VR4CH: Visual recognition voor erfgoedcollecties

eindrapport

MOMU

./DATABLE

Titel	VR4CH: Visual recognition voor erfgoedcollecties. Eindrapport
Auteur(s)	Henk Vanstappen (Datable) Met dank aan Nastasia Vanderperren (PACKED/VIAA) en Dieter Suls (ModeMuseum Antwerpen)
Versie	1.0 (juni 2019)
Licentie	De inhoud van dit document wordt gepubliceerd onder de Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licentie
	Vlaanderen verbeelding werkt

Inhoud

1. Management samenvatting 3

2. Inleiding 4

3. VRS services 6
- 3.1. Descriptoren 6
- 3.2. Modellen en training 6
- 3.3. Interactie 7
- 3.4. Vereisten beeldmateriaal 8
- 3.5. Kosten 8
- 3.6. Selectie VRS 8

4. Methodologie en uitvoering 11
- 4.1. Selectie materiaal 11
- 4.2. Set-up van de workflows 11

5. Beoordeling 13
- 5.1. Kwantitatieve resultaten 13
- 5.2. Type descriptoren 14
- 5.3. Kwalitatieve beoordeling 14
- 5.4. Ruis 15
- 5.5. Granulariteit 15
- 5.6. Snelheid 16

6. Conclusies 17
- 6.1. Conclusies cases study 17
- 6.2. Inzetbaarheid van VRS voor erfgoedcollecties 17
- 6.3. Use cases 17
- 6.4. Toepasbaarheid en verder onderzoek 18

7. Addenda 20
- 7.1. Overzicht VRS 20
- 7.2. Tabel: top 20 descriptoren per VRS model en menselijk descriptoren 22
- 7.3. Tabel: verhouding correcte descriptoren op totaal (ruis) 23
- 7.5. Voorbeelden geanalyseerde beelden 24
- 7.6 Voorbeelden van use cases 25

8. Bronnen 27

1. Management samenvatting

VR4CH was een Innovatief Partnerproject dat in 2018-209 werd uitgevoerd door het MoMu, datable en PACKED/VIAA met een subsidie van de Vlaamse Gemeenschap. Het doel van het project was te onderzoeken in welke mate online beeldherkenningservices (VRS) kunnen worden ingezet bij het doorzoekbaar maken van beeldcollecties uit de erfgoedsector.

Dit verkennend onderzoek gaf een inzicht in het aanbod en de werking van VRS. Uit analyse van de resultaten bleek dat de door een VRS gegenereerde descriptoren van wisselende kwaliteit zijn en veel ruis bevatten. Anderzijds bleken de descriptoren vaak voldoende specifiek om verkennende zoekopdrachten mogelijk te maken, en dit tegen een relatief lage kostprijs en hoge snelheid. Naar aard bleken de descriptoren vaak andere invalshoeken te gebruiken om een beeld te beschrijven. Mede op basis van deze laatste conclusie werden een aantal meer specifieke use case beschreven die het voorwerp kunnen zijn van verder onderzoek.

In dit rapport beschrijven we de typologie en werking van de VRS. In volgende hoofdstukken worden de gehanteerde methodologie beschreven voor het testen van de VRS, gevolgd door een beoordeling en conclusie.

2. Inleiding

De registratie van museale objecten, archieven en andere erfgoedcollecties is doorgaans beperkt tot een aantal formele kenmerken en administratieve gegevens die vereist zijn voor het beheren van het object (de 'basisregistratie'). Manuele registratie is immers een zeer arbeidsintensief proces, waardoor bijkomende registratie van inhoudelijke kenmerken bij iconografische collecties (schilderijen, prenten, foto's) en objecten (bv. kleur van een jurk, een specifieke print, types van kledingstukken,..) vaak slechts beperkt aanwezig zijn.

Sedert de ontwikkeling van de eerste beelddatabanken werd al onderzoek gevoerd naar behoeften van gebruikers. Wanneer beeldmateriaal wordt gezocht over reeds gekende objecten ('known item searches'), volstaan basisgegevens zoals titel of objectnaam doorgaans. Deze 'known item searches' maken echter slechts een minderheid uit van de werkelijke zoekacties (Armitage, 1997). Bij de meeste zoekopdrachten wordt gezocht op inhoudelijke kenmerken naar een vooraf niet gekend beeld (Bakewell, 1988). Hierbij worden verschillende niveaus onderscheiden, die variëren tussen identificeerbare objecten ('de Eiffeltoren'), generieke objecten ('een stoel'), iconologische thema's ('het laatste avondmaal') en abstracte begrippen ('geluk', jeugd') (Shatford, 1986).

Recente evoluties in informatietechnologie bieden nieuwe oplossingen voor deze blinde vlek in informatievoorziening. Sedert een tiental jaren wordt onder meer ingezet op crowdsourcing als middel om omvangrijke beeldcollecties te ontsluiten (Oomen & Aroyo, 2011).

Ontwikkelingen in artificiële intelligentie (AI) blijken recent echter veel krachtiger oplossingen te bieden aan de behoefte om voorwerpen en personen geautomatiseerd te herkennen (Grauman & Leibe, 2010). De technologie is intussen beschikbaar als Software as a Service op meerdere platformen (o.a. IBM, Google, Microsoft, Amazon). Onder meer bij het ontsluiten van commerciële beeldbanken raakt dit soort services snel ingeburgerd.

Toch blijft het gebruik van deze visual recognition services (VRS) binnen de erfgoedgemeenschap nog zeer beperkt (Dee, Hughes, Roderick & Brown, 2016). Er werden een aantal toepassingen ontwikkeld die het herkennen van beelden voor de eindgebruiker ondersteunen (Wein, 2014). Ook werd er onderzoek uitgevoerd naar het herkennen van objecten op schilderijen op basis van getrainde AI-systemen (Crowley & Zisserman 2014). In vele gevallen zijn deze toepassingen echter niet haalbaar of bruikbaar voor erfgoedorganisaties, die niet beschikken over de middelen en expertise om dergelijke services in het dagelijkse beheer van de collectie in te schakelen. In dit onderzoek willen we daarom de vraag stellen of cloud-based VRS, zonder verdere configuratie of training, kunnen helpen bij het ontsluiten van erfgoedcollecties.

Met dit project wil MoMu, met ondersteuning van Datable en PACKED vzw, visual recognition services (VRS)[1] inzetten om na te gaan in hoeverre beeldcollecties van musea efficiënter en beter kunnen worden beschreven en ontsloten, zonder grote investeringen. Om dit te onderzoeken werd een pilot uitgevoerd met beeldmateriaal uit de studiecollectie en de museale collectie van het MoMu.[2]

Uitgaande van de vaststelling dat VRS off-the-shelf beschikbaar is, wil het onderzoeksproject antwoord zoeken op volgende vragen:

[1] In dit document wordt de afkorting VRS zowel voor de enkelvoud- als meervoudsvorm gebruikt.
[2] http://studiecollectie.momu.be/

(1) Welke services zijn er beschikbaar, en welke functionaliteiten bieden ze tegen welke kostprijs?
(2) Wat is de aard en bruikbaarheid van de descriptoren die door een VRS worden geleverd?
(3) Is de kwaliteit van de descriptoren voldoende om - zonder training - relevante informatie toe te voegen?
(4) Waarin verschillen resultaten (kwantitatief, kwalitatief, qua kostprijs) van menselijke registratie?
(5) Hoe kunnen deze praktisch ingeschakeld worden in de bedrijfsprocessen van een erfgoedbeherende organisatie?

Het MoMu, Datable en PACKED vzw ontvingen van de Vlaamse overheid een subsidie binnen het kader van de Innovatieve Partnerprojecten (oproep 2018) voor het project VR4CH.[3] Het project werd uitgevoerd in 2018 - 2019.

