagvlinders leverde deze controleslag ook weer nieuwe trainingsdata op. Wanneer de specialisten bepaalde determinaties als incorrect aanduiden en deze correcties vervolgens soorten betroffen die het model ‘nog niet had gezien’ leverde deze evaluatie nieuwe klassen op voor een volgende versie van het model. Doordat het aantal klassen in het model werd uitgebreid, verbeterde de dekking van het model ten opzichte van het totaal aantal soorten in de collectie.

Terminologie beeldherkenning

- **Classes (klassen)** De categorieën die het model herkent: een reeks van waarden, met bepaalde kenmerken, te onderscheiden door (visueel) onderscheidende eigenschappen.
- **Example** Regel in een dataset. Een *example* bevat een of meer kenmerken en mogelijk een label. Een model leert van gelabelde examples.
- **Label** ‘Antwoord’ of resultaatgedeelte van een klasse of *example*.
- **Support** Het aantal examples (afbeeldingen en metadata behorende tot de labels) waarmee het model getraind is. De uiteindelijke support is ook afhankelijk van het minimum aantal afbeeldingen dat per klasse wordt gekozen en de balans in de trainingsdata.
- **Recall (gevoeligheid)** Verhouding tussen afbeeldingen die correct geïdentificeerd zijn als X en alle afbeeldingen die X zijn.
Formule:
$$\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$
- **Precision (precisie)** Verhouding tussen afbeeldingen geïdentificeerd als X dat ook echt X is.
Formule:
$$\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$
- **Accuracy** Verhouding tussen aantal correcte voorspellingen en het totale aantal (alle) voorspellingen.
Formule:
$$\frac{\text{correct predictions}}{\text{total number of examples (predictions)}}$$

Beoordelingscriteria

Er zijn verschillende maten om de werking van een herkenningmodel te beoordelen, die in het algemeen worden aangeduid met *precision*, *recall* en *accuracy*. Deze waarden zijn per model beoordeeld en tegen elkaar afgewogen om een uiteindelijke keuze te maken welk model beschikbaar wordt gesteld via de *web interface*.

Accuracy wordt gemeten voor het hele model, *recall* en *precision* kunnen ook per klasse gemeten worden en verschillen. Het kan voorkomen dat een model goed werkt voor sommige klassen waarmee het getraind is, maar slechter voor andere klassen.

Ons model ‘kijkt’ bijvoorbeeld naar vorm, kleur en patroon maar kan geen onderscheid maken in werkelijke grootte van de afgebeelde objecten. Mede hierdoor had het vogeleierenmodel moeite met het onderscheiden van bepaalde eieren die uiterlijk slechts in ware grootte te onderscheiden waren, iets wat het model niet ziet op basis van de foto’s.

Wat mag je verwachten van een model?

Ieder herkenningmodel kent beperkingen. Een model accepteert trainingslabels blind als waarheid. De kritische blik van de specialist ontbreekt en daarmee is juiste determinatie van de input (trainingsdata) voorafgaand aan het trainen belangrijk. Daarnaast heeft een model altijd last van tunnelvisie: vooralsnog kan het alleen herkennen wat het ooit gezien heeft, namelijk de specifieke klassen waarmee het getraind is. Hierdoor ‘weet’ het model niet wat het niet kent. Een zeldzame soort dagvlinder die niet in de trainingsset van het model zat (omdat er bijvoorbeeld geen exemplaar van in de collectie zat) zal bijvoorbeeld niet als zodanig door het model herkend worden.

Bovendien genereert een model *altijd* een of meerdere antwoorden, met verschillende predicties, de waarschijnlijkheid dat het antwoord (de identificatie) correct is. Dit wordt wel het ‘open world’-probleem genoemd. Het model zal hierdoor nooit een antwoord geven als: “dit herken ik niet”. Gooi bijvoorbeeld een foto van een auto in het model getraind op kegelslakken, en het model zal de auto proberen te classificeren als de meest op de auto gelijkende kegelslak. Het model kan zelfs zeer overtuigd zijn van ‘foute’ antwoorden, dat wil zeggen dat de voorspellingswaarde ondanks het onjuiste antwoord hoog is. Gebruikers van de modellen worden in de *interface* ook gewezen op deze beperkingen, onder andere op de FAQ-pagina.

Testen met het museumpubliek

Aanvankelijk wilden we de opgeleverde modellen ook testen met het museumpubliek, bijvoorbeeld bij het aandragen van natuurvondsten ter determinatie

tijdens natuursprekuren in Naturalis en Het Natuurhistorisch. Mede als gevolg van COVID-19-sluitingen hebben we nauwelijks de mogelijkheid gehad de prototypen van de modellen met publiek te testen. Toen de musea eenmaal weer openden, was de publieksinterface nog niet opgeleverd. Hierdoor kon deze bijvoorbeeld niet voor het aflopen van de projectperiode op gebruiksvriendelijkheid worden getest.

We zijn van plan hier de komende tijd alsnog energie in te steken. De maandelijkse natuursprekuren in Naturalis zijn weer opgestart. Enkele van de opgeleverde modellen zijn in theorie geschikt voor het determineren van natuurvondsten (eikapsels roggen en haaien, kegelslakken) of artefactvondsten (maskers). De kans is echter klein dat incidentele bezoekers op bijvoorbeeld een natuursprekuren ook daadwerkelijk een object meenemen dat behoort tot één van die modellen. Daarom zou het ook zinvol zijn om een aparte bijeenkomst te plannen met bijvoorbeeld amateurverzamelaars (kegelslakken en maskers). Daarbij is het dan raadzaam om meegebrachte vondsten zowel door het model als door een aanwezige specialist te laten determineren, zodat we de werking goed kunnen beoordelen en vergelijken.

Van collectie naar model

1. Kegelslakken (Conidae)

Kegelslakken (ook wel Conidae genoemd) zijn carnivore zeeslakken waarvan de fraaie huisjes (schelpen) veel verzameld worden. Voordeel

van kegelslakkenhuisjes voor een pilot met beeldherkenning is dat de huisjes vaak een contrastrijk patroon bevatten, een heldere vorm hebben en weinig variatie binnen een soort vertonen. Een mogelijk nadeel is dat de vormvariatie tussen de verschillende clusters binnen deze soortgroep niet altijd heel duidelijk is.

Collecties combineren

Voor het model ter herkenning van de kegelslakken werd in eerste instantie gebruik gemaakt van de collecties van Naturalis en projectpartner Het Natuurhistorisch in Rotterdam. Deze collecties waren beide al op naam gebracht door specialisten van deze instituten en tevens gefotografeerd en gedigitaliseerd. De collectie van Het Natuurhistorisch was bovendien bijzonder uniform gefotografeerd, met ook nog twee aanzichten per specimen. Deze collectie was al gepubliceerd op GBIF, maar omdat de hier gepubliceerde beelden te klein werden bevonden om een model mee te trainen, heeft de conservator van Het Natuurhistorisch voor het project de hogere resolutie beelden voorzien van de juiste naamgeving ten behoeve van de classificatie van het model.

Experimenteren met collectiefoto's

In de foto's van de collectie van Naturalis was meer variatie in beelden, vooral ook met betrekking tot afgebeelde artefacten, zoals verschillende etiketten en kleur- en afmeting-indicatoren. Om die reden is nog een test gedaan met het geautomatiseerd



Lade met Conidae.



Conus ammiralis (collectie Het Natuurhistorisch).

uitsnijden van de specimens uit deze foto's.⁶ Het *croppen* van de beelden in de trainingsdataset betekent echter ook dat alle latere input zal moeten worden gecropt en dit ook dient te gebeuren bij gebruik van het model, intern of door derden.

Buitenlandse collecties als aanvulling

Omdat de dekking van deze gecombineerde dataset - het aantal klassen of soorten - ten opzichte van het totale aantal bestaande soorten kegelslakken nog relatief laag was, gingen we in samenwerking met specialisten op zoek naar andere als betrouwbaar beoordeelde geïdentificeerde en gedigitaliseerde collecties. Zo werd contact gelegd met de conservatoren van het *Muséum national d'Histoire naturelle* in Parijs, dat een belangrijke collectie Conidae beheert, die bovendien goed gedetermineerd en gedigitaliseerd is. Het Parijse museum werkte graag mee aan de pilot, en met het toevoegen van deze dataset aan de trainingsdata werd de support van het model vergroot.

Via het museum in Parijs kregen we een buitengewoon genereus aanbod van de voorzitter van de Franse malacologische vereniging om ook zijn gedetermineerde en gedigitaliseerde privé-collectie te gebruiken om het model te verbeteren. Eerder had de verzamelaar deze collectie ook beschikbaar gesteld aan het AI-platform Fieldguide.⁷ Tenslotte

6 Met behulp van ImageMagick: <https://imagemagick.org/>

7 Kegelslakken op AI-platform Fieldguide: <https://fieldguide.ai/fab444/categories?category=529e33d2389a5e0d-d48cea5a>



Conus pulicarius (collectie Naturalis).

werd nog een bescheiden dataset afkomstig uit de collectie van het National History Museum London na opschoning van niet-relevante beelden toegevoegd aan de trainingsdataset.

