



Faculteit Bedrijf en Organisatie

Het gebruik van Computer Vision API's voor de beschrijving van cultureel-erfgoedcollecties

Nastasia Vanderperren

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Koen Mertens
Co-promotor:
Henk Vanstappen

Instelling: Huis van Alijn

Academiejaar: 2018-2019

Derde examenperiode

Faculteit Bedrijf en Organisatie

Het gebruik van Computer Vision API's voor de beschrijving van cultureel-erfgoedcollecties

Nastasia Vanderperren

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Koen Mertens
Co-promotor:
Henk Vanstappen

Instelling: Huis van Alijn

Academiejaar: 2018-2019

Derde examenperiode

Woord vooraf

Bedankt liefje om dit te willen nalezen! <3

Samenvatting

nog te doen

Inhoudsopgave

1	Inleiding	15
1.1	Probleemstelling	15
1.2	Onderzoeksvraag	16
1.3	Onderzoeksdoelstelling	16
1.4	Opzet van deze bachelorproef	17
2	Stand van zaken	19
2.1	Computer vision	19
2.2	Computer Vision API	21
2.3	Computer vision voor cultureel erfgoed	22
2.3.1	Uitdaging	23
2.3.2	Resultaten	24
2.4	Experimenten met beeldherkenning in musea	25

3	Methodologie	29
3.1	Opstellen use cases	29
3.2	Analyseren van de dataset	30
3.3	Keuze voor API	32
3.4	Beelden beschikbaar maken via een publieke URL	32
3.5	Beschikbare metadata verzamelen in een CSV	33
3.6	Uitvoeren van workflows	34
3.6.1	Workflow 1: Een dataset laten taggen door Clarifai	34
3.6.2	Workflow 2: Een custom model trainen	37
4	Resultaten met het ingebouwd model	51
4.1	Gedetailleerde resultaten per thema	53
4.1.1	Geboorte	53
4.1.2	Huwelijk	56
4.1.3	Sinterklaas	57
4.1.4	Speelgoed	58
4.1.5	Vakantie	60
4.2	Vergelijking met de beschrijvingen van registratoren	60
4.3	Toepassen in de praktijk?	63
4.3.1	Instellen van een drempelwaarde	64
4.3.2	Selectie maken van thema's	64
5	Resultaten met de custom modellen	67
5.1	Themamodel	67
5.2	Periodemodel	68

5.3 Toepassen in de praktijk?	71
5.3.1 Instellen van een drempelwaarde	72
5.3.2 Rekening houden met de waarschijnlijkheidsscore van de tweede tag	72
6 Conclusie	75
A Onderzoeksvoorstel	77
A.1 Introductie	77
A.2 Stand van zaken	78
A.2.1 Computer Vision en (cultureel) erfgoed	78
A.2.2 Theoretisch onderzoek	79
A.2.3 Computer Vision API's	80
A.3 Methodologie	81
A.4 Verwachte resultaten	81
A.5 Verwachte conclusies	82
Bibliografie	85

Lijst van figuren

2.1 <i>The Next Rembrandt</i> , een valse Rembrandt gecreëerd door een AI-algoritme	23
2.2 Installatiezicht van MOMA waarin de AI-diensten van Google Arts & Culture een schilderij van Cézanne herkennen	25
3.1 Staafdiagrammen met het aantal beelden per thema en de verhouding van het aantal periodes per thema	31
3.2 Validatie van het ingebouwd model in Google Sheets.	38
3.3 Staafdiagram met het aantal beelden per periode	39
3.4 Trainingsbeelden toevoegen via de webinterface	40
3.5 Een model maken via de webinterface	42
3.6 Het model trainen via de webinterface	45
3.7 Validatie van het custom model in Google Sheets.	46
4.1 Enkele voorbeelden van foto's die goed of slecht getagged werden door het ingebouwde Clarifai-model	52
4.2 Best en slechtst scorende foto van thema geboorte	55
4.3 Best en slechtst scorende foto van thema huwelijk	56
4.4 Best en slechtst scorende foto van thema Sinterklaas	58
4.5 Best en slechtst scorende foto van thema speelgoed	59

4.6 Best en slechtst scorende foto van thema vakantie	61
4.7 Het aantal ingevulde velden in de bestaande beschrijvingen gemaakt door de registratoren van Huis van Alijn	63
4.8 Vergelijking van waarschijnlijkheidsscores juiste en foute tags	65
5.1 Voorbeeld van foto's van de thema's Vakantie en Speelgoed die gelijkend zijn	69
5.2 Vergelijking van de waarschijnlijkheidsscores van de juiste en foute classificaties van het custom model	73
5.3 Vergelijking van de waarschijnlijkheidsscores van de juiste en foute classificaties van het custom model	74
A.1 Het overzicht van de velden van de basisregistratie en het aantal keer dat ze aanwezig zijn in de collectiedata van MSK Gent.	83
A.2 Vrouw en man met boek worden door Microsoft Computer Vision geïdentificeerd als man en vrouw die een selfie nemen (Roddis, 2018).	84

Lijst van tabellen

3.1	opdeling van het aantal foto's per thema en decennia	30
3.2	De F_1 -scores voor de verschillende iteraties van het Themamodel.	47
3.3	De resultaten (rappel, precisie en F_1 -score) van de vierde iteratie bij het trainen van het Themamodel.	47
3.4	De F_1 -scores voor de verschillende iteraties van het Periodemodel.	48
3.5	De resultaten (rappel, precisie en F_1 -score) van de tweede iteratie bij het trainen van het Periodemodel.	48
4.1	een overzicht van de resultaten per thema na gebruik van het ingebouwde model van Clarifai	53
4.2	een overzicht van de resultaten per periode na gebruik van het ingebouwde model	54
4.3	De dertig meest voorkomende tags van geboortefoto's met het ingebouwde Clarifai-model	54
4.4	De dertig meest voorkomende tags van trouwfoto's met het ingebouwde Clarifai-model	57
4.5	De dertig meest voorkomende tags van Sinterklaasfoto's met het ingebouwde Clarifai-model	58
4.6	De dertig meest voorkomende tags van de speelgoedfoto's met het ingebouwde Clarifai-model	59

4.7 De dertig meest voorkomende tags van de vakantiefoto's met het ingebouwde Clarifai-model	60
4.8 De dertig meest voorkomende termen gebruikt door registratoren van Huis van Alijn	64
4.9 Verhouding tussen drempelwaarde, totaal aantal van de juiste tags en het aantal foute tags	64
5.1 De performantie (rappel, precisie en F_1 -score) van het Themamodel op de volledige dataset van Huis van Alijn.	68
5.2 Confusion matrix voor het Themamodel.	70
5.3 De performantie (rappel, precisie en F_1 -score) van het Periodemodel op de volledige dataset van Huis van Alijn.	70
5.4 Confusion matrix voor het Periodemodel.	71
5.5 Verhouding tussen drempelwaarde, totaal aantal van de juiste classificaties en het aantal foute classificaties	72
5.6 Verhouding tussen een drempelwaarde voor het tweede hoogste concept, totaal aandeel van de juiste tags en het aantal foute tags ..	73

1. Inleiding

In deze bachelorproef wordt onderzocht of Computer Vision API's zoals Clarifai¹, Google Cloud Vision² of Microsoft Computer Vision³ gebruikt kunnen worden om collectiemeewerkers te ondersteunen bij het beschrijven van cultureel-erfgoedobjecten. De digitale fotocollectie van het Huis van Alijn⁴ werd hiervoor als testcase gebruikt.

Dit onderzoek zal als case mee opgenomen worden in het cultureel-erfgoedproject *Beeldherkenning in de registratiepraktijk* van FOMU. In dit project wil het museum, samen met Datable en PACKED, onderzoek voeren naar de toepassing van Visual Recognition Services (of Computer Vision API's) voor de basisregistratie en iconografische ontsluiting van erfgoedobjecten. Hiervoor maken ze gebruik van de eigen collectie, maar leveren ook Netwerk Oorlogsbronnen, MoMu en Erfgoedcel Brugge beelden aan (Derveaux, 2019).

1.1 Probleemstelling

De Vlaamse musea lijden aan een historische achterstand m.b.t. het registreren van de eigen collectie (Gatz, 2016). Het registreren van collectiestukken bestaat onder meer uit het systematisch beschrijven van de historische context en de kenmerken waarmee het object kan geïdentificeerd worden, zoals de titel, een korte beschrijving, het soort object (schilderij, foto, standbeeld, stoel), afmetingen, gebruikte materiaal,... Dit is tijdrovend werk dat door domeinexperten gedaan wordt. Vooral formele en administratieve gegevens

¹ <https://www.clarifai.com/>

² <https://cloud.google.com/vision/>

³ <https://azure.microsoft.com/nl-nl/services/cognitive-services/computer-vision/>

⁴ Voor meer info, zie <http://huisvanalijn.be/nl/collectie/fotoalbum>

worden geregistreerd. Voor het beschrijven van inhoudelijke informatie zoals afgebeelde personen of objecten, emoties en sfeer ontbreekt het de musea aan tijd en personeel, terwijl dat net de informatie is die interessant is voor ontsluiting en onderzoek.