[3] https://cjsm.be/cultuur/themas/cultuur-en-economie/innovatieve-partnerprojecten

3. VRS services

3.1. Descriptoren

Gebruik van artificiële intelligentie (AI) voor geautomatiseerde beeldverwerking (computer vision) kent meerdere, uiteenlopende toepassingen, zoals het het herkennen van tekst (OCR), het reconstrueren van beelden of scènes, het detecteren van bewegingen etc. Eén van de meest gebruikte toepassingen is het geven van een beschrijving van een afbeelding (visual recognition of VR). VR kan verschillende aspecten van een beeld beschrijven:

- object detection: beschrijven van op het beeld aanwezige objecten (of personen), eventueel aangevuld met hun positie, leeftijd, emotie, geslacht;
- algemene beschrijving van het soort beeld ('landschap');
- aard van het beeld ('drawing', 'photo');
- 'aboutness': emoties, sfeer, thema van een beeld ('fashion', 'glamour');
- entity recognition: identificatie van bekende personen of plaatsen (landmarks);
- kleur (globaal, van het hoofdonderwerp, accentkleur);
- gelijkaardige beelden classificeren of clusteren;
- NSFW-score: content waaraan om een of andere reden aanstoot kan worden genomen (erotiek, seksisme, racisme, geweld).[4]

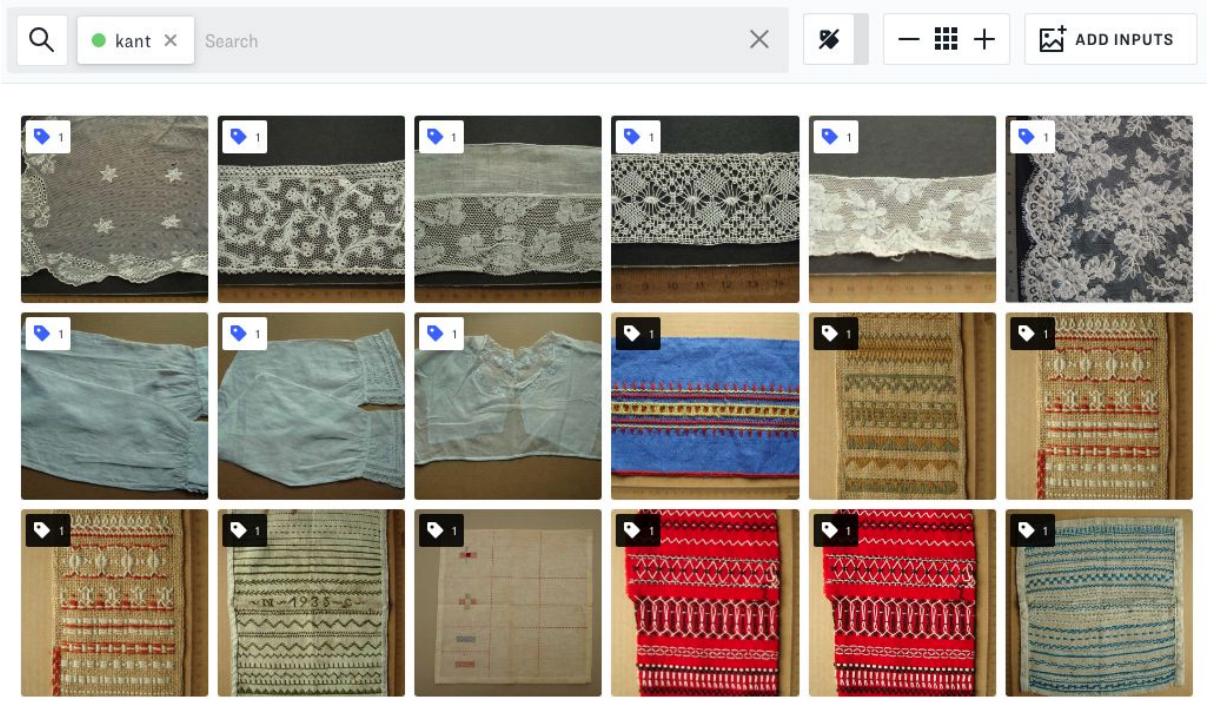
De descriptoren kunnen verschillende vormen aannemen, zoals trefwoorden (tags), een korte zin (tagline) of een code (bv. een kleurcode). Doorgaans zijn de textuele descriptoren in het Engels, al bieden een aantal VRS ook andere talen aan.[5]
Vaak wordt bij een descriptoren een score vermeld die de waarschijnlijkheid aangeeft waarmee de descriptor werd toegekend.

Op basis van de descriptoren kunnen andere functies worden aangeboden, zoals het doorzoekbaar maken of classificeren van de verwerkte beeldcollecties.

3.2. Modellen en training

Een VRS dat als SaaS wordt aangeboden is meestal reeds getraind. Trainen gebeurt op basis van verschillende sets van beelden en bijbehorende descriptoren. Door een andere trainingset te gebruiken, leert de AI-applicatie een verschillend model aan. Elk model heeft eigen functionaliteiten (bv. het al dan niet herkennen van kleuren, creëren van taglines, herkennen van personen) binnen een bepaald domein (bv. mode, celebrities).
De meeste VRS bieden de mogelijkheid een eigen model te creëren, d.w.z. de VRS te trainen met eigen beeldmateriaal en metadata.

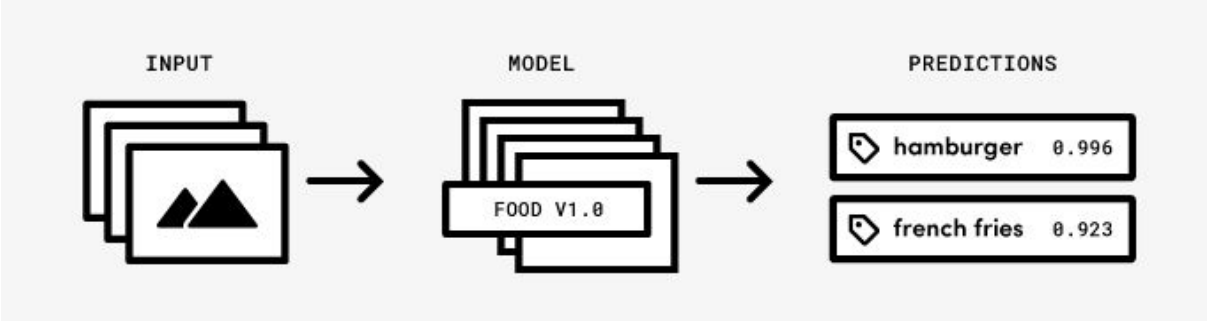
[4] NSFW: "not safe for work"
[5] Om vergelijking tussen de verschillende aanbieders mogelijk te maken werd enkel de Engelstalige versie getest.



Resultaat van upload op een model dat werd getraind om onderscheid te maken tussen kant en merklappen © Clarifai

3.3. Interactie

Interactie met een VR applicatie kan on-premise (met een eigen applicatie, soms verpakt in een hardware component) of via een SAAS toepassing - in dat geval spreken we hier van een Visual Recognition Service of VRS.



Basismodel interactie met VRS © Clarifai

Een aantal VRS bieden een grafische userinterface (GUI) aan, die kan gebruikt worden om de service te verkennen. Voor intensiever gebruik wordt steeds een API voorzien, zodat de gebruiker toepassingen kan schrijven die met de VRS kunnen interageren. Om dit te ondersteunen voorzien de meeste VRS softwarebibliotheken in verschillende programmeertalen, die vrij gebruikt kunnen worden. Zowel de beeldherkenning zelf, als de training die eraan voorafgaat, kan via de API gebeuren.

Het gebruik van de API is doorgaans goed gedocumenteerd. Voor het gebruik van een VRS via de API is uiteraard een account nodig - al laten enkele services (zoals Clarifai) wel toe om via de grafische interface anoniem beelden op te laden.

API-requests verlopen via HTTP (voor de REST API calls). Payloads en responses worden in JSON formaat verstuurd.

Het beeldmateriaal kan op twee manieren aan de VRS bezorgd worden (een aantal VRS ondersteunen beide):

(1) versturen van de URL als onderdeel van de call naar de API;
(2) coderen van het beeld (meestal volgens het base64 algoritme), dat in deze vorm in call naar de API wordt gestuurd.

3.4. Vereisten beeldmateriaal

Een eerste vereiste is uiteraard dat het beeldmateriaal gedigitaliseerd is.[6] Om het coderen en verzenden van beelden als payload bij elke request te vermijden, is het gewenst dat het beeldmateriaal via HTTP bereikbaar is - bij voorkeur via een persistente URL

De beeldkwaliteit zelf heeft slechts een beperkte invloed: hoewel een minimale resolutie van de afbeelding vereist is (Wang e.a. 2016), geven de meeste VRS aan dat een resolutie van ca. 500 pixels volstaat. Hogere resoluties leiden niet per se tot betere descriptoren, maar kunnen daarentegen de performantie van de service vertragen. Voor VRS toepassingen die werken op basis van bepaalde details uit een beeld (bv. het herkennen van emoties) kunnen hogere resoluties uiteraard wel nodig zijn. Beelden die enkel in zeer hoge resolutie beschikbaar zijn, kunnen daarom beter vooraf geschaald worden zodat een kleinere bestandsomvang wordt verkregen.