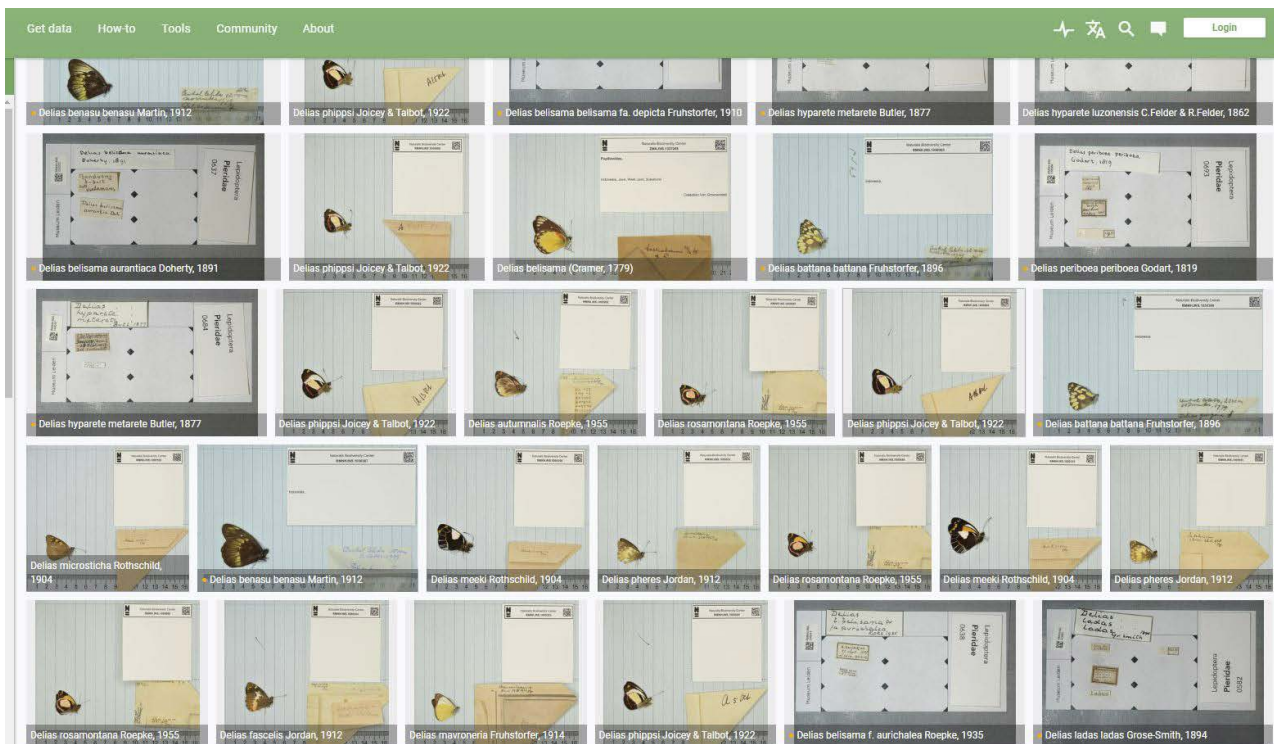
Deze samenwerking en datadonaties resulteerden in een bijzonder grote dataset van bijna 16.000 beelden afkomstig uit vijf collecties (vier museale en een particuliere). Elke collectie hanteert een bepaalde taxonomie en wijze van naamgeving in het registratiesysteem. Hierdoor was het ook bij de Conidae noodzakelijk om de verschillende gehanteerde naamgevingen gelijk te trekken middels de *name resolver* en handmatige opschoning. Behalve de privécollectie zijn al deze collecties te raadplegen op GBIF.⁸

2. Zuidoost-Aziatische dagvlinders (papillotten)

Samenwerking met Papillottenproject

In dit project gingen we aan de slag met een grote collectie Zuidoost-Aziatische dagvlinders uit Naturalis, ook wel aangeduid als de Papillotten-collectie. Dit is een grote deelcollectie gevouwen vlinders in kleine papieren of pergamiijnen zakjes in

8 De betreffende vier museale Conidae-collecties op GBIF: https://www.gbif.org/occurrence/gallery?publishing_org=396d5f30-dea9-11db-8ab4-b8a03c50a862&publishing_org=19456090-b49a-11d8-abeb-b8a03c50a862&publishing_org=3ead5cf3-2840-482e-bbd9-76380825be63&publishing_org=2cd829bb-b713-433d-99cf-64bef11e5b3e&taxon_key=6779



Papillotten, collectie Naturalis op GBIF.

de collectie van Naturalis. Deze vlinders zijn vooral, maar niet uitsluitend, afkomstig uit Zuidoost-Azië en voornamelijk dagvlinders, hoewel de collectie ook nachtvlinders bevat. Deze collectie wordt gedigitaliseerd en geregistreerd met behulp van vrijwilligers. Dit zogenoemde 'Papillottenproject' wordt door Naturalis uitgevoerd met ondersteuning van de Van Groenendael-Krijger Stichting.⁹ De in het Papillotten-project gemaakte foto's worden onder een Creative Commons Zero-licentie (CC-0) ontsloten via Naturalis Bioportal¹⁰ en op GBIF¹¹, en kunnen vrij worden hergebruikt.

Soortenrijke collectie

Deze collectie bestaat uit een grote hoeveelheid verschillende soorten, verdeeld over verschillende genera en families. De vlinders (orde Lepidoptera) in deze collectie behoren grotendeels tot de superfamilie Papilionoidea die weer onderverdeeld is in verschillende families, subfamilies en genera en daarbinnen in vele verschillende soorten.¹² Voor het determineren van zoveel verschillende soorten is veel kennis nodig, kennis die verspreid is over verschillende taxonomen en specialisten in de wereld.

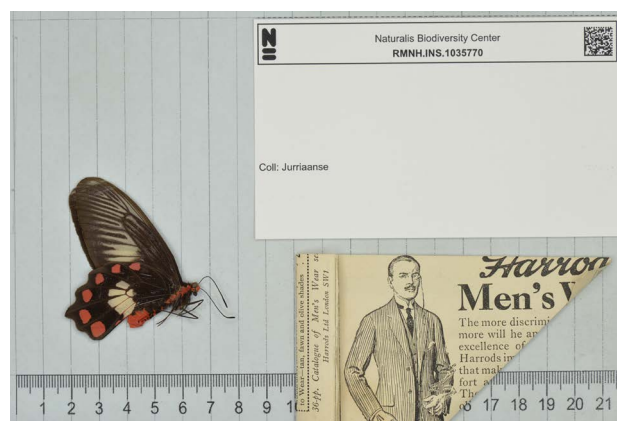


Papillotten: gevouwen vlinders in zakjes.

9 Meer over het Papillottenproject: <https://www.naturalis.nl/collectie/verborgen-vlinders-het-gestroomlijnd-zichtbaar-maken-van-papillotten>

10 Naturalis Bioportal is het venster op de digitale collectie van Naturalis: <https://bioportal.naturalis.nl>

11 Zie onder andere voor alle gedigitaliseerde vlinders in de familie Papilionidae (een deel van de papillottencollectie): https://www.gbif.org/occurrence/gallery?publishing_org=396d5f30-dea9-11db-8ab4-b8a03c50a862&taxon_key=9417



Vlinderspecimen.

12 Wereldwijd zijn er circa 17.268 vlindersoorten bekend, verdeeld over vijf families. Het aantal soorten Zuidoost-Aziatische vlinders wordt geschat op ruim 6.000.

Samenwerking met specialisten

Door de omvang van deze collectie is het onmogelijk om al deze specimens, zo belangrijk voor onderzoek, allemaal 'met de hand' op naam te brengen. Het inzetten van beeldherkenning was daarom een logische stap in de digitalisering en registratie van deze deelcollectie, en resulteerde in een pilot met verschillende modellen als resultaat.

Voor een eerste model werden foto's van gedigitaliseerde vlinders opgestuurd naar verschillende specialisten wereldwijd, die de vlinders tot op soortniveau determineerden. De ontvangen data werd vervolgens geüniformeerd, zodat het model getraind werd met opgeschoonde klassen. Aan de hand hiervan werd een eerste versie van het model getraind. Vervolgens werd een nieuwe set ongedetermineerde vlinders door het model gehaald en werden de AI-determinaties ter verificatie aan de specialisten voorgelegd. De door de specialisten gecorrigeerde onjuiste AI-determinaties werden gekoppeld aan bestaande klassen of als 'nieuwe' soorten (soorten die het model nog niet gezien had) aan de trainingsdata toegevoegd. Zo groeide het aantal klassen en daarmee verbeterde de dekking van het model.

Meer en meer soorten in het model

Doordat de gefotografeerde collectie grotendeels (nog) niet op soortniveau is gedetermineerd, was het niet mogelijk vooraf te selecteren op soorten en zo van elke bestaande soort, ruim 6.000 in totaal, een tweetal foto's aan de trainingsdata toe te voegen. Een model ontwikkelen waarin alle in de collectie aanwezige soorten gerepresenteerd zijn, is daarom een iteratief proces. Dat betekent dat het model steeds verbeterd en uitgebreid dient te worden door het te hertrainen met voor het model 'nieuwe' soorten, van door specialisten aangeleverde identificaties.

Sorteermodellen op hoger niveau

Een deel van de gefotografeerde collectie was weliswaar nog niet gedetermineerd tot op soortniveau maar per verpakkingseenheid wel voorzien van informatie hoger in de taxonomie, bijvoorbeeld op het niveau van genus of subfamilie. Na opschoning en uniformering van deze data werden met die classificaties aanvullende modellen ontwikkeld, met als doel specimens uit specifieke, ondervertegenwoordigde klassen te kunnen sorteren op familie- of genusniveau, om deze vervolgens naar de juiste specialisten te kunnen sturen ter identificatie op soortniveau. Aan de hand van die nieuwe, handmatige determinaties kan het model op soortniveau vervolgens weer verbeterd worden in dekking en dus in de toepasbaarheid in de praktijk.

Zowel de sorteermodellen, op de niveaus van subfamilies en genera, als de laatste versie van het model dat identificeert op soortniveau, zijn toegankelijk via de interface.

3. Vogeleieren

Nederlandse referentiecollectie gedigitaliseerd

Voor het vogeleierenmodel werd de Nederlandse referentiecollectie vogeleieren van Naturalis geselecteerd. Deze bestaat uit op naam gebrachte eieren van soorten vogels die in Nederland voorkomen, en bevat in het algemeen meerdere specimens per soort.

Deze goed gedetermineerde collectie was echter nog niet gedigitaliseerd, en is voor dit project gefotografeerd met speciale instructies ten behoeve van een pilot met beeldherkenning. Voor het project zijn de circa 700 inventarisnummers, elk bestaande uit legfels van één of meerdere eieren, gefotografeerd. Er zijn ruim 2.800 opnamen gemaakt, met daarbij een variatie in achtergrondkleur voor een deelset en daarnaast variatie in aantal specimen(s) per opname, beide met als doel het model te optimaliseren in zowel de werking (mate van herkenning) als de toepassing van het model door andere musea, waar beide factoren eveneens variëren.