Daarnaast is er ook een digitalisering aan de gang waarbij steeds meer collectiestukken gedigitaliseerd worden. Ook deze digitale beelden moeten voorzien worden van metadata of tags. Het doorzoeken of vinden van die digitale beelden is immers moeilijk als je bij de zoekactie geen gebruik kunt maken van trefwoorden. Dit verschilt van digitale tekstbestanden, waarbij op basis van *full text search* bestanden teruggevonden kunnen worden. Ook voor dit werk ontbreekt het de musea aan tijd en mankracht. Het gevolg is dat musea over steeds meer beelden beschikken die niet ontsloten of gebruikt kunnen worden.

Daarom willen we in deze bachelorproef onderzoeken of artificiële intelligentie (AI) de collectiemedewerker kan bijstaan in het beschrijven van de cultureel-erfgoedcollecties. Beeldherkenningssoftware is er de laatste jaren enorm op vooruitgegaan en wordt ook steeds eenvoudiger om te gebruiken.

1.2 Onderzoeksvraag

De centrale vraag in dit onderzoek is of Computer Vision API's gebruikt kunnen worden voor het (inhoudelijk) beschrijven van cultureel-erfgoedcollecties. Aan de hand van één gekozen API zal nagegaan worden of het ingebouwde model voldoende is voor beschrijving van de beelden en of het eenvoudig is om de beeldherkenningssoftware te trainen indien de beschrijving van het ingebouwde model niet voldoet.

We willen hierbij ook te weten komen of de software eenvoudig in gebruik is zodat museum-medewerkers zelf de API's kunnen gebruiken om hun beelden te *taggen*. Gebruiksgemak zal daarom een belangrijk aspect zijn voor de keuze van de API. Een ander belangrijk aspect is de aanwezigheid van goede documentatie en tutorials.

1.3 Onderzoeksdoelstelling

Het onderzoek zal uitgevoerd worden op een deel van de fotocollectie van het Huis van Alijn. Het Huis van Alijn is het museum van het dagelijkse leven en beschikt over een grote fotocollectie die het dagelijkse leven in de twintigste eeuw documenteert⁵. Samen met Huis van Alijn zullen er use cases geformuleerd worden waarvoor beeldherkenning ingezet kan worden.

We focussen ons in de eerste plaats op fotocollecties omdat we vermoeden dat de Computer Vision API's hier het best op scoren. Een *proof of concept* zal opgezet worden om de beelden te laten taggen en de beeldherkenningssoftware te trainen. Vervolgens zullen we

⁵ <http://huisvanalijn.be>

een vergelijking maken tussen de metadata van de beelden die aangeleverd werden door de API en de metadata van de beelden zoals ze door het museummedewerker aangeleverd werden.

Met dit onderzoek wordt de haalbaarheid van het gebruik van Computer Vision API's in de erfgoedsector onderzocht. Tevens zijn de use cases gekend waarvoor beeldherkennings-software een goede partner is.

1.4 Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomain, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte technieken besproken om een antwoord te kunnen formuleren op de onderzoeks vragen.

In Hoofdstuk 4 worden de resultaten besproken van het loslaten van een ingebouwd model van een Computer Vision API op gedigitaliseerde dia's en foto's van het Huis van Alijn.

Vervolgens gaan we een stap verder en wordt een eigen model gemaakt en losgelaten op diezelfde beelden van Huis van Alijn. In Hoofdstuk 5 worden de resultaten hiervan geanalyseerd.

In Hoofdstuk 6, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeks vragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2. Stand van zaken

In dit hoofdstuk wordt uitgelegd wat Computer Vision en Computer Vision API's zijn en inhouden. Tevens wordt een stand van zaken gegeven van het onderzoek naar beeldherkenning voor cultureel-erfgoedcollecties, de uitdagingen hierbij en experimenten met beeldherkenningsstoepassingen in musea.

2.1 Computer vision

Computer Vision (CV) is het onderzoeksgebied waarin technieken ontwikkeld worden om computers te helpen bij het zien en het begrijpen van de inhoud van digitale beelden zoals foto's en video's. Het is een deelgebied van Artificiële Intelligentie en Machine Learning (Wikipedia contributors, 2019a). Het doel van deze techniek is om computers te gebruiken voor taken met digitale beelden of video's die normaal door mensen gedaan zouden worden. Computer Vision wordt momenteel met succes toegepast voor een breed scala van uitdagingen zoals inspectie van machines, geautomatiseerde kassa-afrekeningen in de retail, motion capture, medische beeldvorming, bewaking en politieonderzoek, het herkennen van vingerafdrukken en het assisteren van mensen bij het identificeren van de inhoud van een foto of video (Brownlee, 2019a).

De bovenvermelde uitdagingen gebruiken verschillende CV-taken voor het verkrijgen en analyseren van informatie uit de beelden, zoals beeldherkenning en bewegingsanalyse. Voor het onderzoek van deze bachelorproef is beeldherkenning de belangrijkste taak. Enkele typische voorbeelden hiervan zijn (Wikipedia contributors, 2019a):

- OCR: het identificeren van tekens in beelden van teksten;
- object classification: het classificeren van het beeld in een bepaalde categorie;

- object recognition: het identificeren van objecten in het beeld;
- object detection: het herkennen van een object of entiteit (zoals een persoon) en tevens aanduiden op het beeld waar het zich bevindt;
- face detection en recognition: het herkennen van personen en aanduiden waar ze zich op het beeld bevinden;
- content-based image retrieval: het vinden van beelden in een dataset met een bepaalde inhoud, bv. het vinden van gelijkaardige beelden.

Momenteel zijn algoritmes die gebaseerd zijn op Convolutional Neural Networks (CNN)⁶ het best om deze taken uit te voeren. CNN voor beeldherkenning werd voor het eerst met succes geïntroduceerd in 2012. Sindsdien kent het veld van beeldherkenning een enorme groei. Artificiële neurale netwerken, zoals CNN, leren taken uit te voeren op basis van een trainingsset. Als je wil dat het netwerk een bus herkent, dan bestaat de trainingsset uit verschillende voorbeelden van een bus, maar ook van wat een bus niet is. Wanneer, na de training, het model nieuwe beelden ziet, zou het moeten kunnen zeggen of het afgebeelde object een bus is of niet. Als het netwerk nog een andere taak moet uitvoeren, dan moet het netwerk nieuwe trainingsdata krijgen, bijvoorbeeld om te leren wat een fiets is (Pokharna, 2016).

ImageNet is de dataset die meestal gebruikt wordt voor het ontwikkelen van toepassingen voor Computer Vision. ImageNet is een visuele databank die bestaat uit meer dan veertien miljoen beelden die annotaties hebben over wat er op de beelden staat. De databank voorziet 20.000 concepten, zoals ballon of aardbei. Elk van deze concepten bestaat uit enkele honderden beelden (Wikipedia contributors, 2019b; Brownlee, 2019b). Sinds 2010 vindt jaarlijks de ImageNet Large Scale Visual Recognition Challenge (ILSVRC) plaats. Hierbij wordt een beperkte set van ImageNet (één miljoen beelden voor duizend concepten) gebruikt om de performantie van de verschillende algoritmes te testen. Het valt hierbij op dat de resultaten jaarlijks enorm stijgen. Terwijl er in 2011 een foutenmarge was van 25% voor de uitdaging van image classification, was dit tegen 2017 al teruggevallen tot 5%.

Russakovsky e.a. (2014) stellen vast dat domeinexperten beter blijven dan de CV-modellen, maar dat de CV-modellen significant beter scoren dan mensen die geen expert in het domein zijn. Mensen zijn bijvoorbeeld goed in het onderscheiden van een hond en een kat, maar een CV-model kan veel beter de verschillende hondenrassen herkennen. De modellen waren in het algemeen erg goed in het herkennen van verschillende soorten zoogdieren en hun rassen, maar hadden het veel moeilijker met metalen en doorzichtige voorwerpen. Ze hebben ook moeilijkheden met beelden die in meer dan vijf verschillende concepten te classificeren zijn⁷ en met het identificeren van kleine objecten. Daarnaast hebben ze ook problemen met foto's die vervormd zijn door het gebruik van kleurenfilters of met abstracte voorstellingen van voorwerpen zoals 3D-rendered beelden, schilderijen, tekeningen en schetsen. Aangezien schilderijen, tekeningen, schetsen en objecten algemeen voorkomen in erfgoedcollecties, kan dit een probleem vormen.

⁶ Artificiële neurale netwerken zijn netwerken van verbonden neuronen die het menselijke brein nabootsen. Een CNN is een type neurale netwerk waarbij niet alle neuronen met elkaar verbonden zijn en dat vooral toegepast wordt voor het analyseren van visuele beelden.

⁷ De onderzoekers stelden vast dat ook menselijke classificatoren het hier moeilijk mee hadden.

2.2 Computer Vision API

CV API's zijn REST API's⁸ die het developers mogelijk maken om machine learning toe te passen zonder hier zelf een expert in te zijn. De services zijn zelf al getraind, waardoor het niet nodig is om zelf trainingsbeelden te voorzien en modellen te trainen. De meeste services bieden ook de mogelijkheid aan om eigen modellen te creëren. De diensten zijn zo toegankelijk dat iedere ontwikkelaar zonder machine learning expertise modellen kan bouwen. Volgens Lardinois (2018) zijn de services zo eenvoudig dat zelfs mensen zonder enige programmeerervaring de diensten kunnen gebruiken. De bekendste services zijn Clarifai⁹, IBM Visual Recognition¹⁰, Microsoft Computer Vision¹¹, Amazon Rekognition¹² en Google Cloud Vision¹³.