Alle services ondersteunen het JPEG bestandsformaat. Ook PNG en TIFF worden meestal ondersteund (zie overzichtstabel in het Addendum). Het kan dus nodig zijn het bestandsformaat om te zetten naar een geaccepteerd formaat.

3.5. Kosten

Kosten voor het gebruik van SAAS aanbieders van VRS zijn relatief laag. Doorgaans worden de kosten per call aangerekend: voor elke actie wordt dus een (minimaal) bedrag in rekening gebracht. Elke aanbieder geeft gebruikers een beperkt aantal calls kosteloos. Vaak is dit voldoende om voor een beperkte set training en analyse uit te voeren.

3.6. Selectie VRS

Voor deze case study bestudeerden we een uitgebreid aantal VRS. De belangrijkste spelers zijn Amazon Rekognition, Microsoft Computer Vision, Google Cloud Vision, IBM Watson Visual Recognition en Clarifai.

Google Cloud Vision kan voor zijn image recognition gebruik maken aan de massale hoeveelheid data die via het web zijn geïndexeerd. Een service waarmee Google Vision zich daardoor onderscheid van de andere diensten is de mogelijkheid gelijkaardige beelden op het web terug te vinden. Google Vision zal op basis van gelijkaardige of identieke beelden ook descriptoren voorstellen.[7]

Microsoft biedt image recognition aan binnen het Azure platform. Alle courante services worden ondersteund (object recognition, kleur, gezichtsherkenning, NSFW). Ook trainen is mogelijk, maar de getrainde set moet (i.t.t. bijvoorbeeld Clarifai) naar een eigen instance in Azure worden

[6] De meeste VRS zijn ook geschikt om video te analyseren. Omdat voor deze case study enkel 2D beeldmateriaal werd gebruikt, gaan we niet verder in op de mogelijkheden van video.
[7] https://cloud.google.com/vision/

geëxporteerd. Bijzonder aan Azure is dat ook steeds een 'natural language description' of tagline wordt gegeven ('A cat on a chair').[8]

Clarifai kan niet bogen op een ruim ecosysteem zoals de Google, Amazon of Microsoft, maar legt de focus helemaal op beeld- en videoherkenning.[9] Het positieve gevolg is dat het aanbod zeer overzichtelijk is en - in tegenstelling tot bijvoorbeeld Amazon - niet verdrinkt in een kluwen van diensten en bijbehorende documentatie. De gebruiksvriendelijkheid van Clarifai wordt nog verhoogd door de grafische interface, waarmee de service snel verkend kan worden.

Amazon biedt beeldherkenning aan als onderdeel van het eigen Amazon ecosysteem.[10] Een sterk punt van Amazon Rekognition is de mogelijkheid om personen te identificeren. Werken met Rekognition buiten het ecosysteem van Amazon is minder eenvoudig en kan enkel via een lokaal filesystem, wat de toepasbaarheid en flexibiliteit beperkt. Als enige van de hier besproken VRS laat Amazon geen custom models toe: trainen op basis van een eigen set beeldmateriaal is dus niet mogelijk.

IBM Watson Visual Recognition[11] biedt geen herkenning van video. Ook het aantal modellen dat Watson aanbiedt is relatief beperkt. Naast deze en andere grote spelers (Alibaba[12], EinsteinAI/Salesforce[13]) zijn er talrijke startups die een plaats op de markt proberen te veroveren (bv. Ciliar[14], Imagga[15], Ximilar[16]).

Een aantal services zijn varianten van zogenaamde mechanical turks: Cloudsight[17] (een combinatie van AI en human tagging) en Dataturks[18]. Deze laatste levert human tagging om training sets te creëren en is dus zelf geen AI toepassing. De resultaten zijn wisselend van kwaliteit en onvoorspelbaar. Zowel de responstijden als de kost van deze services liggen doorgaans een stuk hoger.

Enkele services zetten in op één specifiek model, zoals Kairos, dat uitsluitend voor gezichtsherkenning geschikt is.

Ten slotte vermelden we een aantal open source toepassingen zoals OpenCV[19] en TensorFlow (open source software libraries), ImageAI (python library gebaseerd op TensorFlow) en YOLO[20] (een open source applicatie. Deze applicaties zijn zeker interessant, maar worden in deze vorm niet als een SAAS aangeboden. Ook de grote aanbieders maken trouwens vaak gebruik van open source software (bv. TensorFlow bij Google).

[8] https://docs.microsoft.com/en-gb/azure/cognitive-services/Computer-vision/
[9] https://clarifai.com/
[10] https://console.aws.amazon.com/rekognition/home?region=us-east-1#/
[11] https://www.ibm.com/watson/services/visual-recognition/
[12] https://www.alibabacloud.com/product/imagesearch
[13] https://einstein.ai/
[14] https://ciliar.co
[15] https://imagga.com
[16] https://www.ximilar.com
[17] https://cloudsight.ai
[18] https://dataturks.com
[19] https://opencv.org
[20] https://pjreddie.com/darknet/yolo

Uit bovenstaande lijst selecteerden we de services van Google, Microsoft en Clarifai. Elk van deze dienstverleners bieden een voldoende uitgebreid platform, relatief eenvoudige API en goede documentatie. Belangrijk is ook dat bij elk van deze leveranciers een beperkt aantal calls gratis wordt aangeboden, wat de drempel verlaagt voor organisaties die beperkte middelen hebben of eerst met de technologie willen experimenteren.

4. Methodologie en uitvoering

4.1. Selectie materiaal

Voor deze case study werd een aantal sets met beeldmateriaal uit de collectie van het Modemuseum Antwerpen gebruikt, afkomstig uit de studiecollectie en de museale collectie. Het beeldmateriaal bestond uit 327 beelden, waarvan uiteindelijk 164 werden gebruikt voor de evaluatie van de VRS.

De testset voor deze use case bevatte vier types beeldmateriaal:

(1) objecten: deze werden doorgaans geïsoleerd en in hun geheel gefotografeerd tegen een egale (witte of donkere) achtergrond. In een aantal gevallen waren de objecten op paspop gefotografeerd.
(2) details en close ups van objecten
(3) catwalkbeelden
(4) foto's van events (bv. catwalk, backstage bij show, vernissage van tentoonstelling)
(5) scenografie (fotos van tentoonstellingen)

De variatie aan types kan representatief beschouwd worden voor museale beeldcollecties, waar zowel objecten als documentaire foto's aangetroffen worden. Het onderwerp (mode en textiel) is uiteraard zeer specifiek. Ook wat betreft de kwaliteit van de beelden werd gestreefd naar een set die representatief is voor een museale beeldcollectie, d.w.z. van wisselende kwaliteit (scherpte, contrast) en niet steeds even geschikt om door beeldherkenningssoftware te worden geanalyseerd.

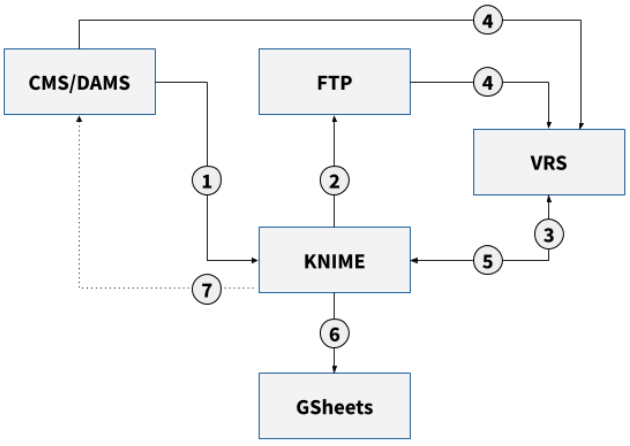
Een deel van deze beelden was voorzien van descriptoren (title, materiaal, objectnaam).

4.2. Set-up van de workflows

Beelden van de studiecollectie waren online gepubliceerd en bereikbaar via de API van het CMS Omeka S.[21] Het CMS beschikt over een API waarmee metadata en media kunnen worden doorzocht en gedownload. Beelden van de museale collectie en events waren beschikbaar via het DAMS ResourceSpace. Ook dit DAMS beschikt over een API, al bleek het niet eenvoudig om dit in een geautomatiseerde workflow aan te bieden aan een VRS.
Zowel in Omeka S als ResourceSpace werd door de collectiebeheerder een set gedefinieerd die overeenstemde met het geselecteerde beeldmateriaal.