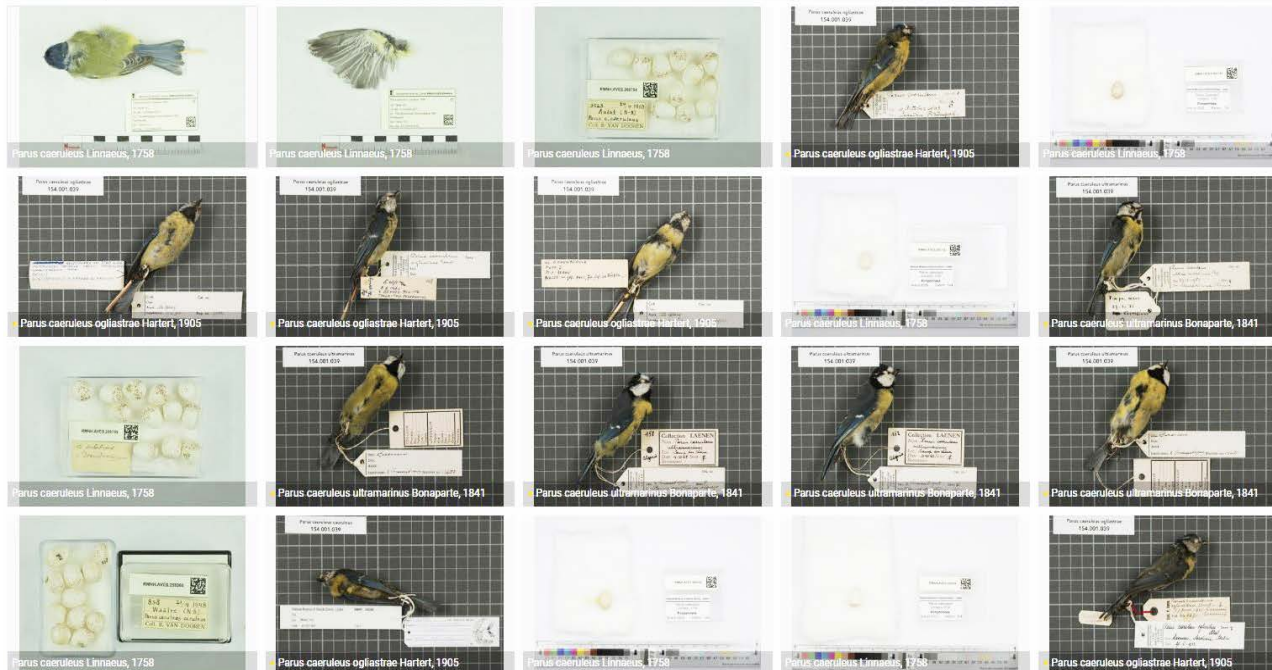
De foto's worden als gedigitaliseerde collectie ontsloten op Naturalis Bioportal onder een Creative Commons Zero (CC-0) licentie en kunnen vrij worden hergebruikt. Ook is deze collectie toegevoegd aan GBIF.¹³



Vogelei uit Nederlandse referentiecollectie.

¹³ De gedigitaliseerde vogeleieren zijn toegevoegd aan de vogeldataset van Naturalis op GBIF: <https://www.gbif.org/dataset/889c91a3-614f-4355-8df8-b6d0260a118c>.

Note that the 'Media type' filter for video or audio does not work while in gallery view



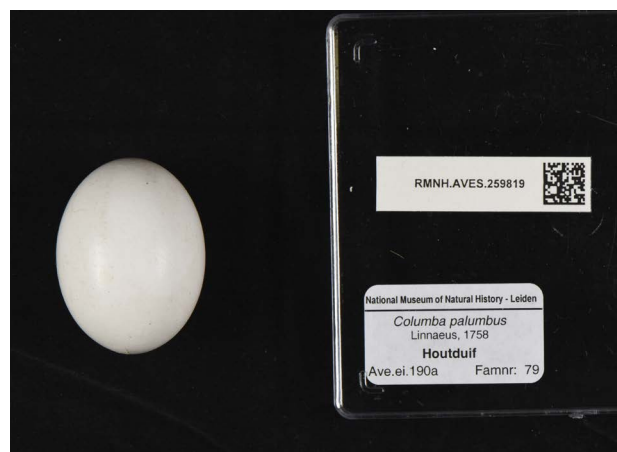
Vogeleieren (pimpelmees), collectie Naturalis op GBIF.

Namen uniformeren

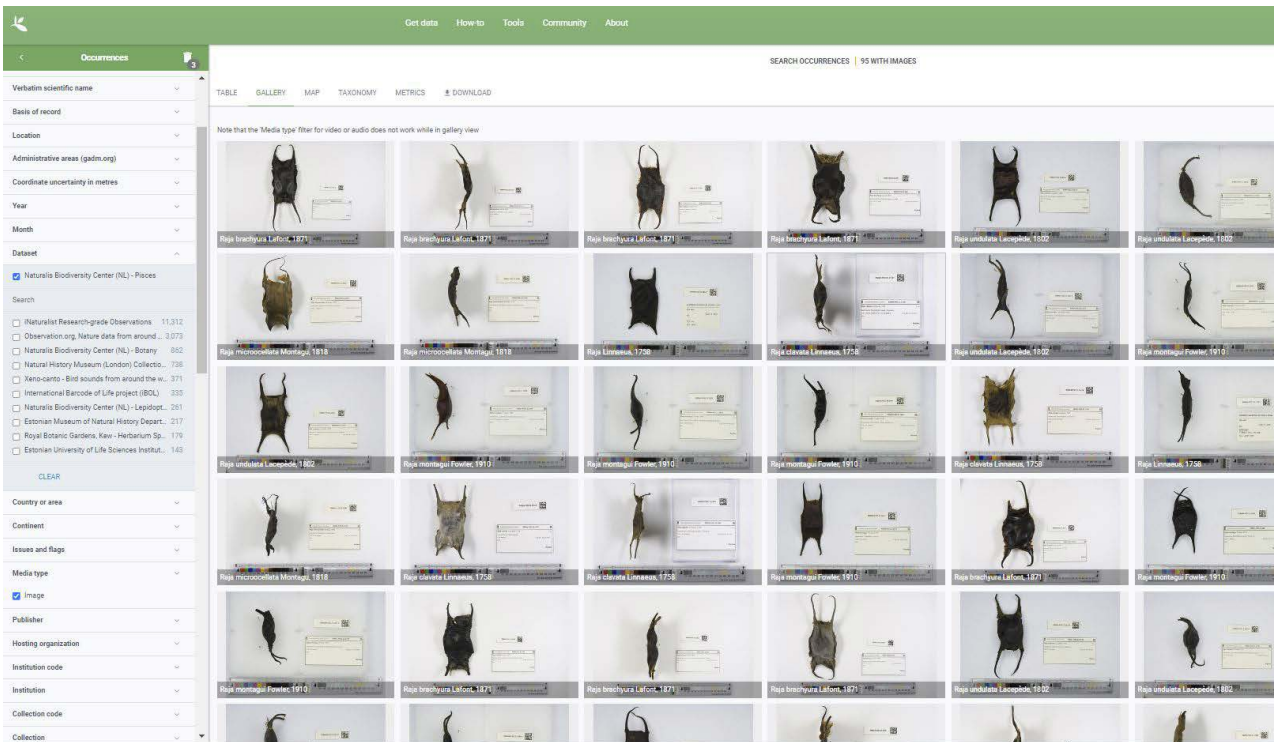
De naamgeving van de soorten in dit model is wederom met een *name resolver* gestandaardiseerd, mede omdat deze collectie deels was geclassificeerd volgens een verouderde soortnamenlijst. Bovendien waren lang niet alle eieren gedetermineerd tot ondersoortniveau. Na consultatie van soortspecialisten is besloten om eventuele ondersoorten niet als aparte klassen op te nemen in het model, maar alle eieren te classificeren op het niveau van de (vogel)soort. Bijvoorbeeld: de staartmees-ondersoorten *Aegithalos caudatus caudatus* (witkopstaartmees) en *Aegithalos caudatus europaeus* (de 'gewone' staartmees) worden samengevoegd tot één klasse, de soort *Aegithalos caudatus* (staartmees).

Collecties combineren

Het uiteindelijke model is getraind op de Nederlandse referentiecollectie, aangevuld met andere afbeeldingen uit de gedigitaliseerde collectie van Naturalis, gefilterd op eieren van in de referentiecollectie voorkomende soorten. In de eerste pilot zijn zes verschillende modellen gedraaid, met verschillende datasets en verschillend minimum aantal afbeeldingen per klasse, wat invloed heeft op de *support*. De toevoeging van circa 1.000 afbeeldingen uit de collectie van Museon-Omniversum leek de prestaties van het model niet substantieel te verbeteren, ondanks dat de *imbalance* wel iets afnam. De mate van *accuracy* heeft de keuze bepaald voor het uiteindelijke model. Voor het testen van het model is een aanvullende



Het verschil tussen het ei van een holenduif en van een houtduif is nauwelijks waarneembaar (collectie Naturalis).



Roggeneieren uit de collectie Naturalis op GBIF.

testset samengesteld via Wikimedia Commons, een collectie museale eieren die vervolgens geschoond werd tot alleen die soorten die het model zou kunnen herkennen, vanwege het 'open world'-fenomeen.

Ei-variatie

De pilot met vogeleieren was door het contrast met de dagvlinders bijzonder leerzaam. Waar de variatie binnen één soort (klasse) vlinder tussen verschillende specimens vrijwel nihil is, kan de variatie in verschijningsvorm tussen eieren binnen één soort bijzonder groot zijn, zelfs binnen één legsel. Anderzijds zijn de eieren van bepaalde verschillende soorten onderling nauwelijks van elkaar te onderscheiden, zeker niet door de herkenningmodellen die wij gebruikten, ook

doordat deze niet naar afmeting kijken. Mogelijk verbetert een substantieel grotere hoeveelheid trainingsdata de werking van het model.

4. Eikapsels roggen en haaien

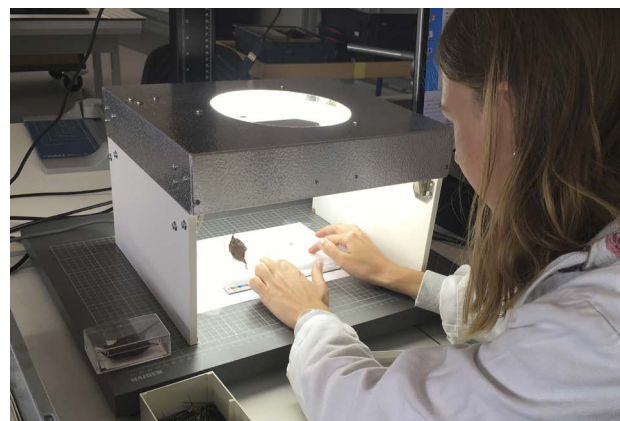
Ook voor gedroogde eikapsels van roggen en haaien aan de Nederlandse kust hebben we een herkenningmodel ontwikkeld. Deze kenmerkende eikapsels, ook wel zeemeerminantasjes genoemd naar het Engelse 'Mermaids purses', kunnen aanspoelen op het strand en de duinen inwaaien.

Een overzichtelijk model

In tegenstelling tot bijvoorbeeld de kegelslakken of dagvlinders, waar sprake is van duizenden soorten,



Binnen een legsel kan de variatie in eieren groot zijn. (Zwarte kraai, collectie Naturalis)



Digitaliseringsteam Naturalis fotografeert roggen-eieren.



Doosjes met verschillende exemplaren roggeneieren.

was het aantal te onderscheiden soorten overzichtelijk. Het model bevat de twaalf soorten die aan de Nederlandse kust aangetroffen kunnen worden.