Welke datasets gebruikt zijn om de CV API te trainen is niet duidelijk. Wel hebben zowel Clarifai, Google als Microsoft deelgenomen aan de ImageNet Challenge. In 2013 was ZFNet (Clarifai) de winnaar van de beeldclassificatietaak, in 2014 behaalde Google topresultaten voor objectdetectie met hun GoogLeNet-model en in 2015 behaalde Microsoft Research topresultaten met het ResNet-model. Op basis hiervan kunnen we veronderstellen dat de services minstens ImageNet gebruikt hebben (Brownlee, 2019b).

Het gebruik van de CV API's is zeer toegankelijk. De meeste van deze CV API's beschikken over een webinterface waarmee de service uitgetest kan worden en waarmee custom modellen gebouwd kunnen worden. Voor intensiever gebruik is er een REST API. Daarnaast hebben ze ook API clients¹⁴ in verschillende programmeertalen (o.a. Java, Python, Javascript/Node.js, C#, PHP) die het de programmeur toelaat om de services te integreren in applicaties of scripts in een programmeertaal naar keuze. Beeldmateriaal kan op verschillende manieren aan de services bezorgd worden:

- via een URL naar het beeld;
- in de vorm van base64;
- opgeslagen in het ecosysteem van de aanbieder (bv. S3 buckets voor Amazon Rekognition, of Google Cloud Storage buckets voor Google Cloud Vision).

De services kunnen verschillende aspecten van een beeld beschrijven:

- OCR: het omzetten van tekens in beelden naar tekst;
- object detection: beschrijven van op het beeld aanwezige objecten of personen;
- algemene beschrijving van het soort beeld;
- de aard van het beeld (foto, tekening, schilderij)
- gezichtsherkenning: identificatie van bekende personen;
- landmarks: identificatie van bekende plaatsen;

⁸API's zijn een set van commando's om met een webservice of computersysteem te communiceren.

⁹<https://www.clarifai.com/>

¹⁰<https://www.ibm.com/watson/services/visual-recognition/>

¹¹<https://azure.microsoft.com/nl-nl/services/cognitive-services/computer-vision/>

¹²<https://aws.amazon.com/rekognition/>

¹³<https://cloud.google.com/vision/>

¹⁴API clients zijn implementaties van een API in een bepaalde programmeertaal.

- aboutness: emoties, sfeer en thema van het beeld;
- gebruikte kleuren, zowel hoofdkleuren als accentkleuren;
- NSFW-score: score op de aanwezigheid van content waaraan aantreft genomen kan worden zoals erotiek, seksisme, racisme en geweld.

De resultaten worden meestal in de vorm van tags gegeven, maar ook taglines (een korte zin die het beeld beschrijft) of codes (kleurcodes) zijn mogelijk, net als URI's (voor gelijkaardige beelden die op het web gevonden zijn). Bij de resultaten wordt steeds een waarschijnlijkheidsscore gegeven (Vanstappen, 2019).

Na een vergelijking van de top vijf van CV API's was Oberoi (2016) onder de indruk van de resultaten. De services waren voldoende om informatie te krijgen over de kern van een beeld op een erg snelle en relatief goedkope manier.

2.3 Computer vision voor cultureel erfgoed

Er is al veel onderzoek gedaan naar het gebruik van Computer Vision voor cultureel-erfgoed. Omwille van de digitalisering blijven de beeldcollecties in musea en cultureel-erfgoedinstellingen groeien waardoor er een enorme registratie-achterstand is. Van CV wordt verwacht dat het werk van de registratoren verlicht en de achterstanden wegwerkt. Computers worden hiervoor getraind als domeinexperten. Ze moeten van beelden van kunstwerken of erfgoedobjecten kunnen bepalen:

- wie de kunstenaar of maker is;
- uit welk materiaal de werken gemaakt zijn;
- in welke periode of jaar het werk gemaakt is;
- welk soort object het werk is;
- en tot welke kunststroming het werk behoort.

In zulke projecten trainen onderzoekers een eigen model of classifier om beelden van erfgoedobjecten te classificeren of te labelen. Een voorbeeld van zo'n onderzoek is het in 2017 gestarte Belgische INSIGHT-project waarin men wil onderzoeken of AI gebruikt kan worden voor het beschrijven voor museale collecties. De digitale collecties van de Koninklijke Musea voor Schone Kunsten van België en de Koninklijke Musea voor Kunst & Geschiedenis worden als testcase gebruikt. Het einddoel van dit project is de ontwikkeling en release van een reeks praktische machine learning tools om digitale collecties te beheren (Universiteit Antwerpen, 2017).

Naast deze onderzoeken wordt ook onderzocht of CV gebruikt kan worden voor het herkennen van vervalsingen (Dickson, 2018), of om te verifiëren of een kunstenaar een bepaald werk gemaakt heeft op basis van penseelstreken en potloodstrepen (ook om vervalsingen te ontdekken) (Elgammal, 2017). Daarnaast wordt AI gebruikt om nieuwe kunstwerken te maken (Dickson, 2019). ING, Microsoft, TU Delft en Mauritshuis hebben in 2016 bijvoorbeeld zo een nieuw kunstwerk van Rembrandt gemaakt op basis van een databank van zijn schilderijen (ING, 2016).



Figuur 2.1: *The Next Rembrandt*, een valse Rembrandt gecreëerd door een AI-algoritme.
Credit: Microsoft/ING/Delft University of Technology

2.3.1 Uitdaging

Een probleem bij het gebruik van computer vision voor erfgoedobjecten zijn de veel kleinere datasets:

- Blessings en Wen (2013) creëerden een dataset van 1.400 beelden via Google Images van de kunstenaars Cézanne, Dali, Dürer, Monet, Picasso, Rembrandt en Van Gogh (tweehonderd beelden per kunstenaar);
- Mensink en van Gemert (2014) ontwikkelden een dataset van meer dan 112.000 beelden van het Rijksmuseum die representatief is voor het museumcollectie. Het bestond uit 1.824 objecttypes waarvoor 406 verschillende soorten materialen gebruikt waren en overspande een periode van het begin van onze jaartelling tot de late 19^e eeuw. De kunstwerken waren afkomstig van meer dan 6.000 kunstenaars. Het ging onder meer over schilderijen, foto's, keramiek, en meubels.
- Sabatteli, Kestemont, Daelemans en Geurts (2018) gebruikten die dataset van Mensink en Gemert, maar vulden dit aan met een kleinere dataset uit het DAMS Antwerpen.¹⁵ In totaal bestond de dataset uit ongeveer 100.000 beelden van kunstwerken die op te delen zijn in 2.016 materialen van kunstwerken, 135.809 beelden van kunstwerken die op te delen zijn in 1.677 objectsoorten en 90.674 beelden van

¹⁵Het DAMS Antwerpen is de online beeldbank van Musea en Erfgoed Antwerpen waarmee de collecties van de musea, bibliotheken, archieven en erfgoedinstellingen van de Stad Antwerpen in de kijker geplaatst wordt. Zie <https://dams.antwerpen.be>.

kunstwerken van 2.099 kunstenaars.

- Elgammal, Mazzone, Liu, Kim en Elhoseiny (2018) gebruikten een dataset van 76.921 beelden van schilderijen afkomstig van 1.119 kunstenaars uit de 15^e eeuw tot de hedendaagse tijd.

Dit stelt weinig voor als je dit vergelijkt met de veertien miljoen beelden van ImageNet. Die beperking heeft te maken met copyright issues op beelden van kunstwerken en de afwezigheid van goede metadata bij de wel beschikbare beelden.

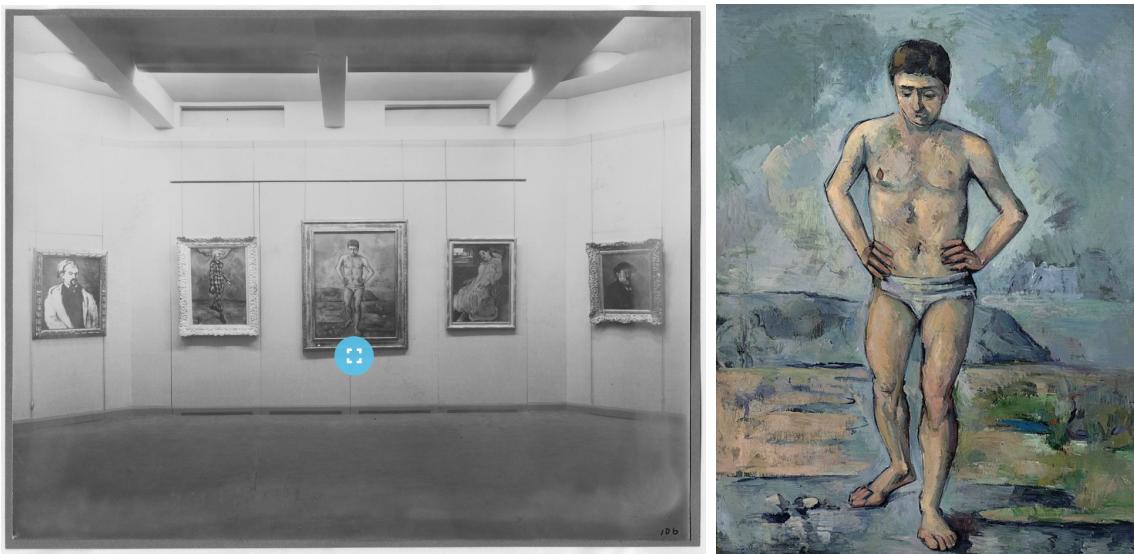
Om dit probleem aan te pakken, verkenden zowel Sabatteli e.a. (2018) als Elgammal e.a. (2018) het veld van Transfer Learning. Hierbij worden geen neurale netwerken *from scratch* gebouwd, maar worden reeds ontwikkelde netwerken verder gefinetuned. De onderzoekers gebruikten hiervoor neurale netwerken die *state of the art* waren op de ImageNet Challenge. De artificiële netwerken werden verder gefinetuned met de testdata. Beide onderzoeken toonden aan dat het beter was om de artificiële netwerken te finetunen dan zelf een model te bouwen (Sabatteli e.a., 2018). Elgammal e.a. (2018) vermoedden dat betere resultaten mogelijk zijn met *from scratch* netwerken als er grotere datasets ter grootte van ImageNet voor erfgoedobjecten ter beschikking zouden zijn.