De gegevens werden verwerkt met behulp van Knime.[22] Dit is een open source tool voor data-integratie, -analyse en -verwerking. De tool laat toe om snel en flexibel data te verwerken en is daarom zeer geschikt voor het maken van proefopstellingen of een proof of concept. Met Knime werden interacties tussen verschillende componenten (DAMS, CMS, VRS, FTP, visualisatie) uitgewerkt.
Omdat interactie tussen ResourceSpace en de VRS niet lukte zonder het beveiligingsniveau van ResourceSpace aan te passen, werden beelden uit dit DAMS tijdelijk gepubliceerd op een FTP server.
Voor een visuele representatie van de resultaten en de evaluatie van de gegevens werd gebruik gemaakt van Google Sheets.

[21] Omeka S is een CMS waarmee erfgoedorganisaties hun collectie online kunnen publiceren (bv. webtentoonstellingen te maken) https://omeka.org/s/
[22] http://knime.org



Basisproces beeldherkenning

1. Haal beelden op uit CMS/DAMS
2. Plaats beelden op FTP
3. Stuur URLs naar VRS
4. Haal beelden uit CMS of FTP
5. Haal resultaten op
6. Verwerk resultaten en publiceer op Google Sheet
7. importeer resultaten in CMS/DAMS

In een eerste iteratie werden verschillende modellen van de VRS getest. Op basis van de eerste resultaten werd een verdere selectie gemaakt van de modellen en de descriptortypes die binnen een model werden aangeboden (zie tabel). In enkele gevallen werd een threshold of limiet ingesteld om minder relevante resultaten te filteren. Bij thresholds werd een minimale score ingesteld van de descriptor (d.w.z. de mate van waarschijnlijkheid waarmee de descriptor aan het beeld kan worden toegekend). Wanneer dit niet mogelijk was of wanneer het aantal descriptoren te hoog lag (met oog op menselijke controle), werd een maximaal aantal descriptoren bepaald.

VRS - model - descriptortype	**threshold / limiet**	**voorbeeld**
Clarifai - general	limiet 20	elegant
Clarifai - color	limiet 3	#fdfdfd
Clarifai - accessories	limiet 4	Blouse
Google Vision - webDetection - bestGuesLabel	nvt	day dress
Google Vision - webDetection - webEntities	-	Cocktail dress
Google Vision - description	-	Clothing
Microsoft Azure - dominant foreground color	-	White
Microsoft Azure - description	-	a person wearing a black dress
Microsoft Azure - text	-	skirt
Microsoft Azure - tags	score > 0.25	shoe

In een tweede iteratie werd de volledige, geparametriseerde workflow uitgevoerd voor een uitgebreide set beelden. Ter illustratie geven we in bijlage de meest frequent voorkomende termen voor elk van de gebruikte services (uitgezonderd Clarifai-color) en de reeds aanwezige metadata.

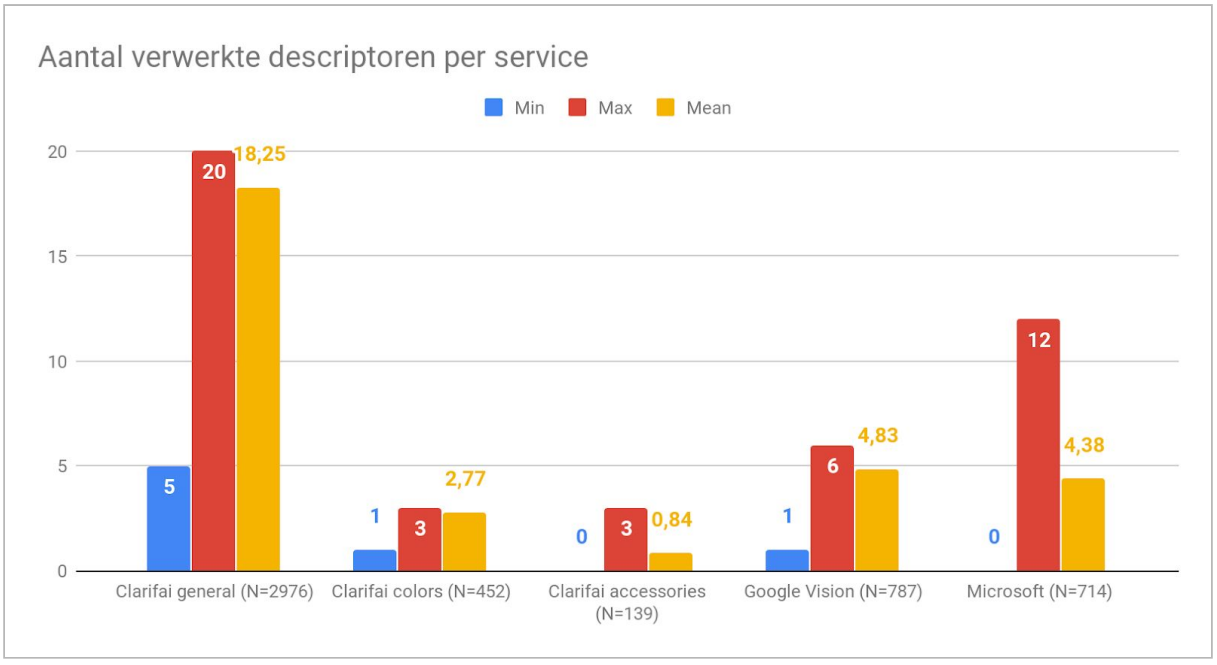
5. Beoordeling

Beoordeling van VRS gebeurt doorgaans aan de hand van 'gestandaardiseerde' sets van afbeeldingen die van descriptoren zijn voorzien, zoals de NORB en de MINIST datasets (Lin, 2018). Andere analyses richten zich op zeer specifieke problemen, zoals het bekende Chihuahua/Muffin fenomeen (Yuo, 2017) of zijn beperkt in omvang en geven geen kwantificering van de resultaten (Filestack, 2019).
Voor dit onderzoek werkten we een eigen vereenvoudigde methodologie uit, die toelaat de kwantiteit te meten en de kwaliteit van de descriptoren te kwantificeren, waarbij we zoveel mogelijk rekening trachtten te houden met de context waarin beeldmateriaal en metadata in de erfgoedsector wordt gebruikt.

Een gedetailleerde analyse met vergelijking tussen de door de VRS aangeleverde descriptoren en menselijke beschrijvingen was moeilijk haalbaar: de menselijke descriptoren waren - voor zover beschikbaar - steeds in het Nederlands. Zelfs na vertaling zou het nog moeilijk zijn om te bepalen of twee descriptoren dezelfde concept vertegenwoordigen. We merken bovendien hierbij nogmaals op dat VRS ook andere types descriptoren aanlevert (sfeer, kleur, …) dan de meer klassieke objectregistratie, waarbij vooral objectnamen en materialen worden gegeven. Een vergelijking op dit niveau zou dus weinig zinvol zijn.

5.1. Kwantitatieve resultaten

Onderstaande tabel geeft een overzicht van het aantal descriptoren per VRS. Let wel dat dit minder is dan het totaal aantal descriptoren dat in de eerste iteratie werd ontvangen: het betreft hier enkel de descriptorentypes die als relevant werden weerhouden.



Het aantal descriptoren verschilt dus aanzienlijk - zowel tussen services onderling als tussen verschillende beelden. Ter informatie: voor de 164 geanalyseerde beelden waren 133 door een mens gecreëerde descriptoren beschikbaar.

5.2. Type descriptoren

Naar aard leverden de services verschillende types descriptoren aan:

- Clarifai general gaf steeds een groot aantal descriptoren met betrekking tot allerlei aspecten van het beeld, zowel herkende objecten ('dress', 'footwear') en materialen ('cotton') als algemene onderwerpen ('fashion', 'exhibition') en bijvoeglijke naamwoorden ('old', 'beautiful'). Kleuren werden niet gegeven in dit model.
- Clarifai accessories gaf telkens slechts enkele descriptoren met betrekking tot objecten die op de foto werden herkend.
- Google Vision gaf net als Clarifai general een breed spectrum aan descriptoren ('fashion', 'bag', 'illustration', …), maar ook kleuren (blue) en named entities ('Raf Simons'). Deze laatste zijn steeds afkomstig van de webDetection-types - d.w.z. descriptoren die gebaseerd zijn op kennis die Google eerder op het web heeft verzameld.
- Microsoft Computer Vision gaf eveneens descriptoren van het type onderwerp, objectnaam, kleur en adjectieven ('standing') zoals hoger vermeld aangevuld met taglines. Kleuren werden zowel gecodeerd als in de vorm van tekst gegeven. In twee gevallen bevatte de tagline ook een named entitiy (waarvan één ook correct was).