Eikapsels digitaliseren

Voor het vergaren van trainingsdata voor het model van rogg- en haaieneikapsels werden door de collectiebeheerder gedetermineerde specimens uit de collectie van Naturalis geselecteerd. Het digitaliseringsteam maakte foto's van de specimens, zowel van losse exemplaren als van meerdere exemplaren per inventarisnummer. Ook hier werden vanuit het project speciale instructies opgesteld voor de fotografen. De specimens zijn, met behulp van kleine naalden om de eikapsels mee te fixeren, vanuit verschillende aanzichten - ventraal en dorsaal - gefotografeerd om karakteristieke kenmerken vast te leggen.

De foto's van de eikapsels, voorzien van QR-code en inventarisnummer, zijn door de collectiebeheerder gekoppeld aan het registratiesysteem en deze zijn ontsloten via onder andere Naturalis Bioportal en GBIF onder een Creative Commons Zero-licentie (CC-0) en kunnen vrij worden hergebruikt.

Voor het testen van het model zijn - voor zover aanwezig en gedetermineerd - andere specimens uit de collectie gefotografeerd. Dat gebeurde met zowel een smartphone als een compactcamera, om de werkwijze van en gebruik door het publiek mee te simuleren.

Gedroogde en natte eikapsels

Het model werd getraind op gedroogde exemplaren uit de collectie van Naturalis, maar met het idee om deze zowel geschikt te maken voor herkenning van museale specimens als voor vondsten van het publiek. Voor het determineren van publieksvondsten bestaat o.a. ook een zoekkaart ontwikkeld door de Dutch Shark Society, waarbij wordt geadviseerd het exemplaar nat (geweekt) te



Specimen kleinoogrog (*Raja microocellata*, collectie Naturalis).

meten en te determineren.¹⁴ Voor het determineren van museale exemplaren is weken in water geen optie. Ook voor door het publiek naar het museum meegebrachte (droge) eikapsels kan het model gebruikt worden, bijvoorbeeld op een natuursprekkuur.

5. Maskers

Cultuurhistorische collectie

Het project beoogde de inzet van beeldherkenning in museale collectieregistratie te verkennen. Om die reden is ook specifiek een niet-natuurhistorische collectie meegenomen in de pilots. Projectpartner Museon-Omniversum bracht hiertoe de collectie etnografische maskers in. Deze collectie bestaat uit diverse typen maskers uit verschillende continenten en uiteenlopende culturen.

Voor het trainen van een eerste versie van een herkenningmodel werd de collectie van Museon-Omniversum ingedeeld in afbeeldingen per type masker per cultuur. De software-ontwikkelaar trainde met deze relatief kleine dataset een model op niveau van type, maar ook op hogere niveaus zoals regio (West-Afrika, Centraal-Afrika, etc.) en continent. Die laatste twee experimenten hebben we enerzijds gedaan om te kijken of dat interessante resultaten zou opleveren, en anderzijds om simpelweg meer afbeeldingen per klasse te hebben als trainingsmateriaal.

Grote verscheidenheid

Anders dan bijvoorbeeld bij de vlinders, vertonen maskers ook binnen een type of klasse vaak een visuele verscheidenheid. Anderzijds zijn er ook weer

¹⁴ De *Determinatiegids voor eikapsels van haaien, vleten en rogg* is onderdeel van de Grote Eikapseljact georganiseerd door de Dutch Shark Society: <https://www.dutchsharksociety.org/wp-content/uploads/2022/01/DeGroteEikapselJacht.pdf>.

typen maskers die weliswaar uit heel verschillende streken en culturen afkomstig zijn, maar desondanks sterk op elkaar lijken. Dat betekent dat het vereiste minimum aantal afbeeldingen per klasse voor het trainen van het model groter dient te zijn dan bij bijvoorbeeld de vlinders.

Tijdens het trainen van meerdere versies (iteraties) van ieder model zijn we op zoek gegaan naar de optimale instelling van de parameters waar je tijdens het trainen mee kan sturen. Om de resultaten van het model verder te verbeteren hebben de collega's van Museon-Omniversum extra data verzameld, onder andere uit de collecties van het Nationaal Museum van Wereldculturen. Daarnaast is nog extra trainingsmateriaal online verzameld, beoordeeld en gecategoriseerd tot de bestaande klassen in het model. Met deze data zijn de modellen hertraint en vervolgens getest met een online verzamelde dataset.

Modellen ontsluiten voor gebruik

Gebruikersinterface

Het openstellen van de modellen voor gebruik door andere partijen, zoals musea, verzamelaars etc., was een van de doelstellingen van het project. Hiertoe is een gebruikersinterface ontwikkeld

waar de modellen kunnen worden gebruikt om specimens te identificeren. We hebben uiteindelijk, mede in overleg met publieksafdeling Educatieve Ontwikkeling van Naturalis één *interface* gebouwd voor alle doelgroepen genoemd in de projectaanvraag.

De *interface* is onder vermelding van projectpartners en financiers gepubliceerd als website op <https://museum.identify.biodiversityanalysis.nl/>.

Voor de verschillende pilots is per model een pagina gemaakt met een omschrijving van het model en een mogelijkheid om het model te gebruiken. Daarbij is ieder model voorzien van een link naar een lijst met de klassen waarmee het model is getraind. Ook is de interface voorzien van instructies voor het gebruik alsmede voor de interpretatie van de getoonde resultaten en een veelgestelde vragen-pagina (FAQ). De website is functioneel in beheer gebracht bij de Naturalis-afdeling Collectie Informatie. Er zijn nog enkele bekende verbeterpunten in de interface die de komende tijd door Naturalis opgelost zullen worden.

Open voor zover ethisch verantwoord

In samenspraak met de projectpartners is er voor gekozen het vogeleierenmodel niet openbaar te zetten; dit model is afgesloten voor publiek om



Masker in collectie Museon-Omniversum.

Tabel: Beschikbare modellen per collectie

Model	Herkomst datasets	Interface link
Zuidoost-Aziatische dagvlinders (soortniveau)	Naturalis	https://museum.identify.biodiversityanalysis.nl/models/papilionidae
Zuidoost-Aziatische dagvlinders (genusniveau)	Naturalis	https://museum.identify.biodiversityanalysis.nl/models/papilionidae_genera/classes
Zuidoost-Aziatische dagvlinders (subfamilieniveau)	Naturalis	https://museum.identify.biodiversityanalysis.nl/models/papilionidae_subfamilies
Kegelslakken (Conidae)	Het Natuurhistorisch; Naturalis; Muséum national d'Histoire naturelle (Parijs); Privécollectie (Frankrijk); Natural History Museum (Londen)	https://museum.identify.biodiversityanalysis.nl/models/conidae
Maskers (typeniveau)	Museon-Omniversum	https://museum.identify.biodiversityanalysis.nl/models/maskers/classes
Maskers (continentniveau)	Museon-Omniversum	https://museum.identify.biodiversityanalysis.nl/models/maskers_continent
Vogeleieren Nederland	Naturalis Museon-Omniversum	niet publiek toegankelijk
Eikapsels roggen en haaien Nederland	Naturalis	https://museum.identify.biodiversityanalysis.nl/models/haaien_roggen_eieren

nestverstoring en -roof niet aan te moedigen (dit is bij wet verboden). Het model is wel op aanvraag bij Naturalis te gebruiken door instellingen, zoals natuurhistorische musea.

API en open source code

De herkenningmodellen zijn behalve via een website ook via een openbare API beschikbaar gemaakt.¹⁵ Met deze API kunnen bijvoorbeeld informatici, data scientists en softwareontwikkelaars de herkenningmodellen aansluiten op andere programma's en applicaties, of in bulk afbeeldingen door een model laten identificeren.

De code van de website, de API en de trainings-pipeline voor de herkenningmodellen zijn conform het Naturalis-beleid omtrent open source gepubliceerd als repository op Gitlab.¹⁶

15 Documentatie API: <https://museum.identify.biodiversityanalysis.nl/site/page/api>. Deze API is geschreven in Python 3, en maakt gebruik van Flask en Unicorn voor het webserver-gedeelte. TensorFlow en Keras worden gebruikt voor voor het deel dat de inference doet. Ook de API draait op Linux (specifiek Ubuntu 20.04).

16 Zie <https://gitlab.com/naturalis/bii/beeldherkenning-museumproject> (API, trainingspipeline) en <https://gitlab.com/naturalis/bii/beeldherkenning-museumproject-gui> (website).

Het technisch beheer is ondergebracht bij de Naturalis-ICT-afdeling Applicatieontwikkeling en -Beheer (zie tabel).¹⁷

Verdiepende soort- en objectinformatie

De determinaties die onze modellen leveren, hebben we verrijkt met verdiepende soort- en objectinformatie. Hieronder verstaan we in dit project korte teksten (beschrijvingen) over soorten en objecttypen vertegenwoordigd in de verschillende modellen. Hoe zien deze soorten of objecten eruit? Waar komen ze voor? Tot welke familie of groep behoren ze? Antwoorden op dergelijke vragen helpen de gebruiker om determinaties die modellen genereren te verifiëren en om hun kennis te verdiepen (een vorm van *capacity building*).

Samenwerking

Voor het schrijven en vastleggen van deze beschrijvingen zijn instructies opgesteld en drie projectmedewerkers ingehuurd. De projectmedewerkers zijn aan de slag gegaan met aanvullende

17 Deze website is geschreven in PHP, gebruikmakend van het CodeIgniter-framework (v4) en een MariaDB-database. De website draait op Linux (Ubuntu 20.04).

informatie voor de kegelslakken (Conidae), Zuidoost-Aziatische dagvlinders (Papilionoidea) en vogeleieren. We maakten daarbij gebruik van gevalideerde (literatuur)bronnen. De teksten voor roggen- en haaieneikapsels zijn opgepakt door de projectcoördinator, voortbordurend op eerder opgeleverde teksten.¹⁸ Aanvullende informatie over de maskers is opgesteld en beschikbaar gemaakt door Museon-Omniversum.

Duurzame opslag en beheer

Informatie voor dit project hebben we zoveel mogelijk toegevoegd aan bestaande soort-informatie-databases van Naturalis¹⁹ en van daaruit gekoppeld aan de gebruikersinterface voor de herkenning-modellen. Op deze manier bouwden we voort op bestaande databases zoals het Nederlands Soortenregister²⁰ en Malesian Butterflies én is duurzame inbedding in de ICT-infrastructuur van Naturalis gegarandeerd, ook na afloop van dit project.