2.3.2 Resultaten

Voor het classificeren van kunstwerken per kunstenaar behaalden Blessings en Wen (2013) goede resultaten (85% correctheid). Hun onderzoek concentreerde zich wel maar op zeven kunstenaars. Bij de onderzoeken met meer kunstenaars, zoals bij van Mensink en van Gemert (2014) en Sabatteli e.a. (2018) waren de resultaten lager. Bij die laatste werd een maximumscore van 69% behaald.

Het classificeren van werken per materiaal werd beschouwd als de eenvoudigste classificatie, onder meer omdat het aantal klassen vijf keer kleiner is dan bij de kunstenaarsclassificatie en typeclassificatie. Sabatteli e.a. (2018) vermoedden dat deze uitdaging ook het meest overeenstemt met de uitdagingen voor de ImageNet Challenge. Ze haalden hier een resultaat van 92,95% op. Mensink en van Gemert (2014) behaalden een resultaat van 94%. Ook voor het classificeren van kunstwerken per objecttype werden in beide onderzoeken gelijkaardige resultaten behaald.

Tot slot was ook het classificeren op basis van kunststroming succesvol. De uitdaging bestond uit het classificeren van 76.921 beelden in twintig kunststromingen. Er werd een score behaald van maximum 73%. Ondanks dat het CV-model geen kennis heeft van kunstenaars, creatiedatum of geschiedenis van stijlen, worden de schilderijen uit dezelfde periode bij elkaar geplaatst en is er een geleidelijke overgang van stijlen te zien vanaf de Renaissance (15^e eeuw) tot de abstracte kunst (20^e eeuw) (Elgammal e.a., 2018).



Figuur 2.2: In een installatiezicht van de tentoonstelling "Cézanne, Gauguin, Seurat, Van Gogh" uit 1929 in het MOMA werd het werk *De Bader* van Cézanne herkend (foto tentoonstelling: Photographic Archive. The Museum of Modern Art Archives, New York. IN1.7. Foto door Peter A. Juley.).

2.4 Experimenten met beeldherkenning in musea

Verschillende musea zijn reeds aan de slag gegaan met beeldherkenning via Computer Vision API's. De use cases zijn anders dan bij het (academisch) onderzoek. Musea willen de bezoekers van hun website nieuwe ervaringen aanbieden om hun beeldcollectie te ontdekken, maar wensen tevens hun beelden beter doorzoekbaar te maken. Uit onderzoek blijkt namelijk dat gebruikers de behoefte hebben om beelden te kunnen zoeken op basis van inhoudelijke kenmerken, zoals identificeerbare objecten (*Eiffeltoren*), generieke objecten (*stoel*), iconologische thema's (*Het Laatste Avondmaal*) en abstracte begrippen (*geluk*) (Vanstappen, 2019). Om deze ervaringen aan te bieden, beginnen musea meer en meer te experimenteren met beeldherkenning.

The Museum of Modern Art (MoMA) gebruikt AI-diensten van Google om historische foto's van afgelopen tentoonstellingen te koppelen aan de beelden uit de kunstcollectie die te zien zijn op die tentoonstellingszichten. Het CV-model analyseerde hiervoor alle foto's van tentoonstellingen. Wanneer het model een kunstwerk op de foto herkende, dan legde het een koppeling met het beeld van het kunstwerk. (MoMA, 2018).

Voor het Noorse Nasjonalmuseet werd het deep learning framework Caffe¹⁶ gebruikt om compositionele gelijkenissen te zoeken tussen de kunstwerken en om ze te classificeren op basis van Iconclass¹⁷. Dit resulteerde in een vernieuwde publiekstoegang waarbij kunstwerken op basis van gelijkenissen gevisualiseerd werden. Hoe meer gelijkenissen een kunstwerk heeft, hoe dichter de kunstwerken bij elkaar staan (Nasjonalmuseet, 2017).

¹⁶zie <http://caffe.berkeleyvision.org/>

¹⁷Iconclass is een classificatiesysteem voor kunst en iconografie, zie <http://www.iconclass.nl/home>

Daarnaast proberen een aantal musea via CV een oplossing te vinden voor de grote massa aan beelden die niet manueel beschreven raakt. Om zowel intern als extern gebruik van de beelden te stimuleren, zijn minstens een aantal trefwoorden nodig.

- Wellcome Collection heeft zelf een model gemaakt en getraind om beelden te categoriseren en gelijkaardige beelden te clusteren. Het wordt intern gebruikt om ongewenste beelden te verwijderen van de collectiewebsite en om de registratieworkflow te verbeteren. In de toekomst wil men het mogelijk maken dat bezoekers van de collectiewebsite gelijkaardige beelden vinden op basis van een beeld (Pim, 2018).
- Auckland Museum deed een test op tweeduizend collectiebeelden om via Microsoft Computer Vision de beelden automatisch te laten taggen en een basisrecord per beeld te creëren. Tags met een lagere waarschijnlijkheidsscore dan 60% werden verwijderd om te vermijden dat beschamende of misleidende records gepubliceerd worden (Moriarty, 2018).
- Het Britse webbedrijf CogApp gebruikte drie CV API's (Clarifai, Google Cloud Vision en Microsoft Computer Vision) om een visuele zoekmachine te maken waarmee bezoekers op basis van de inhoudelijke kenmerken van de beelden doorheen een collectie kunnen zoeken.¹⁸ Het deed de test met tweeduizend beelden van schilderijen van het Zweedse Nationalmuseum (Hindle, 2017).
- Sarjeant Gallery maakte zijn collectie op de nieuwe collectiewebsite doorzoekbaar via tags van een CV API.¹⁹ De collectie kan doorzocht worden op basis van kleur, beeldoriëntatie en tags. De tags werden verkregen via de Google Cloud Vision. Foutieve tags worden verborgen door de collectiemedewerkers. (Rowe, 2017)
- Ook The Swedish National Heritage Board heeft een webinterface gecreëerd waarbij de collectie doorzoekbaar gemaakt wordt met onder meer tags van Google Cloud Vision (Haskiya, 2019).

De resultaten met die experimenten waren goed. Soms zelfs beter dan verwacht. Sarjeant Gallery dacht de tags enkel te gebruiken voor intern gebruik. Collectiemedewerkers konden er sets van beelden rond een bepaald onderwerp mee maken. Men vond de tags echter zo goed dat besloten werd om ze ook op de website te publiceren. Doordat veel kunstwerken geen onderwerpbeschrijving hadden, konden de tags gebruikt worden om gerelateerde kunstwerken te vinden. (Rowe, 2017). Ook Moriarty (2018) concludeert dat CV API nuttig zijn om snel basisrecords te creëren voor beelden waarmee zowel interne als externe gebruikers zich een weg kunnen zoeken doorheen de collectie.

Er zijn wel nog vragen rond de controle op correctheid van de tags en hoe (en of) je aan de gebruiker moet laten weten dat de tags door een AI-systeem gecreëerd werden (Moriarty, 2018). Volgens Hindle (2017) zijn foute beschrijvingen een gevolg zijn van het trainen van de CV API met hedendaagse beelden, terwijl erfgoedbeelden historisch zijn.

In Vlaanderen wordt eveneens geëxperimenteerd met CV API. Vanstappen (2019) heeft voor MoMu onderzocht in welke mate off-the-shelf CV API zonder training bruikbaar zijn voor erfgoedcollecties. Het was een vergelijkend onderzoek waarbij verschillende modellen van Clarifai, Google Cloud Vision en Microsoft Computer Vision vergeleken werden op

¹⁸Zie <http://labs.cogapp.com/iiif-ml/>.

¹⁹Zie <https://collection.sarjeant.org.nz/explore>

een set van 164 beelden van objecten (inclusief close-ups en detailfoto's), catwalkbeelden, eventfoto's en scenografie. Vanstappen (2019) concludeerde dat de verkregen tags voor een dergelijk gespecialiseerd onderwerp geen hoge kwaliteit hadden, maar dat ze wel bruikbare resultaten gaven voor een globale beschrijving van de erfgoedcollecties. Het sterke punt is vooral de performantie. Een API is veel sneller en goedkoper dan een menselijke registrator. Bovendien geeft de beeldherkenningssoftware ook onderwerpen die buiten de traditionele collectiebeschrijving vallen, zoals kleur en sfeer.

3. Methodologie

In dit hoofdstuk wordt de methode toegelicht die gevuld werd voor het oplossen van de onderzoeksvragen. In samenspraak met Huis van Alijn werden use cases opgesteld die uitgetest werden met de CV API. Het museum leverde ook een dataset aan om te gebruiken voor dit onderzoek en waarmee twee workflows verschillende keren doorlopen werden: een voor het laten labelen van de fotocollectie door de CV API en een voor het trainen van een eigen model om foto's door de CV API te laten classificeren.