5.3. Kwalitatieve beoordeling

Om een kwalitatieve beoordeling mogelijk te maken werden de ruim 5000 descriptoren beoordeeld door een domeinexpert. Elke descriptor werd vergeleken met het geanalyseerde beeld en beoordeeld als waar/onwaar. De waarde 'waar' werd toegekend voor elke descriptor die als correct kon worden gevalideerd - onafgezien van de intrinsieke relevantie van de descriptor. Het voordeel van deze neutrale, binaire aanpak is dat deze validatie relatief snel kon worden uitgevoerd. Meer nuances in de beoordeling zou het evalueren complexer maken.

Bij de evaluatie door de domeinexpert werden een aantal opmerkingen gegeven die een eerste indruk geven van de kwaliteit van de toegekende descriptoren.

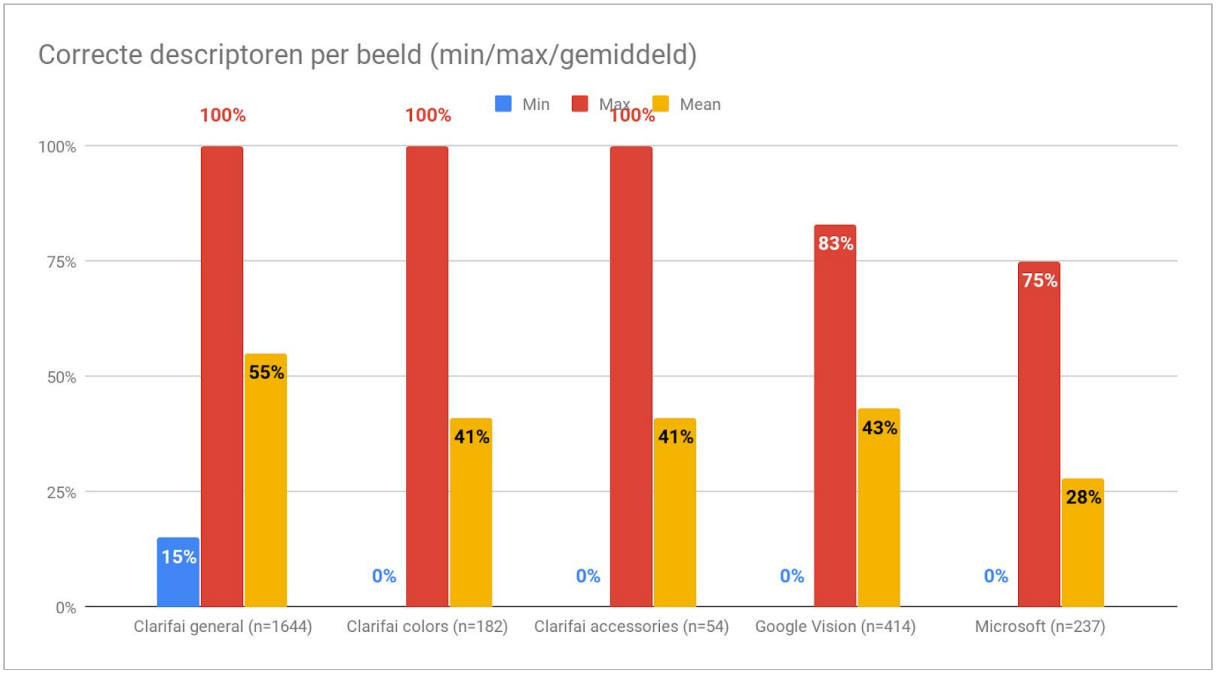
Clarifai color maakt geen onderscheid tussen voorgrond- en achtergrondkleur. Een kleur wordt als een 'gemiddelde' kleur weergegeven, bijvoorbeeld een fijn zwart-wit patroon wordt grijs. Vaak wordt de hoofdkleur van een object niet herkend.
Clarifai accessories werkt niet goed bij het herkennen van objecten in 'drukke scenes' (vernissages e.d.). Veel objecten zijn niet getagd.
Clarifai-general geeft erg veel generieke, vaak weinigzeggende descriptoren. Veel descriptoren met gelijkaardige betekenis (bv. retro - vintage - antique - old). Soms wordt het hoofdonderwerp gemist, hoewel er wel andere relevante descriptoren gegeven worden.

Microsoft geeft een beperkt aantal descriptoren in vergelijking met Clarifai, maar ze zijn soms wel informatiever, specifieker of relevanter in vergelijking met Clarifai (zie echter ook verder voor een kwantificering van de kwaliteitsdimensie.) Er zijn relatief veel beelden niet van descriptoren voorzien. Description (tagline) is soms interessant, maar meestal slaat deze de bal mis (slechts 32 van de 148 taglines werden als correct beoordeeld).

Bij Google Vision tenslotte valt vaak op dat de beelden door Google al gekend zijn. Soms worden zelfs Nederlandstalige descriptoren gegeven ('jarretellengordel'). Ook foute informatie wordt overgenomen. Google geeft voor één beeld vaak identieke descriptoren in verschillende descriptortypes.

5.4. Ruis

Op basis van de evaluatie door de domeinexpert kan ruis gemeten worden door de verhouding correcte termen op het totaal te meten (zie grafiek). Onderstaande tabel geeft per VRS telkens het totaal aantal descriptoren, het aantal correcte descriptoren en de verhouding tussen beide waarden.



De relevantie score (verhouding tussen correcte aantal en totaal aantal descriptoren) ligt meestal onder de helft van het totaal. De beste score wordt behaald door Clarifai-general (55%), tegenover 28% bij Microsoft. Omgekeerd geformuleerd betekent dit dat er een hoge mate van ruis (tussen 42 en 72%).
Een tweede vaststelling is dat enkel Clarifai general steeds minstens één correcte descriptor geeft.

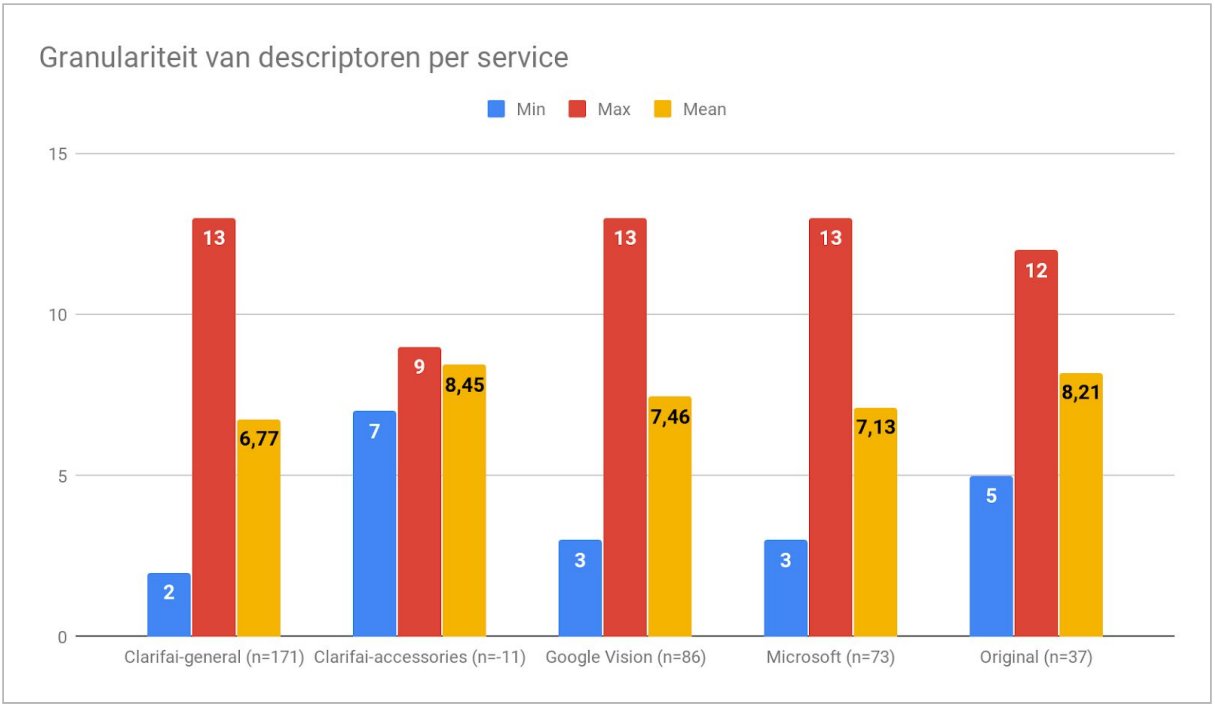
5.5. Granulariteit

Granulariteit is een maatstaf voor de specificiteit van de gegeven descriptoren. Hoe hoger de granulariteit, hoe nauwkeuriger en dus 'informatiever' de descriptor.