Teksten kegelslakken (Conidae)

Voor het model voor de kegelslakken is aanvullende informatie verzameld uit literatuur²¹ om de determinaties te voorzien van aanvullende context, ter verificatie van de determinaties. In overleg met een specialist van de Nederlandse Malacologische Vereniging is er voor gekozen om aanvullende informatie over de verspreiding van elke soort en de afmeting van slakkenhuizen van volwassen exemplaren op te nemen in een Linnaeus-database²² en deze te tonen bij de determinaties om deze zodoende te kunnen verifiëren.

Teksten Zuidoost-Aziatische dagvlinders

Voor het model voor de Zuidoost-Aziatische dagvlinders is veel gedetailleerde soortinformatie beschikbaar bij de determinaties op soort- en ondersoortniveau en deze zal in de toekomst nog verder worden uitgebreid.

Binnen dit project is uitgebreide informatie over onder andere uiterlijk (morfologische kenmerken) en verspreiding van zo'n 300 soorten uit de subfamilie Pieridae verzameld uit geselecteerde specialistische literatuur en vervolgens gestructureerd en herschreven door een projectmedewerker.

18 Afkomstig uit het in 2016 door Naturalis opgeleverde project 'Dierendeterminatie', dat mede werd gefinancierd door Mondriaan Fonds en Prins Bernhard Cultuurfonds.

19 Draaiende op het door Naturalis ontwikkelde systeem voor soortinformatiemanagement Linnaeus: <https://linnaeus.naturalis.nl/>

20 Nederlands Soortenregister: <https://www.nederlandsesoorten.nl>

21 De voornaamste bron is het tweedelige werk van C. Roux, E. Monnier, L. Limpalaer, en A. Robin: *A Taxonomic Iconography of the Living Conidae*, uit 2018.

22 <https://conidae.linnaeus.naturalis.nl>

Door een samenwerking aan te gaan met het lopende, door de Van Groenendael-Krijger Stichting gefinancierde, Maleise dagvlinder-project is deze informatie duurzaam toegevoegd aan de Linnaeus-database Malesian Butterflies.²³ Ook is informatie over soorten behorende tot andere families vanuit het Maleise dagvlinder-project beschikbaar gesteld om te koppelen aan de determinaties van onze modellen. Zo is de informatie over de soorten in de familie Papilionidae die binnen dit lopende project al was verzameld beschikbaar voor het beeldherkenningsproject en gekoppeld aan de determinaties. Beide projecten versterken elkaar zo. In de nabije toekomst zal daar de informatie over de soorten uit de familie Hesperidae aan worden toegevoegd, en worden getoond bij de determinaties uit deze familie.

Informatie vlindersoort *Appias cardena*

Morphology

Male

Hindwings The hindwing underside has its basal half uniformly yellow coloured and the yellow markings are divided by black veins. The yellow markings become white-ish near the marginal border (Yata and Morishita, 1979).

Female

Females of *Appias cardena perkana* are similar to males, but with heavier and broader black margins of the hindwing underside (D'Abrera, 1982).

Subspecies and variation

Three subspecies are described (D'Abrera, 1982; Yata and Morishita, 1979). The males and females of ssp. *cardena* are similar to ssp. *perkana*, but are smaller, with less intense black markings on the underside wing surface (D'Abrera, 1982). Ssp. *hagar* has the apical black markings better developed and the yellow markings on the hindwing underside are paler. Ssp. *perkana* has the yellow marking deeper (Yata and Morishita, 1979).

Similar species

This species is similar to *Appias waltraudae* (Yata and Morishita, 1979).

https://bseai.linnaeus.naturalis.nl/linnaeus_ng/app/views/species/taxon.php?id=808

23 <https://bseai.linnaeus.naturalis.nl/>

Informatie vogelsoort merel (*Turdus merula*)



Herkenning 25 cm. Mannetje eenvoudig herkenbaar aan geheel zwarte verenkleed met gele of oranje snavel en oogring. Vrouwje meer variabel, donker aardbruin, met lichte keel en warmbruine of rossigbruine borst en buik, onduidelijk zwart gevlekt en gestreept. Lange afgeronde staart, vleugels ronder dan van andere lijsters. Juveniel lijkt op adult vrouwje, maar is meer oranjebruin en zwaarder gevlekt en gestreept, met, ook op de bovendelen, lichtgele vlekjes. Meestal alleen of in paren, maar soms op trek in kleine groepjes. Zingt van geëxponeerde zangpost, zoals dak, antenne of boomtop.

Verspreiding en voorkomen Komt voor in Europa, Rusland, rond de Zwarte Zee, Midden-Oosten en Noord-Afrika. Deels trekvogel. In Nederland uiterst talrijke

broedvogel, jaarrond aanwezig, doortrekker en wintergast in uiterst groot aantal.

Biotopen Diverse habitats, zolang er voldoende dekking aanwezig is in vorm van goed bebladerde bomen en dicht struikgewas. Komt voor in open loofbos, parken, groene voorsteden, maar ook hoger in de bergen en in afgelegen gebieden. In dergelijke gebieden zijn de dichtheden lager.

Voedsel Ongewervelden, waaronder veel regenwormen, en vruchten.

Eieren Aantal eieren in legsel meestal 4-5, soms 3-9. Buikig. Glad en glanzend. Lichtblauw meestal. Overvloedig bezet met spikkels en fijne vlekjes in licht roodachtig-bruin, die soms de schaal een algemene bruine tint geven, soms schaarser. Op schaars getekende eieren zijn de tekens groter en zwaarder, bij uitzondering herinneren zij aan de eieren van de beflijster. De tekens staan soms dichter op elkaar aan de stompe pool. Zo nu en dan effen, of met een witte grondkleur. Formaat 29,4 x 21,7 mm.

Geluiden Verschillende roepen, zoals 'pink pink', 'tjsoek tsjoek', dun 'tsrie' en luide ratel. Zang gevarieerd met melodieuze fluitende tonen, niet herhalend zoals zanglijster dat doet.

https://www.nederlandsesoorten.nl/linnaeus_ng/app/views/species/nsr_taxon.php?id=139714

Teksten Nederlandse broedvogels en hun eieren

Voor het model voor de herkenning van vogeleieren zijn gegevens over verspreiding, biotoop, voedsel en habitat van de soort, herkenning van de soort en kenmerken van de eieren en legsels toegevoegd aan het Nederlands Soortenregister, voor vrijwel alle soorten waarop het model getraind is. Deze informatie wordt getoond bij de determinaties die het model genereert. Enerzijds kunnen hiermee de determinaties geverifieerd worden, van belang bij eieren die uiterlijk erg op elkaar lijken maar van verschillende soorten afkomstig zijn, anderzijds om de gebruiker te voorzien van extra informatie over de soort.

Hoewel het vogeleierenmodel niet openbaar toegankelijk is om mogelijke aanzet tot nestverstoring te voorkomen (dit is namelijk bij wet verboden) is de soortinformatie voor eenieder duurzaam toegankelijk gemaakt via Linnaeus en te raadplegen op de website Nederlands

Soortenregister. Uiteindelijk zal het vogeleierenmodel ook achter een afgeschermd login beschikbaar gemaakt worden, waarbij ook de verdiepende informatie gekoppeld zal worden.

Teksten eikapsels roggen en haaien

Informatie over de eikapsels van de twaalf aan de Nederlandse kust voorkomende roggen en haaien was al eerder gepubliceerd in een voorgaand project van Naturalis.²⁴ Deze informatie wordt beheerd en beschikbaar gemaakt via het Nederlands Soortenregister. De eerder verzamelde teksten zijn voor dit project aangevuld met informatie over morfologische kenmerken,

²⁴ Gmelig Meyling, A.W. & M.C. Cadée, 2016. Soortzoeker Eikapsels roggen en haaien van Nederland. Stichting ANE-MOON & Naturalis Biodiversity Center, Leiden. https://determineren.nederlandsesoorten.nl/linnaeus_ng/app/views/matrixkey/index.php?p=eikapselsroggenhaaien



Masker in collectie Museon-Omniversum.

habitat, verspreiding, voedsel etc. van de betreffende soort rog of haai. Deze informatie is zowel gekoppeld aan de determinaties als direct te benaderen via de site Nederlands Soortenregister.

Teksten maskers

Voor het herkenningmodel voor etnografische maskers heeft Museon-Omniversum ruim 70 beschrijvende teksten ontwikkeld bij de verschillende typen maskers. Elke klasse heeft een pagina op de website van het museum, met meerdere afbeeldingen per type masker, en een omschrijving waarin het gebruik in een bepaalde cultuur en de uiterlijke verschijningsvorm van het masker verder wordt toegelicht.²⁵ Deze informatie is gekoppeld aan de determinaties in het maskermodel middels links naar de beschrijvingen op de site van Museon-Omniversum.

²⁵ De website van Museon-Omniversum waarop alle beschrijvende teksten en aanvullende afbeeldingen per type masker worden getoond: <https://beeldherkenning.museon-omniversum.nl/>. De informatie per type masker is gekoppeld aan de determinaties die het maskermodel genereert.



Lat. Aquilegia
Bel. akelei
Gal. ankolis

Akelei, tekening album De Boodt,
in particulier bezit - in bruikleen bij het Rijksmuseum,
RP-T-BR-2017-1-9-44, Publiek Domein.

Crowdsourcing-campagne voor één deelcollectie

Tekeningen Anselmus de Boodt

Projectpartner Rijksmuseum bracht in dit project een collectie tekeningen in om te onderzoeken of het mogelijk was middels kunstmatige intelligentie soorten planten en dieren te herkennen op illustraties. De collectie bestaat uit een aantal albums met tekeningen van dieren, planten en paddenstoelen die tussen 1596 en 1610 gemaakt en verzameld werden door de Vlaamse humanist en kunstenaar Anselmus Boëtius de Boodt. Deze 'Historia Naturalis' is in particulier bezit, in bruikleen bij het Rijksmuseum en ontsloten via de Rijksstudio.²⁶

Beeldherkenning van soorten op illustraties

Deze tekeningen waren weliswaar ten dele voorzien van opschriften over de afgebeelde soorten, maar met deze data kon geen model worden getraind. De opschriften waren geen betrouwbare determinaties van de afgebeelde soorten. Ook was de kwantiteit van het aantal afbeeldingen per soort (klasse) veel te laag voor het trainen van een model.

Voordat kunstmatige intelligentie kan worden ingezet op illustraties van planten- en diersoorten, zijn er voldoende voorbeelden nodig van afbeeldingen met gedetermineerde soorten, bijvoorbeeld uit verschillende perioden en vervaardigd door verschillende tekenaars. Niet alleen variatie in stijl en techniek maar ook in aanzicht en achtergrond zal moeten worden gedekt door een grote dataset van illustraties per klasse.