3.1 Opstellen use cases

Huis van Alijn heeft een grote fotocollectie over het dagelijkse leven in België in de 20^e en 21^e eeuw. Een deel hiervan zijn foto's die ontvangen werden via schenkingen of die gekocht werden op rommelmarkten, online of in kringloopwinkels. Die foto's komen daar na een overlijden of door een verhuis. Soms werd er op de foto's of dia's een jaartal, plaatsnaam of tekst geschreven die het de collectieregistrator mogelijk maakt om metadata toe te voegen aan de foto, maar vaak is er amper contextuele informatie en is er niet geweten wie of wat op de foto staat (Heemkunde Vlaanderen, 2011).

Om de foto's te kunnen ontsluiten of te doorzoeken, heeft het museum nood aan metadata die een idee geven van wat er op het beeld staat. Uit een gesprek met Astrid Vergauwe, digitaal strateeg van Huis van Alijn en Industriemuseum, bleek dat het museum de foto's indeelt in een aantal thema's (bv. *huwelijk, vakantie en speelgoed*) en in het decennium waarin de foto's getrokken worden (bv. 50s, 60s). Het zou voor het museum een enorme hulp zijn als via beeldherkenning de beelden voorzien worden van beschrijvende metadata en ingedeeld worden in de thema's en periodes.

	Geboorte	Huwelijk	Sint	Speelgoed	Vakantie	Totaal
00s	6	1	0	2	0	9
10s	10	4	0	2	0	16
20s	4	13	0	4	11	32
30s	2	20	0	10	22	54
40s	6	26	13	16	0	61
50s	15	82	82	14	16	209
60s	18	169	0	11	28	226
70s	21	55	2	19	17	114
80s	11	14	0	17	19	61
90s	3	9	0	6	3	21
onbekend	0	7	0	0	35	42
Totaal	96	400	97	101	151	845

Tabel 3.1: De opdeling van het aantal foto's per thema (kolommen) en per decennia (rijen).

Aan Huis van Alijn werd voorgesteld om volgende use cases te onderzoeken:

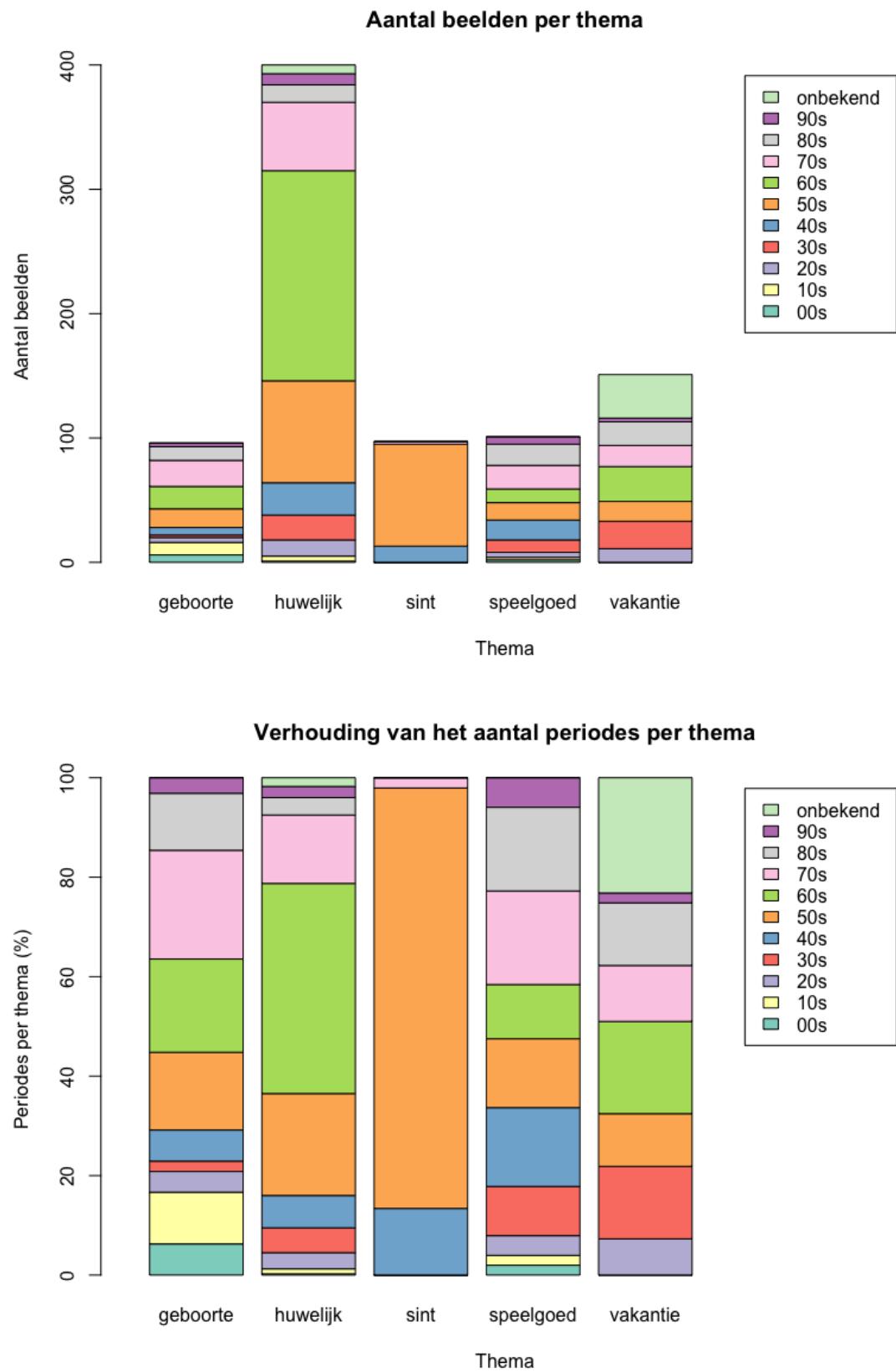
- het automatisch metadateren van iedere foto door het ingebouwde model van de CV API.
- het classificeren van de foto's in de thema's die door Huis van Alijn voorgesteld worden. Hiervoor wordt een *custom* model gecreëerd en wordt de CV API getraind. We gaan ook na of het mogelijk is om die classificatie te doen op basis van de tags uit het ingebouwde model van de CV API. Vermoedelijk zal het slagen hiervan afhangen van hoe vertrouwd het model met het thema is.
- het indelen van de foto's in de decennia waarin ze gemaakt werden. Ook hiervoor wordt een custom model ontwikkeld en wordt de CV API getraind.

3.2 Analyseren van de dataset

Het Huis van Alijn leverde een set van 845 beelden aan uit de *Anonieme snapshots* collectie die bestaat uit vijf thema's en tien decennia. Alle beelden hadden reeds metadata zodat de tags van de CV API vergeleken kon worden met de bestaande beschrijvingen. In tabel 3.1 en figuur 3.1 wordt de spreiding van de foto's per thema en periode getoond.

Het valt hierbij op dat de dataset ongelijk verdeeld is:

- bijna de helft van de dataset bestaat uit huwelijksfoto's. De andere thema's zijn meer gelijk verdeeld met grofweg honderd foto's per thema.
- daarenboven concentreren de huwelijksfoto's zich op drie periodes. Bijna 75% van de foto's zijn afkomstig uit de periode 50s, 60s en 70s.



Figuur 3.1: **Boven:** Het aantal beelden per thema met opsplitsing van de beelden per periode. Het aantal huwelijksfoto's springt er duidelijk uit. **Onder:** de verhouding van het aantal periodes per thema. Ook hier is het duidelijk dat dit ongelijk verdeeld is.

- de Sinterklaasfoto's zijn afkomstig uit slechts drie periodes (40s, 50s en 70s) waarvan bijna 85% uit één periode (50s).
- de ongelijke verdeling van de foto's per thema heeft ook een gevolg op de foto's per periode. Er zijn vooral foto's afkomstig uit de periodes 50s, 60s en 70s. Voor 50s gaan bijna 80% van de foto's over huwelijk en Sinterklaas, voor de 60s bijna 75% over huwelijk en voor de 70s zijn 50% huwelijksfoto's. Van de vroegste periodes (00s en 10s) zijn dan weer weinig foto's (zie ook figuur 3.3).
- zowel Sinterklaas- als vakantiefoto's komen niet in alle periodes voor.

Bij het interpreteren van de resultaten moet rekening gehouden worden met deze ongelijke verdeling.

3.3 Keuze voor API

Clarifai werd als API gekozen om de onderzoeksvervraag te beantwoorden.

Clarifai is gesticht in 2013 door Matthew Zeiler. Hij had in 2013 deelgenomen aan de ImageNet Challenge als doctoraatsstudent samen met zijn promotor Dr. Rob Fergus. Hun architectuur, ZFNet, waarop ook Clarifai gebouwd is, was de winnaar van de ImageNet Challenge (Tsang, 2018).