Om granulariteit te bepalen werden de als correct bestempelde descrip- toren gereconcilieerd met Art & Architecture Thesaurus (AAT).[23] De AAT is een thesaurus van erfgoed- concepten. De structuur van de thesaurus wordt opgebouwd door meer specifieke

[23] http://www.getty.edu/research/tools/vocabularies/aat/index.html

concepten als narrower term te koppelen aan meer algemene concepten (broader term). Door de positie van elk gereconcilieerd concept ten opzichte van de top van de thesaurus te meten, kan de mate van granulariteit worden bepaald (zie grafiek en illustratie).



Op basis van deze meting kan worden besloten dat Clarifai-accessories gemiddeld de meest specifieke resultaten geeft (positie 8.45) en daar ook het meest consistent in is (standaarddeviatie 0.68). Daarbij moet rekening gehouden worden met het beperkt aantal descriptoren en het beperkt aantal (11) matches van Clarifai-accessories met een AAT concept.. Zoals bij de kwalitatieve beoordeling door de domeinexpert al was opgemerkt, geeft Clarifai-general de meest generieke descriptoren. Het hoge aandeel correcte descriptoren van Clarifai-general blijkt dus relatief: ze zijn wel correct, maar ook minder informatief.

Een tweede vaststelling is dat de specificiteit van de VRS-descriptoren niet veel afwijkt van die van de reeds aanwezige ‘menselijke’ descriptoren. Clarifai-accessories scoort hier zelfs iets beter met descriptoren die gemiddeld zowat even specifiek zijn als degene die door een registrator werden toegekend.

5.6. Snelheid

De performantie (snelheid) van de services werd in deze case study niet gemeten. Ander onderzoek toont aan dat responstijden sterk kunnen variëren (van 0.4 seconden tot een minuut), met gemiddelden tussen 1 en enkele seconden (Filestack, 2018). In de proefopstelling werd een timeout van 2 seconden gebruikt, wat slechts in enkele gevallen werd overschreden. Er werden geen opvallende verschillen vastgesteld tussen de verschillende services. Onderzoek (Yao, 2017) wijst er op dat zeer oppervlakkige beschrijving door menselijke ‘mechanical turks’ 10 to 20 maal trager gaat. Zonder hier een quotering op te kunnen geven, is het duidelijk dat een VRS het qua snelheid steeds wint van een menselijke registrator.

6. Conclusies

6.1. Conclusies cases study

Het analyseren van een collectie beelden rond een redelijk gespecialiseerd onderwerp levert geen descriptoren van hoge kwaliteit (veel ruis, inconsequenties, missers). Een VRS legt het dus duidelijk af tegen een domeinexpert. Een VRS mist immers de gespecialiseerde en technische en historische kennis en context om een object gedetailleerd en correct te beschrijven op basis van één enkel beeld. Voor zover enige mate van ruis acceptabel is, levert een VRS 'out-of-the-box' toch vaak bruikbare resultaten met een onverwacht hoge mate van granulariteit. Afhankelijk van de toepassing volstaan de resultaten dus mogelijk voor een indicatieve, globale beschrijving van erfgoedcollecties.

Het sterkste punt van een VRS is de performantie: een VRS is duidelijk sneller én goedkoper dan een menselijke registrator: binnen enkele seconden worden beelden van descriptoren voorzien, waarvoor een registrator veel meer tijd nodig heeft.
Een bijkomende toegevoegde waarde van een VRS is dat het ook onderwerpen en categorieën geeft die buiten traditionele scope van registratie vallen (kleur, stijl) - hoewel ook hier meer ruis te verwachten is.

6.2. Inzetbaarheid van VRS voor erfgoedcollecties

Om de vraag te kunnen beantwoorden of VRS bruikbaar zijn voor collectiebeheer en -ontsluiting, moeten meerdere aspecten in beeld worden gebracht, zoals de aard, omvang en kwaliteit van de beeldcollectie, nabewerking, eventuele training van de VRS en uiteraard de gewenste kwaliteit en het precieze doel waarvoor de VRS wordt ingezet (de use cases).

Voorafgaand aan het uitvoeren van beeldherkenning is het noodzakelijk de workflow op de juiste manier te parametriseren om zo een optimaal resultaat te verkrijgen. Dit kan enerzijds door selectie van de VRS modellen die het best aansluiten bij de inhoud van het beeldmateriaal en de verwachte uitkomst, en anderzijds door instellen van de gewenste thresholds om ruis te minimaliseren, op basis van de probability score van de descriptoren.

Resultaten van de ruwe output van een VRS kunnen vervolgens op verschillende manieren verwerkt worden om de kwaliteit te optimaliseren, in functie van de gestelde use case. Enkele voorbeelden:

(1) handmatig beoordelen van descriptoren;
(2) geautomatiseerd filteren van de descriptoren om ongewenste of irrelevante descriptoren te verwijderen d.m.v. blacklisting of whitelisting;
(3) clusteren van de descriptoren om beelden te classificeren op inhoudelijke of visuele kenmerken;
(4) opzoeken van de descriptor in een trefwoordenlijst of authority om verrijking (bv. vertaling) mogelijk te maken. Een voorbeeld is de reconciliëring met de AAT (zie hoger bij beoordeling van de granulariteit).

6.3. Use cases

De vraag stelt zich bij welke praktische toepassingen VRS kan worden ingezet in een erfgoedcontext. Deze use cases worden enerzijds bepaald door de aard en omvang van de collectie, anderzijds door het gebruik dat men wil maken van de descriptoren die de VRS genereert.

Naar type collectie kan onderscheid worden gemaakt tussen:

(1) foto's van objecten uit de collectie;
(2) digitale reproducties van tweedimensionale objecten (schilderijen, foto's, gravures e.d.);
(3) archief en documentatie, incl. het eigen beeldarchief van de organisatie;
(4) bibliotheekcollecties, zoals gedigitaliseerde tijdschriften en boeken.

Naar gebruik kan onderscheid gemaakt worden tussen:
(1) beschrijven van een collectie fysieke objecten op basis van beeldmateriaal;
(2) pre-iconografisch beschrijven van beeldmateriaal;
(3) analyseren en op collectieniveau beschrijven van beeldcollecties;
(4) clusteren van beeldmateriaal in subcollecties;
(5) herkennen van personen, plaatsen (entity recognition)
(6) iconologisch beschrijven

In elk van deze gevallen moet worden bepaald welke kwaliteit men van de VRS verwacht en welke aanvullende (door mensen uitgevoerde) acties men wil voorzien. Wanneer de VRS als vervanging van menselijke arbeid wordt gezien, zal er een hogere foutenmarge moeten aanvaard worden. Een VRS kan ook ingezet worden als ondersteuning bij of aanvulling op menselijke analyse. In dat geval kunnen procedures ontwikkeld worden waarbij de resultaten van de VRS worden gevalideerd en waar nodig bijgestuurd. Hieronder geven we enkele use cases waarvoor VRS een ondersteunende rol kan spelen.

6.4. Toepasbaarheid en verder onderzoek

Om het gebruik van VRS voor erfgoedcollecties verder ingang te doen vinden, moeten een aantal aspecten verder worden onderzocht:
(1) mogelijkheden en impact van het trainen van VRS;
(2) alternatieven voor SAAS VRS, zoals het gebruik van open source toepassingen, het ontwikkelen van eigen toepassingen op basis van open source bibliotheken, on premise toepassingen, edge computing etc.;
(3) verwerking van output van VRS zodat deze aansluit bij de interne werkprocessen van erfgoedbeheerders:
    (a) workflows voor validatie van de output;
    (b) inzet van crowdsourcing voor validatie
    (c) integratie van data in de collectiebeheerssystemen;
    (d) aanpassingen aan het datamodel van objectregistratie om het statuut van door AI gegenereerde data te documenteren;
    (e) reconciliëren van de output van VRS met gestandaardiseerde woordenlijsten en thesauri.
(4) Gebruik van output van VRS voor andere vormen van ontsluiting, d.w.z. op basis van kleur, gelijkaardige beelden.