Crowdsourcing-campagne *Tag the species*

Het determineren van soorten op de tekeningen van De Boodt draagt bij aan het vormen van een dataset waarmee in de toekomst mogelijk een model kan worden getraind dat soorten herkent op illustraties. Om hier een begin mee te maken is voor

de tekeningen uit de collectie van het Rijksmuseum daarom tussen 1 oktober en 31 december 2021 een aparte campagne, *Tag the species*, georganiseerd waarin de dieren, planten en paddenstoelen op de tekeningen van De Boodt met hulp van het publiek werden geïdentificeerd. Iedereen met enige kennis van soorten planten, dieren of paddenstoelen kon meedoen aan het *taggen*: crowdsourcing als opmaat naar een herkenningmodel dus.

Scans van de serie tekeningen waren al geplaatst op Wikimedia Commons²⁷, de beeldbank die onder meer wordt gebruikt voor Wikipedia. Afbeeldingen op Wikimedia Commons kunnen gekoppeld worden aan gestructureerde data uit Wikidata (de database achter alle Wikipedia's), en zo bijvoorbeeld voorzien worden van wetenschappelijke soortnamen. Door deze koppeling wordt informatie over de tekeningen beter gestructureerd en zijn de afgebeelde soorten sneller vindbaar in de beeldbank.

Wikidata

Bijna elk denkbaar onderwerp (een *item*) op Wikidata heeft een uniek nummer, een ID. Zo is bijv. [Q148929](#) de wilde akelei of *Aquilegia vulgaris*. Bij dit zogeheten "Q-item" staan allerlei eigenschappen (*properties*), zoals de wetenschappelijke naam, maar ook de volksnaam in verschillende talen. Ook informatie over de taxon-auteur en -datum en over de zeldzaamheid worden vermeld. De meeste items bevatten links naar Wikipedia-artikelen en naar andere afbeeldingen van de betreffende soort op Commons, maar ook naar bijvoorbeeld het [Nederlands Soortenregister](#) of naar GBIF.

²⁶ Tekeningen Anselmus De Boodt in de Rijksstudio: <https://www.rijksmuseum.nl/nl/zoeken?p=1&ps=12&involvedMaker=Anselmus%20Bo%C3%ABtius%20de%20Boodt&st=Objects&ii=1>

²⁷ Gedigitaliseerde tekeningen De Boodt op Wikimedia Commons: https://commons.wikimedia.org/wiki/Category:Historia_Naturalis_van_Rudolf_II

We maakten voor de uitvoer van de campagne gebruik van de Wiki-infrastructuur. Enkel het hebben of aanmaken van een account voor de Wikimedia-platforms was benodigd voor deelname. Het publiek kon op deze manier kennis maken met deze schitterende tekeningen en tegelijkertijd bijdragen aan het project én aan verschillende Wiki-platforms.

Opschriften en hulp van het publiek

Sommige tekeningen bevatten namen of omschrijvingen uit de tijd van De Boodt (laat zestiende, begin zeventiende eeuw), dus nog voorafgaand aan de classificatie die Carl Linnaeus (1707-1778) gebruikte om planten en dieren in te delen volgens een specifieke naamgeving. Later, in de negentiende eeuw, zijn de tekeningen gedeeltelijk voorzien van de toen gangbare namen van de afgebeelde soort(en) in verschillende talen, maar deze determinaties bleken niet altijd correct. Daarnaast was geregeld sprake van verouderde namen. De vraag aan het publiek was: welke soort of soorten is of zijn er daadwerkelijk te zien?

GLAM-projectpagina en ISA-tool

Voor de campagne werd een Wikipedia-pagina gecreëerd met uitleg voor de deelnemers.²⁸ De campagne zelf werd aangemaakt met behulp van de ISA-tool om gestructureerde data aan bestanden op Wikimedia Commons te koppelen.²⁹ Met deze tool zijn ook de resultaten te downloaden als CSV, wat inlezen in een collectiebeheersysteem mogelijk maakt voor - in dit geval - het Rijksmuseum.³⁰

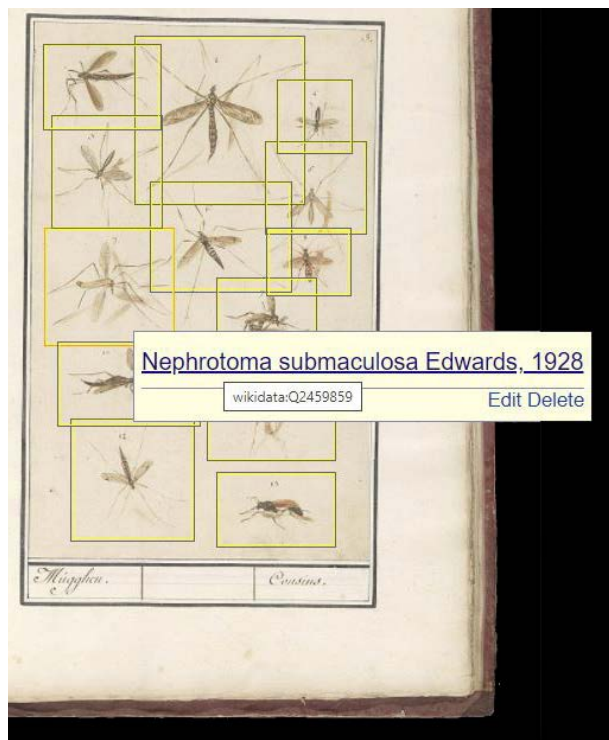
Ook na de campagnelooptijd (einde 31 december 2021) zijn er nog, buiten deze ISA-tool om, identificaties aan tekeningen toegevoegd door via de property P180 (= 'beeldt af') items uit Wikidata te koppelen aan de afbeeldingen. Om ook de Wikidata-items die buiten de ISA-tool om zijn gekoppeld aan de afbeeldingen te exporteren, en van alle items aanvullende gegevens (zogenaamde *properties* zoals *scientific name*, *vernacular name* etc.) te exporteren, is gebruikt gemaakt van de Wikimedia Commons Query Service (Bèta) en SPARQL, waarmee query's kunnen worden uitgevoerd op Wikidata en op de gestructureerde data die zijn gekoppeld aan Wikimedia Commons.³¹

28 Projectpagina Naturalis voor de crowdsourcing-campagne: https://nl.wikipedia.org/wiki/Wikipedia:GLAM/Naturalis/ISA-campagne_soortenherkenning_tekeningen_Anselmus_de_Boodt

29 ISA - Easily Add Information about Images on Wikimedia Commons: <https://isa.toolforge.org/>

30 ISA-campagne *Tag the species*
1 oktober 2021 - 31 december 2021.

31 De aan De Boodt's tekeningen gekoppelde Q-id's met Nederlandse labels via sparql: <https://w.wiki/5Spe>.



Ook meerdere soorten konden op één tekening worden getagged (RP-T-BR-2017-1-6-75, particulier bezit - in bruikleen bij het Rijksmuseum).

Aanzet tot model

De publieksbijdragen zijn niet per definitie *gevalideerde data*, en zijn niet gecontroleerd door soortspecialisten (al waren sommige deelnemers wel degelijk specialist of liefhebber). Voor het ontwikkelen van een model dat in staat is specifieke soorten te herkennen op illustraties, is betrouwbare *labelling* nodig, aan de hand van gevalideerde identificaties. Ook moeten de illustraties kwalitatief goed genoeg zijn: de afbeelding dient een getrouwe weergave te zijn van de afgebeelde soort. Een goede selectie van de trainingsdata op beide aspecten is noodzakelijk voor de ontwikkeling van een dergelijk model, terwijl ook de dataset kwantitatief groot genoeg moet zijn om de onderlinge verschillen in aanzicht, stijl, techniek en dergelijke te compenseren.

Deze verrijkte set van tekeningen van De Boodt is een aanzet om tot een voldoende grote dataset te komen van soorten op illustraties. In combinatie met andere datasets zou in de toekomst mogelijk een model kunnen worden getraind voor het herkennen van specifieke soorten op prenten en tekeningen, op platen en in boeken.³²

32 De KB organiseerde in 2020 een ISA-campagne voor de Nederlandsche Vogelen van Nozemann & Sepp, waarin de afbeeldingen zijn gekoppeld aan de afgebeelde soorten op Wikidata. Deze set zou kunnen worden meegenomen in een herkenningmodel. <https://isa.toolforge.org/campaigns/72>

Communicatie en publiciteit

Projectpagina

Als centrale online plek voor het project is op de website van Naturalis een projectpagina aangemaakt met informatie over de doelstellingen, de pilots, projectpartners, fondsen, en een contactmogelijkheid met de project-leider en projectcoördinator.³³ Voor het afsluitende symposium is een onderliggende pagina aangemaakt met programma en aanmeldmogelijkheid, waar na afloop een samenvatting en de slides van de presentaties van de verschillende sprekers toegankelijk zijn gemaakt. Hier is ook de link naar de interface waar de modellen gebruikt kunnen worden, gepubliceerd.³⁴

Presentaties

Gedurende het project zijn er voor verschillende gremia binnen en buiten Naturalis presentaties gegeven over het project als geheel of over één of meer deelprojecten of specifieke modellen. Vanwege de COVID-19-maatregelen de afgelopen jaren vond het merendeel van deze presentaties online plaats.

Het biodiversiteitswerkveld

Het project werd vlak voor de eerste coronalockdown geïntroduceerd voor het Nederlandse biodiversiteitswerkveld op de jaarlijkse NLBIF-relatiedag op 3 maart 2020 (Jaarbeurs, Utrecht).³⁵

In 2021 en 2022 werden presentaties gegeven voor de online bijeenkomsten van en voor de *interest group* AI4Bio, die gericht is op het kennis delen over de inzet van AI bij natuurhistorische collecties en biodiversiteitsonderzoek. Toehoorders bestaan uit Naturalis-medewerkers en onderzoekers aan verschillende universiteiten die zich met kunstmatige intelligentie bezig houden.