De keuze viel op deze API om een aantal redenen:

1. De website bevat goede en duidelijke documentatie om met de API aan de slag te gaan. De uitleg wordt telkens voorzien van een voorbeeld in de ondersteunde programmeertalen.
2. De API heeft clients in verschillende programmeertalen die het de ontwikkelaar mogelijk maken om de API te gebruiken in de programmeertaal naar keuze. JavaScript, Python, Java, C# en PHP hebben officiële clients die door Clarifai ondersteund en ontwikkeld worden. De REST API kan ook rechtstreeks gevraagd worden zonder gebruik van een API client (Clarifai, n.d.-a).
3. De Clarifai website beschikt over een grafische interface waarin gebruikers beelden kunnen opladen om ze te laten taggen door een van de ingebouwde modellen of een nieuw model kunnen trainen. Dit verhoogt het gebruiksgemak, wat, zoals besproken in 1.1, een belangrijke factor was bij de keuze voor een API.
4. Er zijn verschillende ingebouwde modellen die gebruikt kunnen worden om beelden te taggen, gaande van zeer algemene modellen (het general model) tot zeer specifieke modellen, zoals een model om modeconcepten te herkennen of een model om elementen met betrekking tot voeding en vaatwerk te detecteren.

3.4 Beelden beschikbaar maken via een publieke URL

Het Huis van Alijn bracht ons de beelden aan op een harde schijf. De beelden waren immers nog niet online beschikbaar. Om op een geautomatiseerde manier de API van

Clarifai te kunnen bevragen, moest een publieke URL voor de beelden voorzien worden. Het waren hoge resolutie beelden waarvan het merendeel in TIFF-formaat was en een kleiner deel in JPEG-formaat. De meeste beelden hadden daarom een bestandsgrootte (10 à 13 MB per beeld) die te groot is voor Clarifai. Clarifai wil immers beelden met een grootte van maximum 3,6 MB.

De beelden werden daarom op een beeldenserver geplaatst waarmee met een URL het beeld opgevraagd kon worden in een kleinere bestandsgrootte in het JPEG-bestandsformaat. Clarifai kan immers niet omgaan met beelden groter dan 3,6 MB. Een andere optie was om de beelden te converteren naar JPEG, ze daarbij te verkleinen naar een breedte van 512 pixels en ze vervolgens op een webserver of FTP-server te plaatsen. Clarifai verkleint immers zelf alle beelden tot 512 pixels en hoe dichter je bij dat formaat komt, hoe sneller de API calls verlopen (Clairbot, 2019). Dit laatste ontdekten we echter pas nadat de test al uitgevoerd was. Het was namelijk niet opgenomen in de documentatie.

De beelden werden opgeladen in een Amazon AWS EC2 Ubuntu instance waarop de beeldenserver (Cantaloupe) geïnstalleerd was. Op die manier kreeg ieder beeld een publieke URL die gebruikt kon worden bij het aanroepen van de Clarifai API.

3.5 Beschikbare metadata verzamelen in een CSV

De 845 beelden van Huis van Alijn waren voorzien van metadata die uit de bestandsnaam van de digitale bestanden gehaald konden worden. Er werd een CSV gecreëerd waarin alle beschikbare metadata per beeld verzameld werden:

- bestandsnaam of ID van het beeld;
- de extensie of het oorspronkelijke bestandsformaat van het beeld: .tif of .jpg;
- type: foto, diapositief of onbekend;
- thema;
- periode (decennium)

Op de harde schijf waren de afbeeldingen verzameld in een mapje (directory) per thema. Informatie zoals extensie, type, thema en periode werden uit de bestandsnaam gehaald.

Listing 3.1: een fragment van de CSV. De bestandsnaam heeft de logica *type-periode-nr* waaruit het type van het originele beeld (foto of dia) en de periode gehaald kan worden.

```
1 filename,extension,type,theme,period
2 DIA-80-00249.tif,diapositief,vakantie,80s
3 DIA-80-00358.tif,diapositief,huwelijk,80s
4 DIA-90-00009.tif,diapositief,vakantie,90s
5 FO-00-00015.tif,foto,geboorte,00s
6 FO-00-00034.tif,foto,huwelijk,00s
```

3.6 Uitvoeren van workflows

Er werden twee workflows een aantal keer doorlopen bij het uitvoeren van dit onderzoek:

- een workflow om de beelden te laten taggen door Clarifai;
- een workflow om via training een custom model te creëren.

3.6.1 Workflow 1: Een dataset laten taggen door Clarifai

De Clarifai API aanroepen

Clarifai beschikt over API-clients in verschillende programmeertalen (Java, Python, Javascript, C#, Objective-C en PHP), maar doordat we zelf het meeste voeling hebben met shell scripting, werd gekozen om de REST API te bevragen via cURL, een command-line programma voor het verkrijgen of verzenden van bestanden met behulp van een URL.

Een API key is nodig om de API te bevragen. Daarom werd eerst een account bij Clarifai aangemaakt. Automatisch krijg je dan een gratis key.

Het general model van Clarifai werd gebruikt om de volledige dataset van Huis van Alijn te taggen. Dit is namelijk het meest uitgebreide model met de meeste concepten. Het bevat ongeveer 11.000 verschillende concepten waaronder objecten, thema's, emoties in verschillende talen. Daardoor kan het model gebruikt worden voor een reeks van zeer uiteenlopende beelden (Clarifai, n.d.-b).

Om via de Clarifai API tags te krijgen voor een beeldbestand, moet commando 3.2 ingegeven worden in de command line. \${API_key} en \${url_naar_het_beeld} moeten vervangen worden door respectievelijk de API key en de URL van het beeld. In de URL moet je verwijzen naar de id van het model en de id van de versie van het model die je wenst te gebruiken. (Clarifai, n.d.-a). Voor het general model is dit aaa03c23b3724a16a56b629203edc62c en aa7f35c01e0642fda5cf400f543e7c40.

Listing 3.2: bash commando om een beeld door Clarifai te laten taggen.

```

1 #!/bin/bash
2 curl -X POST
3   -H 'Authorization: Key "${API_key}"'
4   -H 'Content-Type: application/json'
5   -d '{
6     "inputs": [
7       "data": {
8         "image": {
9           "url": "${url_naar_het_beeld}"
10          }
11        }
12      ]
13    }',
14  
```

```
15 "https://api.clarifai.com/v2/models/${model_id}/versions/
  ${version_id}/outputs"
```

Het bevragen van de API verliep vlot. Gemiddeld waren er per beeld twee à drie seconden nodig. Bij drie beelden mislukte het aanroepen omdat ze een te groot bestandsvolume hadden. Deze beelden werden verkleind waarna de API opnieuw aangeroepen werd.

Tags van API in een overzicht verzamelen

Na het bevragen van de Clarifai API werd per beeld een API response in JSON ontvangen met daarin onder meer de tags die aan de beelden gegeven werden.

Listing 3.3: een ingekorte versie van een ontvangen API response met voorspellingen van Clarifai .

```
1 {
2   "status": {
3     "code": 10000,
4     "description": "Ok",
5     "req_id": "7a70efb773234ba9a3eaf219328782b4"
6   },
7   "outputs": [
8     {
9       "id": "94b039d533454d53b01f2d3f825a097e",
10      "status": {
11        "code": 10000,
12        "description": "Ok"
13      },
14      "created_at": "2019-04-01T16:14:28.886109228Z",
15      "model": {
16        "id": "aaa03c23b3724a16a56b629203edc62c",
17        "name": "general",
18        "created_at": "2016-03-09T17:11:39.608845Z",
19        "app_id": "main",
20        "output_info": {
21          "message": "Show output_info with: GET /models/${model_id}/output_info",
22          "type": "concept",
23          "type_ext": "concept"
24        },
25        "model_version": {
26          "id": "aa7f35c01e0642fda5cf400f543e7c40",
27          "created_at": "2018-03-06T21:10:24.454834Z",
28          "status": {
29            "code": 21100,
30            "description": "Model trained successfully"
31          }
32        }
33      }
34    }
35 }
```

```
32     "display_name": "General"
33   },
34   "input": {
35     "id": "d74a5f9d8961406eb5a15e9172be9399",
36     "data": {
37       "image": {
38         "url": "http://ec2-18-191-252-182.us-east-2.
39           compute.amazonaws.com:8182/iiif/2/F0
40           -30-00197/full/full/0/default.jpg"
41     }
42   },
43   "data": [
44     {
45       "concepts": [
46         {
47           "id": "ai_18TKp2h5",
48           "name": "people",
49           "value": 0.999308,
50           "app_id": "main"
51         },
52         {
53           "id": "ai_bmls4LpL",
54           "name": "group",
55           "value": 0.9847269,
56           "app_id": "main"
57         },
58         {
59           "id": "ai_sd6DKdXp",
60           "name": "group together",
61           "value": 0.9833066,
62           "app_id": "main"
63         },
64         {
65           "id": "ai_RQccV41p",
66           "name": "woman",
67           "value": 0.9795268,
68           "app_id": "main"
69         },
70         {
71           "id": "ai_VPmHr5bm",
72           "name": "adult",
73           "value": 0.97893196,
74           "app_id": "main"
75         }
76       ]
77     }
78   }
79 }
```

De gegevens werden verzameld in een CSV-bestand die geïmporteerd werd in Google Sheets. Google Sheets heeft als voordeel dat resultaten eenvoudig gedeeld kunnen worden met de museummedewerkers en bezit bovendien over ingebouwde reken- en queryfuncties.

die het analyseren van de resultaten vereenvoudigen. Bovendien kan Google Sheets afbeeldingen in een cel tonen via de IMAGE-functie²⁰. Dit maakt het veel eenvoudiger om de gegevens te valideren.