In dit onderzoek lag de focus op de output van VRS in de vorm van - doorgaans textuele - metadata, als een aanvulling op of alternatief voor menselijke registratie. Ook het gebruik van andere vormen van output (gelijkende beelden, cropping hints, NSFW-score e.d.) kan verder worden onderzocht en nieuwe opportuniteiten bieden voor het gebruik van erfgoeddata.

Het ModeMuseum zal op basis van de ervaring die in dit project werd opgedaan verder onderzoek doen naar de praktische inzetbaarheid van VRS om zijn collectie, archief en documentatie verder te ontsluiten. Daartoe worden verschillende use cases opgezet, waarvan een aantal zullen worden uitgevoerd in het kader van een subsidieproject voor cultureel-erfgoedwerking op landelijk en

internationaal niveau, met steun van de Vlaamse overheid en gecoördineerd door het Antwerpse Fotomuseum (FOMU).

Use cases kunnen onder meer betrekking hebben op:

- ontsluiten van beelden van museumobjecten door geautomatiseerd toekennen van descriptoren die gematched worden met de Europeana Fashion Thesaurus.[24] Hiervoor kan een model worden getraind op basis van beeldmateriaal dat reeds met deze Fashion thesaurus is beschreven.
- identificeren en beschrijven van illustraties in modetijdschriften waarin zowel tekst als beeld voorkomen;
- automatisch selecteren van beelden op basis van esthetische en technische beeldkwaliteit (vb. compositie, helderheid, beeldscherpte, etc.) in functie van gebruik van het beeldmateriaal voor communicatiedoeleinden.

In Addendum 7.6 worden mogelijke use cases verder toegelicht.

Bij dit onderzoek zal niet alleen de bruikbaarheid en kwaliteit van VRS worden beoordeeld, maar er zal ook worden onderzocht hoe de output van een VRS kan worden ingeschakeld in de werkprocessen en beheerssystemen van het MoMu.

[24] http://thesaurus.europeanafashion.eu/

7. Addenda

7.1. Overzicht VRS

Onderstaande tabel geeft een overzicht van de functies van verschillende VRS systemen die als SAAS worden aangeboden. De gegevens werden verzameld in januari - februari 2019.

NAAM	Google Cloud Vision	IBM Watson Visual Recognition	Amazon Rekognition	Microsoft Computer Vision API	Clarif.ai
URL	https://cloud.google.com/vision/	https://www.ibm.com/watson/services/visual-recognition/	https://console.aws.amazon.com/rekognition/home?region=us-east-1#/	https://docs.microsoft.com/en-gb/azure/cognitive-services/Computer-vision/	https://clarifai.com/
Technologie					
REST API	y	y	y	y	y
GUI	y	y	y	y	y
Software libraries	y	y	y	y	y
local	n	n	n	n	y
FOSS	n	n	n	n	n
Image source: url	y	y	y (https)	y	y
Image source: offline	y	y	y	y	y
max size	20 MB	10 MB	5MB	4MB	3.6MB
max rate (sec)	30	100	50	10	10
supported formats	JPEG, PNG, (animated) GIF, BMP, WEBP, RAW, ICO, PDF, TIFF	JPEG, PNG, GIF, TIFF	JPEG, PNG	JPEG, PNG, GIF, BMP	JPEG, PNG, TIFF, BMP, WEBP, TIFF
requirements	Aanbevolen grootte is 640x800, 1600x1200 voor face detection en 1024x768 voor text detection	Minimum aanbevolen pixel density is 32x32, maar betere resultaten met minstens 224x224 pixels	Aanbevolen resolutie van minstens 640x480 pixels	Beelden moeten groter zijn dan 50x50 pixels; voor OCR tussen 50x50 en 4200x4200 px	aanbevolen 512 pixels (grotere beelden worden gedreduceerd)
Functionaliteit					
type	y	n	n	y	y
face detection	y	y	y	y	y
face: sex	n	y	y	y	y

face: age	n	y	y	y	y
face: emotion	y	n	y	y	n
face: similar	n	n	y	y	n
face recognition (matching)	n	n	y	y	n
object detection	y	y	y	y	y
image tagging	y	y	y	y	y
image categorization	n	n	n	y	n
text (OCR)	y	y	n	y	n
text (label detection)	y	y	y	y	y
text (logo)	y	n	n	n	y
color (term)	y	y	y	y	y
color (code)	y	n	n	y	y
ER: people	n	n	y	y	y
NER: landmarks	y	n	n	y	n
aboutness: sfeer	y	n	y	n	y
content moderation (nsfw)	y	y	y	y	y
NL-description	n	n	n	y	n
language: EN	y	y	y	y	y
language: NL	n	n	n	n	y
models: default	y	y	y	y	y
models: custom	y	y	n	y	y
confidence ratio	y	y	y	y	y
web search	y	n	n	n	n
video analysis	y/n	n	y	y	y
Prijs					
#gratis calls	1000/maand	1000/maand	1000/maand	5000/maand; 20/min	5000/maand; 5/sec
prijs per 1000 calls	$ 1,50	$ 2,00	$ 1,00	€ 2,11	$ 1,20
formule(s)	https://cloud.google.com/vision/pricing?hl=nl	https://www.ibm.com/watson/services/visual-recognition/pricing/index.html#pricing	https://aws.amazon.com/rekognition/pricing/	https://azure.microsoft.com/en-us/pricing/details/cognitive-services/computer-vision/	https://clarifai.com/pricing

7.2. Tabel: top 20 descriptoren per VRS model en menselijk descriptoren

Overzicht van de meest frequent voorkomende descriptoren voor elk van de gebruikte services (uitgezonderd Clarifai-color) en de reeds aanwezige metadata. Tussen haakjes het aantal keer dat ze voorkomen.

Clarifai - accessories	**Clarifai - general**	**Google Vision**	**Azure**	**ORIGINAL**
bra (5)	wear (121)	clothing (51)	fashion (36)	handtas (8)
women's hat (5)	fashion (111)	dress (27)	clothing (23)	schoen (8)
cocktail dress (4)	no person (104)	fashion model (17)	black (18)	onderjurk (7)
pumps (3)	one (81)	fashion (16)	person (18)	jurk (6)
satchel (3)	woman (72)	runway (15)	white (18)	walter van beirendonck (5)
socks (3)	retro (64)	day dress (13)	accessory (9)	weefsel (5)
wide leg pants (3)	elegant (52)	handbag (13)	brown (9)	ann demeulemeester (4)
women's scarf (3)	isolated (46)	fashion show (10)	blue (8)	japon (4)
blouse (2)	design (45)	footwear (10)	footwear (8)	kraag (4)
casual dress (2)	adult (44)	shoe (10)	grey (8)	beha (3)
men's hat (2)	vintage (41)	pattern (9)	red (7)	dries van noten inspirations (3)
sleepwear (2)	style (36)	sandal (7)	dress (6)	hoed (3)
umbrella (2)	mannequin (34)	briefs (5)	shoe (6)	jak (3)
women's sandals (2)	dress (32)	brown (5)	art (5)	kapmantel (3)
women's short shorts (2)	fashionable (29)	cobalt blue (5)	skirt (5)	kostuum (man) (3)
blazer (1)	old (29)	needlework (5)	coat (4)	kous (3)
bodysuit (1)	people (27)	outerwear (5)	pattern (4)	raf simons (3)
capris (1)	man (23)	sleeve (5)	textile (4)	sjaal (3)
fur coat (1)	antique (21)	textile (5)	a person wearing a dress (3)	stalenboek (3)
gloves (1)	decoration (20)	yellow (5)	book (3)	waaier (3)

7.3. Tabel: verhouding correcte descriptoren op totaal (ruis)

Clarifai general

	Min	Max	Mean
valid tags (N=1644)	2	18	10.08
total tags (N=2976)	5	20	18.25
relevance score	0.15	1.0	0.55

Clarifai colors

	Min	Max	Mean
#valid (N=182)	0	2	1.11
#tags (N=452)	1	3	2.77
relevance score	0.0	1.0	0.41

Clarifai accessories

	Min	Max	Mean
valid tags (N=54)	0	3	0.32
total tags (N=139)	0	3	0.84
relevance score	0.0	1.0	0.41