33 Projectpagina <https://www.naturalis.nl/wetenschap/beeldherkenning-als-instrument-voor-museumcollecties>

34 <https://www.naturalis.nl/wetenschap/beeldherkenning-als-instrument-voor-museumcollecties/symposium-beeldherkenning>

35 <https://www.nlbif.nl/wp-content/uploads/NLBIF-relatiedag-20200303-programma-website.pdf>

Andere presentaties binnen Naturalis werden onder andere gegeven voor de sector Collectie tijdens het collectie-colloquium en voor de collega's van het door NWO-gefinancierde ARISE-project waarin beeldherkenning een voorname rol speelt.³⁶

De Wiki-community

Ook buiten de muren van Naturalis werden verschillende presentaties gegeven over het project. Op 13 november 2021 werd het project voor het voetlicht gebracht in een online presentatie voor de Wikimedia-gemeenschap op de WikiconNL, georganiseerd door Wikimedia Nederland. Focus hier was de inzet van tools en data van de Wikimedia-platforms in het beeldherkenningsproject, zoals de ISA-tool voor crowdsourcing en de museale collecties op Commons als testdata, en tegelijkertijd het publiek, deels afkomstig uit de GLAM-gemeenschap, kennis te laten maken met AI en museumcollecties.³⁷

Aandacht voor de vlinders

Specifieke aandacht was er voor het herkenningsmodel van de Zuidoost-Aziatische dagvlinders. In een sessie over het Papillottenproject voor de Van Groenendael-Krijger Stichting, financier van dat project, werd de beeldherkennings-pilot met de *Papilionoidea* gepresenteerd. Tijdens een tweedaagse, internationale workshop georganiseerd door Naturalis voor musea in Indonesië, Papua en de Filipijnen werden de vlinderherkenningsmodellen en de *interface* toegelicht voor toekomstig gebruik door deze musea in de presentatie *Outsourcing species recognition: Automated species recognition*. Het doel van de presentatie tijdens de online *Workshop Malesian Butterflies* op 9 mei 2022 was de natuurhistorische musea in Indonesië en de Filipijnen in het bezit van collecties Zuidoost-Aziatische dagvlinders te laten kennismaken met

36 <https://www.arise-biodiversity.nl/>

37 Presentatie Wikicon 2021: https://commons.wikimedia.org/wiki/File:Presentation_Annika_Hendriksen_Naturalis-Project_AI_museumcollections_WikiconNL_2021_Dutch.pdf

de toepassingsmogelijkheden van de ontwikkelde modellen in de collectieregistratie, en eventueel te participeren in de doorontwikkeling van deze modellen met behulp van datadonaties.

Bekendheid met het project middels deze presentatie en samenwerking met het Maleise dagvlinder-project heeft er toe geleid dat enkele musea al datasets hebben gedoneerd of hebben aangegeven te willen helpen onze modellen met datadonaties te verbeteren in *performance* en dekking.

Aandacht voor crowdsourcing-campagne met Rijksmuseum

Publiciteit in de media werd met name door het crowdsourcing-project met de tekeningen uit het Rijksmuseum gegenereerd. Op de website Nature Today werd tweetalig gepubliceerd over het deelproject met de vroeg zeventiende eeuwse tekeningen waarvoor crowdsourcing werd ingezet om soorten te determineren (*taggen*).³⁸ Vervolgens wijdde ook het BNN-VARA radioprogramma Vroege Vogels er een item aan, opgenomen in het Rijksmuseum.³⁹ Ook de website van Vroege Vogels en de website van Roeg.tv besteedden in november 2021 aandacht aan deze campagne.⁴⁰

Eindsymposium Museumcollecties & AI

Ter afsluiting van het project en met het specifieke doel de opgedane kennis te delen met andere instituten en AI-geïnteresseerden, werd op 16 mei 2022 in Naturalis een symposium georganiseerd dat werd bezocht door circa 70 personen afkomstig uit verschillende disciplines en instellingen uit het hele land.

Tijdens het symposium werden doel, resultaten en *lessons learned* van het project gedeeld met het publiek. De verschillende processtappen van het trainen van een model en de technische achtergrond van beeldherkenning kwamen aan bod. Presentaties over de totstandkoming van het vlindermodel en resultaten van het maskermodel werden gevolgd door lezingen over andere actuele AI-projecten bij het Rijksmuseum en Museum Volkenkunde. Ook kwam AI in het biodiversiteitsonderzoek aan bod, gevolgd door een korte paneldiscussie over het

thema AI in museale context. De presentaties zijn - vanwege auteursrechten in aangepaste vorm - gepubliceerd op de projectpagina op de website van Naturalis.⁴¹

Om het gebruik van de modellen te promoten werden flyers en visitekaartjes uitgedeeld aan de bezoekers met daarop de QR-code van de website (zie bijlage).

Netwerk Digitaal Erfgoed

Het beeldherkenningsproject kreeg ruime aandacht op de website van het Netwerk Digitaal Erfgoed (NDE) waar op 26 april 2022 een interview met de projectcoördinator over dit project werd gepubliceerd.⁴² Ook werden het project en het symposium onder de aandacht gebracht in de nieuwsbrief van NDE, wat resulteerde in een aantal extra aanmeldingen uit het erfgoedveld voor het symposium.

Naar aanleiding van het project en het eindsymposium werd Naturalis gevraagd deel te nemen aan een door het NDE en de NL AI Coalitie georganiseerde middag over AI en de culturele sector. Deze bijeenkomst *Maak jezelf AI-ready! Lancering AI-cursus voor culturele en creatieve professionals* vond plaats op 21 juni 2022 in Teylers Museum. Hier werden praktijkvoorbeelden gepresenteerd aan het publiek, waaronder het project *Automatische beeldherkenning als instrument voor museumcollecties*.⁴³



Eindsymposium Museumcollecties & AI

38 NatureToday-artikel over *Tag the species* gepubliceerd op 23 oktober 2021: <https://www.naturetoday.com/intl/nl/nature-reports/message/?msg=28333>

39 Vroege Vogels radio over *Tag the species*, 21 november 2021: <https://www.bnnvara.nl/vroegevogels/artikelen/natuurkenners-gezocht-herken-de-soorten-op-tekeningen-uit-het-rijks>

40 Roeg.tv over *Tag the species*: <https://www.roeg.tv/roeg-nieuws/herken-jij-deze-illustraties-van-plant-en-diersoorten/>

41 <https://www.naturalis.nl/wetenschap/beeldherkenning-als-instrument-voor-museumcollecties/symposium-beeldherkenning>

42 <https://netwerkdigitaalerfgoed.nl/nieuws/kun-je-met-beeldherkenning-de-collectieregistratie-versnellen-naturalis-onderzoekt-de-mogelijkheden/>, 26 april 2022.

43 Opnamen van de bijeenkomst zijn door NDE geplaatst op Youtube Lancering AI Cursus voor de Creatieve Industrie - 6 Panelgesprek en op de website van NDE: <https://netwerkdigitaalerfgoed.nl/nieuws/maak-jezelf-ai-ready-volg-de-online-ai-cursus-voor-culturele-en-creatieve-professionals/>

Terug- en vooruitblik

Wat hebben we geleerd?

We hebben met dit project goede stappen kunnen maken met het koppelen van de praktijk van museale en vooral natuurhistorische collecties en de mogelijkheden van beeldherkenning. We leerden veel over de mogelijkheden en de moeilijkheden. We ondervonden andermaal dat ‘schone’ en gestructureerde collectiedata van uiterst belang is als je met digitale collecties aan de slag gaat. We zagen dat inbedden in dagelijkse processen zoals digitalisering alleen werkt door nauw samen te werken met de betrokken collectiebeheerders, registratoren én specialisten, en hun vragen en wensen goed te leren begrijpen.

We zijn van mening dat beeldherkenning de specialist niet zal vervangen; beeldherkenning is in sommige gevallen een zeer bruikbaar determinatiemiddel als *aanvulling* op de specialist. Specialist op het gebied van soorten en andere museale objecten (zoals conservatoren) blijven van groot belang en zijn juist nodig bij het ontwikkelen en evalueren van modellen. Zonder hen wordt het lastig om de inzet van automatische beeldherkenning in museumcollecties tot een succes te maken.

Samenwerking en kennisuitwisseling met andere projecten en instituten is ook essentieel. Er gebeurt inmiddels zoveel op het gebied van *artificial intelligence* en automatische beeldherkenning dat het risico op zelf het wiel uitvinden of werk dubbel doen groot is. Tijdens ons project hebben we behalve met de projectpartners ook binnen Naturalis kennis uitgewisseld (met bijvoorbeeld programma's en projecten zoals ARISE en LegaSea). En we zochten contact met meerdere andere instituten en personen uit binnen- en buitenland. Zij leerden niet alleen van ons, maar wij ook van hen.

De inzet van beeldherkenning in de praktijk van museaal collectiebeheer stopt niet met de resultaten van dit project. Voor Naturalis is automatische beeldherkenning een speerpunt

Algemene leerpunten automatische beeldherkenning en museumcollecties

1. Gestandaardiseerde data - via thesauri en/of identifiers - is onmisbaar
2. Niet alle beeldmateriaal is geschikt
3. Niet alle typen objecten zijn geschikt
4. Sommige soorten/typen behoeven meer trainingsdata dan andere soorten/typen
5. Aanvullende informatie voor definitieve determinatie is soms noodzakelijk
6. Er gebeurt veel op dit gebied: kijk wat er elders gebeurt en wissel kennis uit!
7. Automatische beeldherkenning is een middel, niet altijd een of de beste oplossing...

dat de komende jaren nog in meerdere projecten een voorname rol speelt. We willen de komende tijd nog stappen zetten voor toepassing van de opgeleverde modellen in de praktijk. Ook delen we actief kennis met andere lopende projecten waarin beeldherkenning ook een rol speelt, zodat zij van ons kunnen leren. In de volgende paragrafen lichten we een aantal van de concrete verbeterpunten en andere vervolprojecten van Naturalis en partners toe.

Onze modellen in de werkpraktijk

Verbeteringen in proces

Om de modellen optimaal in te kunnen zetten in de praktijk van collectieregistratie worden de komende tijd nog een aantal verbeteringen verder uitgewerkt. Aan de hand van de feedback van gebruikers

zullen bijvoorbeeld de exportmogelijkheden van de determinatie-output en de robuustheid van de modellen worden verbeterd. Implementatie in het werkproces en een efficiënte strategie voor het hertrainen van de modellen (in afstemming met ICT) worden verder ontwikkeld.

Idealiter wordt in een collectiebeheersysteem vastgelegd dat een determinatie middels AI - in tegenstelling tot een specialist - is vastgesteld, wanneer en met welk model, en daarbij ook het zekerheidspercentage van de determinatie. We willen faciliteren dat deze metadata worden meegegeven in de output van de modellen (*interface* en API). Het doel is een export te genereren dat door musea eenvoudig kan worden ingelezen in hun collectiebeheersysteem. In Naturalis wordt het papillotten-model gebruikt voor de herkenning van een grote collectie gevouwen dagvlinders. Beoogd wordt de metadata van de AI-determinaties te kunnen inlezen in het CRS, het gebruikte collectieregistratiesysteem.