Gegevens valideren

Om de kwaliteit van de tags te controleren, werd dezelfde methodologie gebruikt als in Vanstappen (2019). Bij ieder tag werd een checkbox geplaatst die door de validator aangevinkt moest worden indien de tag correct was. Dit had als voordeel dat de validatie relatief snel kon gebeuren en dat ook het verwerken van de resultaten eenvoudig was.

Tijdens het valideren bleek het aanvinken van correcte tags niet altijd eenvoudig te zijn. Sommige tags vereisten enige interpretatie die moeilijk is zonder de context van het beeld te kennen. Het gaat om tags zoals *happy*, *music* en *love*. Daarvoor konden we ons baseren op de thema's van de foto's. We gingen ervan uit dat mensen op huwelijksfoto's en geboortefoto's vol liefde waren, en dat zij, net als mensen op vakantie, gelukkig zijn.

Een andere moeilijkheid was het herkennen van het geslacht van jonge kinderen op (oudere) foto's, of weten wanneer de mensen op de foto beschouwen konden worden als bejaard. We zijn bij het corrigeren van deze tags niet te streng geweest. Bij twijfel, wanneer bijvoorbeeld zowel de tags *boy* als *girl* voorkwamen, hebben we steeds de term met de grootste waarschijnlijkheidsscore goedgekeurd.

Ten slotte waren er ook nog de tags *group*, *several* en *many*. We beschouwden een groep als een groep als er drie of meer mensen op de foto stonden. De andere tags werden gevoelsmatig goed- of afgekeurd.

3.6.2 Workflow 2: Een custom model trainen

Voor het trainen van de API door de creatie van een custom model werd de workflow gevuld van Clarifai (n.d.-a). Voor dit onderzoek werden twee custom models gemaakt:

- een model die de beelden van Huis van Alijn kan classificeren volgens hun thema (geboorte, huwelijk, Sinterklaas, speelgoed en vakantie), vanaf nu het themamodel;
- een model die de beelden van Huis van Alijn kan classificeren volgens het decennium waarin de foto getrokken was (1900s t.e.m. 1990s), vanaf nu het periodemodel.

Voor ieder model werden een aantal stappen doorlopen:

1. het maken van een trainingsset en validatieset;
2. het opladen van de trainingsbeelden naar Clarifai;
3. het creëren van een custom model met de concepten uit de trainingsbeelden;
4. het trainen van het model;
5. het valideren van het model met de validatieset.

²⁰Voor meer info over de IMAGE-functie in Google Sheets, zie: <https://support.google.com/docs/answer/3093333?hl=en>

beeld	URL	type	theme	period	tag 1	tag 2	tag 3	tag 4	tag 5
	http://ec2-18-191-252-182.us-east-2.amazonaws.com/bachelorsproof/2003-036-018	onbekend	huwelijk	onbekend	people	<input checked="" type="checkbox"/> portrait	<input checked="" type="checkbox"/> man	<input checked="" type="checkbox"/> adult	<input checked="" type="checkbox"/> lid
	http://ec2-18-191-252-182.us-east-2.amazonaws.com/bachelorsproof/2012-029	onbekend	huwelijk	onbekend	groom	<input checked="" type="checkbox"/> wedding	<input checked="" type="checkbox"/> bride	<input checked="" type="checkbox"/> people	<input checked="" type="checkbox"/> woman
	http://ec2-18-191-252-182.us-east-2.amazonaws.com/bachelorsproof/2012-030-001	onbekend	huwelijk	onbekend	people	<input checked="" type="checkbox"/> group	<input checked="" type="checkbox"/> adult	<input checked="" type="checkbox"/> administration	<input type="checkbox"/> woman
	http://ec2-18-191-252-182.us-east-2.amazonaws.com/bachelorsproof/2012-030-001								<input checked="" type="checkbox"/>

Figuur 3.2: Validatie van het ingebouwd model in Google Sheets.

De getrainde modellen werd vervolgens gebruikt om de hele dataset van Huis van Alijn te classificeren in respectievelijk thema en periode.

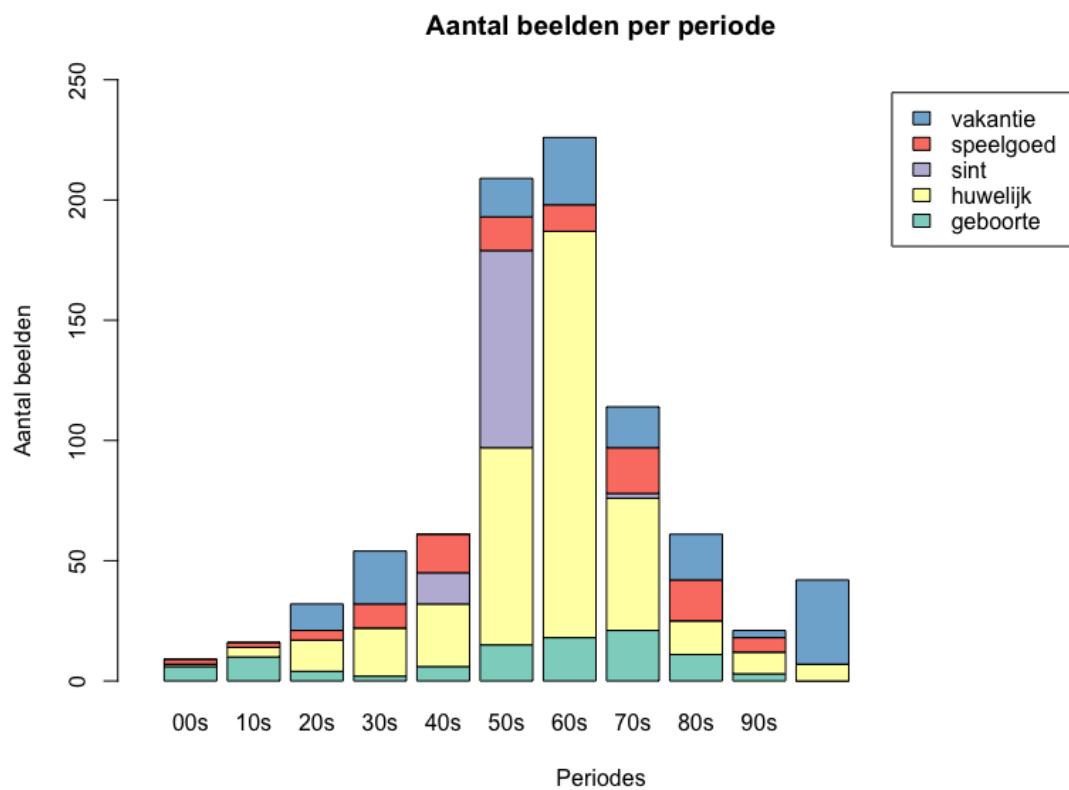
Het is in Clarifai mogelijk om een custom model te maken via de API of via een webinterface. Voor dit onderzoek hebben we gebruik gemaakt van de API.

Training- en validatieset maken

Uit de dataset van Huis van Alijn werd een trainingset en een validatieset gemaakt. De trainingsbeelden dienden om het model te trainen; de validatiebeelden om de verschillende iteraties van het model te valideren en te beoordelen welke iteratie het meest performant is (Lievens, 2017; Brownlee, 2017). Door de kleine dataset (845 beelden) wilden we slechts 50% van de beelden gebruiken voor training; 20% werd gebruikt als validatieset.

Trainingsbeelden en validatiebeelden werden willekeurig geselecteerd uit de dataset. Gong (2017) meldt dat het erg belangrijk is dat de trainingsbeelden overeenkomen met de beelden waarop het model toegepast zal worden. Er werd daarom gezorgd dat, in de mate van het mogelijke, in de beelden voor respectievelijk het themamodel en het periodemodel alle decennia en thema's (gelijkmataig) aanwezig waren. Omdat een huwelijk in de jaren 1900 er nu eenmaal anders uitziet dan dat in de jaren 1950 of 1990 is het immers belangrijk dat het model zoveel mogelijk beelden te zien krijgt van hoe een huwelijk eruit kan zien doorheen alle periodes. Het periodemodel moet dan weer weten dat een foto uit een bepaald decennium kan bestaan uit zowel huwelijks-, geboorte-, Sinterklaas-, speelgoed- als vakantiefoto's.

Voor het themamodel wilden we ieder thema evenwaardig trainen en dus evenveel trai-



Figuur 3.3: Staafdiagram die het aantal beelden per periode weergeeft. De meest rechtse staaf toont het aantal beelden waarvan de periode niet gekend is.