Google Vision

	Min	Max	Mean
valid tags (N=414)	0	5	2.54
total tags (N=787)	1	6	4.83
relevance score	0.0	0.83	0.43

Microsoft

	Min	Max	Mean
valid tags (N=237)	0	6	1.46
total tags (N=714)	0	12	4.38
relevance score	0.0	0.75	0.28

7.5. Voorbeelden geanalyseerde beelden

Opname op paspop © RV

Stalenboek © RV

Mode-illustratie © RV

Textiel © RV

Backstage fashion event © RV

Detailopname © RV

Catwalkopname © Catwalkpictures

Scenografie © RV

Kostuumbroek © RV

7.6 Voorbeelden van use cases

Ondersteuning bij objectregistratie

Gefotografeerde objecten worden door VRS van metadata voorzien om objectregistratie te ondersteunen of te vervangen.	
resultaat:	- descriptoren per object
vereist:	- individueel gefotografeerde objecten, d.w.z. goed te onderscheiden van achtergrond - collectie bij voorkeur met grote diversiteit - relatief courant voorkomende objecten
aanvullende acties:	- filteren adhv black- of whitelist - reconciliëren met thesaurus (AAT)
voorbeelden:	- industrieel erfgoed - gebruiksvoorwerpen

Tagging met op basis van Linked data trainingset

Een VRS wordt getraind op basis van een externe beeldcollectie die werd beschreven met Linked Data. Het getrainde model wordt toegepast op een vergelijkbare beeldcollectie, die daardoor aan de hand van Linked Data beschreven is.	
resultaat:	- Linked data descriptoren per object
vereist:	- individueel gefotografeerde objecten, d.w.z. goed te onderscheiden van achtergrond - inhoudelijk vergelijkbare collectie met metadata die als Linked Open Data gepubliceerd is.
aanvullende acties:	-
voorbeelden:	- Europeana Fashion collection met Fashion Thesaurus als trainingset, toegepast op beeldbank van het MoMu.[25, 26]

Analyse van beeldkwaliteit

Beeldkwaliteit wordt geëvalueerd op basis van esthetische en technische kenmerken (scherpte, resolutie,, contrast, kleurdiepte …)	
resultaat:	- ‘esthetische’ kwalificatie per beeld
vereist:	- beeldmateriaal van wisselende kwaliteit

[25] http://thesaurus.europeanafashion.eu/
[26] https://www.europeana.eu/portal/nl/collections/fashion

aanvullende acties:	- vertalen naar (en eventueel implementeren van) filter in het beeldbeheerssysteem
voorbeelden:	- beeldbank gebruikt voor communicatie en publicatie - selectie van (primair) beeld dat op collectiewebsite wordt getoond

Pre-iconografische en iconografische beschrijving van beelden

Beeldmateriaal wordt geanalyseerd en voorzien van descriptoren die de inhoud (afgebeelde objecten), aard, sfeer en onderwerp van de afbeelding weergeven	
resultaat:	- descriptoren
vereist:	- inhoudelijk geschikt voor iconografische beschrijving (narratief)
aanvullende acties:	- filteren adhv black- of whitelist - reconciliëren met thesaurus (AAT, Iconclass)
voorbeelden:	- kunsthistorische collecties (bv. grafiek) - documentaire fotocollecties

Automatische classificatie van beeldcollecties d.m.v. topic detection

Beeldmateriaal wordt van descriptoren voorzien. Op basis van deze descriptoren worden de beelden geclusterd. Vooraf wordt het aantal clusters bepaald, maar het onderwerp (topic) wordt door de AI toepassing bepaald.	
resultaat:	- aan elk beeld wordt een topic toegekend, met aanduiding van de waarschijnlijkheid
vereist:	- omvangrijke, diverse collectie
aanvullende acties:	- optimaliseren van de topic detection door parameters aan te passen - visualiseren van clusters
voorbeelden:	- inhoudelijk clusteren van omvangrijke beeldcollecties

Identificeren van beeldmateriaal in tekstdocumenten

VRS analyseert pagina's van kranten, tijdschriften of boeken om hierin aanwezig beeldmateriaal te vinden en van descriptoren te voorzien. Descriptoren kunnen betrekking hebben op inhoud of op formele kenmerken (zwart/wit vs. kleur, binair vs grayscale, …)	
resultaat:	- dataset met referentie naar paginas met afbeelding, eventueel met descriptoren
vereist:	- omvangrijke, gedigitaliseerde collectie publicaties
aanvullende acties:	- extraheren van beeldmateriaal - toekennen

voorbeelden:	- collectie tijdschriften ontsluiten

Doorzoekbaarheid van omvangrijke collecties ondersteunen door tagging van deelcollecties

VRS kent descriptoren toe aan omvangrijke beeldcollecties (bv. fotoarchief) met het oog op snel, goedkoop en efficiënt vindbaar maken van het beeldmateriaal. Daarbij wordt een grote mate van ruis aanvaard. De descriptoren worden gebundeld per deelcollectie	
resultaat:	- descriptoren toegekend aan beelden - descriptoren gebundeld per deelcollectie
vereist:	- omvangrijke beeldcollectie met deelcollecties rond verschillende thema's - beeldcollectie is ingedeeld in deelcollecties (op basis van bestaande metadata, bestandsnaam, locatie …)
aanvullende acties:	-
voorbeelden:	- fotoarchief doorzoekbaar maken op deelcollectieniveau

Entity recognition

Herkennen van personen en locaties ('landmarks')	
resultaat:	- beeldcollectie voorzien van descriptoren en identifiers die naar persoon en/of landmark verwijzen
vereist:	- hoge kwaliteit - beeldcollectie met gemakkelijk te identificeren personen en/of locaties
aanvullende acties:	- training van VRS voor personen/locaties die niet gekend zijn. - reconciliëren met personen authority (VIAF, WD)
voorbeelden:	- fotoarchief doorzoekbaar maken op persoon (politicus, kunstenaar, model) - fotoarchief doorzoekbaar maken op locatie

8. Bronnen

Armitage, L. H., & Enser, P. G. (1997). Analysis of user need in image archives. Journal of information science, 23(4), 287-299.

Bakewell, Elizabeth, William O. Beeman, & Carol McMichael Reese. 1988. Object, Image, Inquiry: The Art Historian at Work. Santa Monica: The Getty Art History Information Program.

Dee, H.M., Hughes, L.M. , Roderick, G.L. and Brown, A.D. (2016) Visual digital humanities: using image data to derive approximate metadata. In: Foster, A. and Rafferty, P. (eds.) Managing Digital Cultural Objects. facet publishing. ISBN 9781783301027

Crowley, E. & Zisserman, A. 2014, The State of the Art: Object Retrieval in Paintings using Discriminative Regions. in British Machine Vision Conference, 2014.

Filestack. 2019. Comparing Image Tagging Services: Google Vision, Microsoft Cognitive Services, Amazon Rekognition and Clarifai. https://blog.filestack.com/thoughts-and-knowledge/comparing-google-vision-microsoft-cognitive-amazon-rekognition-clarifai/ [2019-05-07]

Grauman, Kristen, and Bastian Leibe. 2010. "Visual Object Recognition." Synthesis Lectures on Artificial Intelligence and Machine Learning 5(2): 1–159. http://www.morganclaypool.com/doi/abs/10.2200/S00332ED1V01Y201103AIM011.

Lin, Yiqi. 2018, Evaluation Methods and Applications in Image Recognition based on Convolutional Neural Networks, in Journal of Physics: Conference Series, 1087 062023

Oomen, Johan, and Lora Aroyo. 2011. "Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges." In *Proceedings of the 5th International Conference on Communities and Technologies*, C&T '11, New York, NY, USA: ACM, 138–49. http://doi.acm.org/10.1145/2103354.2103373.

Shatford, Sara. 1986, "Analyzing the Subject of a Picture: A Theoretical Approach" Cataloging & Classification Quarterly, v. 6(3), Spring 1986

Wang, Z., Chang, S., Yang, Y., Liu, D., & Huang, T. S. 2016, Studying very low resolution recognition using deep networks. In CVPR. IEEE.

Wein, Leonard. 2014. "Visual Recognition in Museum Guide Apps: Do Visitors Want It?" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, New York, NY, USA: ACM, 635–38. http://doi.acm.org/10.1145/2556288.2557270.

Yao, Mariya. 2017. Chihuahua Or Muffin? Searching for the best computer vision API. https://www.topbots.com/chihuahua-muffin-searching-best-computer-vision-api/