Doorontwikkeling modellen

Daarnaast willen we met name de modellen die ontwikkeld zijn voor de herkenning van de vlinders/papillotten uit Zuidoost-Azië verder doorontwikkelen, omdat we zien dat de digitale collectie van Naturalis daar echt een meerwaarde is op wereldschaal. Door herhaaldelijk nieuwe determinaties toe te voegen aan de trainingsdata worden de dekking van het aantal soorten uitgebreid en de robuustheid van het model verbeterd. Dankzij internationale contacten via het tevens vanuit Naturalis gecoördineerde Maleise dagvlinder-project zijn datasets verworven in Indonesië die het papillotten-model kunnen verbeteren. We verwachten dat dergelijke datadonaties met meer bekendheid van de projecten en actieve doorontwikkeling van de modellen frequenter zullen voorkomen en zowel bijdragen aan de bruikbaarheid van de modellen als aan kennisdeling over de betreffende soorten.

Inzet modellen in publiekstoepassing

Het inzetten van een of meer modellen in een publiekstoepassing was tevens een doel van dit project. Het eikapselmodel van roggen en haaien leent zich goed om publiek met strandvondsten te helpen het eikapsel te determineren, bijvoorbeeld op een natuurspreekuur. Deze toepassing kan zowel op zaal live worden ingezet als online, op de website van Naturalis of via andere kanalen. Ook het kegelslakkenmodel kan interessant zijn voor het grotere publiek, omdat dit type schelp al van oudsher veel verzameld wordt.

AI en Naturalis

Binnen Naturalis lopen verschillende beeldherkenningsprojecten voor diverse soorten collecties en datasets. En hoewel sommige typen objecten of collecties binnen ons project niet tot een herkenningmodel hebben geleid, worden er nieuwe projecten uitgerold waarbij de in ons project opgedane kennis van nut blijkt.

Tijdens het verzamelen en analyseren van data bleek dat een model voor herkenning van mineralen en een pilot met herkenning van botten en botmateriaal niet haalbaar waren binnen de looptijd van het project. De hoeveelheid benodigde data is sterk afhankelijk van het type object, en hoe meer variatie in type, aanzicht en verschijningsvorm, hoe hoger het benodigde aantal beelden. Dit bleek voor beide typen objecten nog onvoldoende en/of ongeschikt.

Fossielen en LegaSea

In het geval van fossielen en botten is de benodigde hoeveelheid data om een werkbaar model te trainen relatief groot, door de kwantiteit aan vormvariatie per soort, als gevolg van 1) het aantal botten per individu (soort), 2) het voorkomen van botfragmenten en zodoende de grote variatie in breuken en breukvlakken, en 3) variatie in aanzichten van deze driedimensionale objecten.

Toch wordt er in de aankomende jaren in Naturalis een beeldherkenningsproject uitgerold voor de herkenning van fossielen en botmateriaal. Het door NWO gefinancierde LegaSea-project betreft het contextualiseren van fossiel botmateriaal afkomstig uit de Noordzeebodem. Aan de hand van *machine learning* (AI) en *citizen science* worden meerdere collecties geanalyseerd om tot een beter inzicht te komen van de fauna uit de laatste ijstijd. Naast de collectie van Naturalis, zullen onder anderen de (digitale) collecties van Het Natuurhistorisch, de Oervondstchecker⁴⁴ en diverse verzamelaars worden meegenomen. De komende vier jaar wordt er gewerkt aan het realiseren van dit project, met als doel een voortlevend product neer te zetten.⁴⁵ Onze in het beeldherkenningsproject opgedane ervaringen, kennis en ideeën hebben we gedeeld met betrokkenen bij het LegaSea-project.

44 De Oervondstchecker werd in 2020 door Naturalis overgenomen van het Havenbedrijf Rotterdam en vormt een rijke en groeiende bron van gedetermineerde fossielen: <https://www.oervondstchecker.nl/>

45 LegaSea: An AI-based citizen-science approach to contextualize and understand ex-situ Quaternary fossils from the North Sea Basin. <https://www.nwo.nl/nieuws/veertien-vernieuwende-onderzoeksprojecten-van-start-open-competitie-enw-m>



Natuur op spreekuur, Naturalis.

Andere ideeën vanuit de organisatie

Uit de enquête die we in september 2021 hielden onder collectiebeheerders en andere collega's (van Naturalis en projectpartners), bleek grote interesse in de toepassing van beeldherkenning in collectieregistratie en publiekseducatie. Er werden diverse collecties aangedragen die mogelijk geschikt zijn om versneld te identificeren en registreren met behulp van beeldherkenning. Ook andere toepassingen werden genoemd, die ofwel bijdragen aan een efficiëntere workflow (zowel collectieregistratie en -ontsluiting als dienstverlening) of aan publiekstoepassingen. Toepassingen die werden genoemd waren collectie sorteren, specimens tellen en vergelijkbare objecten in de collectie vinden. Als externe doelgroepen werden naast museumpubliek en geïnteresseerden in natuur en biodiversiteit ook archeologen en (amateur)paleontologen genoemd. Een toepassing van beeldherkenning buiten de museale context zou bijvoorbeeld detectie door de douane van illegale handel in dieren en planten kunnen zijn.

Kennis en modellen delen

Kennis en ervaringen uitwisselen over AI tussen en binnen erfgoed- en onderzoeksinstituten zal de komende jaren alleen maar belangrijker worden. Binnen Naturalis is dit uitgegroeid tot een terugkerende (online) bijeenkomst met geïnteresseerde medewerkers uit eigen organisatie en van partners van verschillende universiteiten: AI4Bio. De verschillende AI-projecten en initiatieven worden op deze manier in kaart gebracht, belangrijk om te voorkomen dat er geen dubbel werk wordt gedaan maar men elkaar versterkt.

AI verschuift zo langzaam maar zeker van incidentele toepassing naar een structureel onderdeel van werkprocessen en onderzoeksmethoden. Zo wordt

bijvoorbeeld binnen het ambitieuze project ARISE, een initiatief om alle in Nederland voorkomende soorten in kaart te brengen en te kunnen herkennen, op meerdere manieren gebruik gemaakt van automatische beeldherkenning.⁴⁶ Binnen ARISE maar ook daarbuiten is er de roep om modellen én de onderliggende trainingsdatasets centraal te delen, zodat makkelijker uitgewisseld kan worden wat al beschikbaar is en we kunnen voortbouwen daarop.⁴⁷ Met ons project onderschrijven we deze noodzaak. We hebben ook een begin gemaakt om de modellen open beschikbaar te stellen, zodat ze in de toekomst bijvoorbeeld naar een organisatie- of sectorbrede centrale repository overgeheveld kunnen worden. We houden het moment dat dit zich aandient actief in de gaten. Wordt ongetwijfeld vervolgd.

We zijn het Mondriaan Fonds, NLBIF en het Prins Bernhard Cultuurfonds zeer erkentelijk voor hun steun. Met behulp van deze financiële bijdragen hebben we herkenningmodellen ontwikkeld voor verschillende typen collecties. In samenwerking met de projectpartners wisselden we kennis en data uit. Ook konden we met de bijdragen de opgedane kennis delen met de erfgoedsector. We zijn ervan overtuigd dat we met dit project andere instellingen, groot en klein, hebben kunnen helpen of inspireren voor de inzet van AI voor de registratie en ontsluiting van hun collecties.

⁴⁶ Automatische beeldherkenning in het project ARISE: <https://www.arise-biodiversity.nl/post/digital-species-identification>

⁴⁷ Zie bijvoorbeeld Greeff M, Caspers M, Kalkman V, Willemse L, Sunderland BD, Bánki O, Hogeweg L (2022) *Sharing taxonomic expertise between natural history collections using image recognition*. Research Ideas and Outcomes 8: e79187. <https://doi.org/10.3897/rio.8.e79187>



APR - - 1938
E11 - DJAMPEA
Labosan-Mengnatti
Snd. and. Salseler

1938

Bijlage

Flyer en visitekaartje

Coll: Jurriaanse

Beeldherkenning als instrument voor museumcollecties

In dit project onderzoekt Naturalis samen met projectpartners of en hoe kunstmatige intelligentie (AI) kan bijdragen aan het proces van museale objectregistratie, met focus op natuurhistorische collecties in Nederland. In deze collecties is juiste identificatie van de soort of het type object essentieel.

In het project werd dit onderzocht aan de hand van verschillende typen objecten uit meerdere collecties. Op basis van *machine learning* zijn modellen ontwikkeld voor het herkennen van o.a.

- Zuidoost-Aziatische dagvlinders
- Huisjes van kegelslakken (Conidae)
- Eikapsels van roggen en haaien
- Maskers uit verschillende culturen

 **Probeer de modellen zelf uit!**
Ze zijn medio mei 2022 open beschikbaar via museum.identify.biodiversityanalysis.nl

Projectpartners
Naturalis Biodiversity Center (Leiden), Museon-Omniversum (Den Haag), Het Natuurhistorisch (Rotterdam), Natuurmuseum Brabant (Tilburg) en Rijksmuseum (Amsterdam)

Mede mogelijk gemaakt door

Flyer pag. 1.

Programma Symposium Museumcollecties & AI

16 mei 2022 Naturalis Biodiversity Center 13.00-18.30 uur

13.00	13.30	Inloop	
			Automatische beeldherkenning als instrument voor museumcollecties
13.30	13.45	Welkomstwoord en inleiding project	Sander Pieterse, Naturalis
13.45	14.05	Projectresultaten	Annika Hendriksen, Naturalis
14.05	14.20	Technische toelichting herkenningsmodellen	Maarten Schermer, Naturalis
14.20	14.35	Papillotten-project en beeldherkenning	Mónica Guimarães Cruz, Naturalis
14.35	14.50	Beeldherkenning van maskers	Corine Bliëk, Museum-Omniversum
14.55	15.15	Pauze	
15.15	15.35	AI-beeldherkenning en de fotocollectie Nederlands-Indië	Marjolijn van Beelen, Museum Volkenkunde
15.35	15.55	Database en geautomatiseerde beeldherkenning van keurmerken op (Nederlands) zilver	Ellen van Bork, Rijksmuseum
15.55	16.10	Fotoherkenning van wilde planten, dieren en paddenstoelen in ObsIdentify en Waarneming.nl	Laurens Hogeweg, Naturalis
16.10	16.25	Animal sound & AI [English]	Dan Stowell, Naturalis/ Tilburg University
16.30	17.00	Paneldiscussie [English]	Rob Erdmann, Rijksmuseum e.a.
17.00	18.30	Borrel	



Het actuele programma is te vinden op:
naturalis.nl/symposiumbeeldherkenning

Flyer pag. 2.



In dit project onderzoekt Naturalis samen met projectpartners of en hoe kunstmatige intelligentie (AI) kan bijdragen aan het proces van museale objectregistratie, met focus op natuurhistorische collecties in Nederland. In deze collecties is juiste identificatie van de soort of het type object essentieel.



Probeer de modellen zelf uit!

Ze zijn medio mei 2022 open beschikbaar via museum.identify.biodiversityanalysis.nl

Visitekaartje.