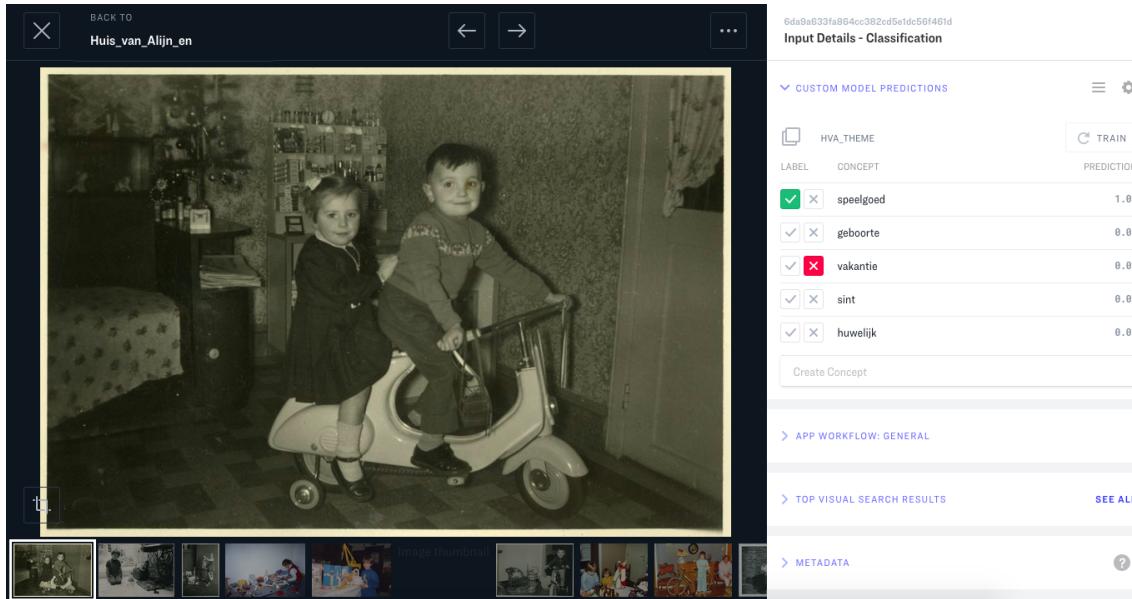
ningsbeelden per thema gebruiken. Doordat het aantal beelden per thema ongelijk verdeeld is (vierhonderd huwelijksfoto's, ± honderd beelden voor de andere thema's, zie tabel 3.1), beslisten we om het aantal foto's per thema te beperken tot honderd. Dit hield in dat voor ieder thema slechts vijftig foto's gebruikt zouden worden als trainingsbeelden en dat de validatieset uit twintig foto's per thema bestaat.

Bij het periodemodel was het niet mogelijk om iedere periode evenwaardig te trainen. Het aantal beelden per decennium was hiervoor te ongelijk verdeeld (zie figuur 3.3). Er waren bijvoorbeeld minder dan dertig foto's voor de jaren 00, 10 en 90, en meer dan tweehonderd foto's voor de jaren 50 en 60. Beperken van het aantal foto's per periode was daardoor onmogelijk. De trainingset zou dan te marginaal zijn.

Trainingsbeelden opladen naar de computer vision service

Om een model te maken, moeten eerst trainingsbeelden opgeladen worden die reeds het concept bevatten waarvan we willen dat het model het herkent (Clarifai, n.d.-a).

We beslisten om het model in iteraties te trainen. Clarifai (n.d.-a) stelt immers voor om te starten met tien beelden per concept en vervolgens meer beelden per concept toe te voegen indien dit nodig blijkt. Na iedere iteratie werd het model gevalideerd. Voor het themamodel



Figuur 3.4: Trainingsbeelden kunnen via de webinterface van Clarifai opgeladen en voorzien worden van een of meerdere concepten.

werd gewerkt in iteraties van tien trainingsfoto's per thema. Voor het periodemodel werd dit opgetrokken tot iteraties van maximum dertig trainingsfoto's per periode²¹, omdat het periodemodel een grotere variëteit aan beelden moet herkennen.

Listing 3.4: bash commando om een beeld met een concept naar Clarifai op te laden.

```

1 #!/bin/bash
2 curl -X POST \
3   -H 'Authorization: Key "${API_key}"' \
4   -H 'Content-Type: application/json' \
5   -d '{
6     "inputs": [
7       "data": {
8         "image": {
9           "url": "${url_naar_het_beeld}"
10          },
11         "concepts": [
12           "id": "${concept}",
13           "value": true
14         ]
15       }
16     ]
17   }'
18 https://api.clarifai.com/v2/inputs

```

²¹Waar mogelijk, voor de decennia 00s, 10s, 20s, 30s en 90s waren geen dertig trainingsbeelden beschikbaar.

Listing 3.5: Het antwoord van de Computer Vision API in JSON na het opladen van een beeld met een concept

```

1  {
2      "status": {
3          "code": 10000,
4          "description": "Ok",
5          "req_id": "f8db86f0ff564086a6611c7dd34e91e3"
6      },
7      "inputs": [
8          {
9              "id": "153005f7c8b04d949356dc36c1bbbee9",
10             "data": {
11                 "image": {
12                     "url": "http://ec2-18-191-252-182.us-east-2.
13                         compute.amazonaws.com:8182/iif
14                         /2003-036-018/full/922,/0/default.jpg"
15                 },
16                 "concepts": [
17                     {
18                         "id": "huwelijk",
19                         "name": "huwelijk",
20                         "value": 1,
21                         "app_id": "1b9b01460b944923a9eb1578668d4ea6"
22                     }
23                 },
24                 "created_at": "2019-07-13T13:57:15.569447471Z",
25                 "modified_at": "2019-07-13T13:57:16.399324215Z",
26                 "status": {
27                     "code": 30000,
28                     "description": "Download complete"
29                 }
30             }
31         }
32     }
33 }
```

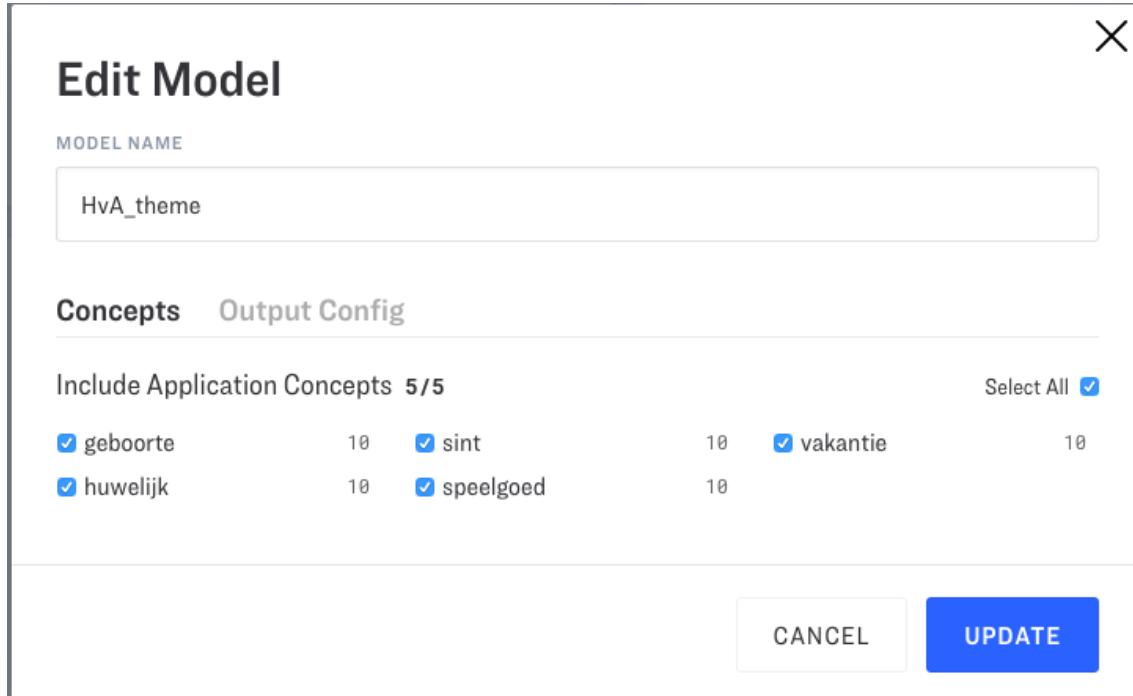
Het is in Clarifai ook mogelijk om negatieve concepten toe te voegen. Dit zijn concepten die niet op de foto te zien zijn. Op een huwelijksfoto kan bijvoorbeeld aangeduid worden dat dit geen speelgoed-, vakantie-, geboorte- en Sinterklaasfoto is. Om te testen of negatieve concepten een invloed hebben op de resultaten van de validatieset, werd na de eerste iteratie voor alle trainingsbeelden alle negatieve concepten toegevoegd.

Het model maken

Nadat de beelden van de eerste iteratie opgeladen werden naar Clarifai, kan het het model aangemaakt worden. Aan het model moet een naam worden gegeven en de concepten toegevoegd die op de trainingsbeelden te zien zijn.

Daarnaast kan het model nog geconfigureerd worden. In onze modellen werd ingesteld dat

een beeld slechts één concept bevat²² (parameter `concepts_mutually_exclusive`) en dat er enkel trainingsbeelden gebruikt worden waarop een concept te zien is (parameter `closed_environment`)



Figuur 3.5: Een custom model kan via de webinterface van Clarifai gecreëerd en aangepast worden.

Listing 3.6: bash commando om een custom model met vijf concepten te creëren

```

1 #!/bin/bash
2 curl -X POST \
3   -H 'Authorization: Key "${API_key}"' \
4   -H 'Content-Type: application/json' \
5   -d '{
6     "model": {
7       "id": "HvA_theme",
8       "output_info": {
9         "data": {
10           "concepts": [
11             {"id": "'${concept1}'"},
12             {"id": "'${concept2}'"},
13             {"id": "'${concept3}'"},
14             {"id": "'${concept4}'"},
15             {"id": "'${concept5}'"}]
```

²²Een huwelijksfoto kan bijvoorbeeld nooit een vakantiefoto zijn.

```

19         "id": "${concept5}",
20     }]
21   },
22   "output_config": {
23     "concepts_mutually_exclusive": true,
24     "closed_environment": true
25   }
26 }
27 }
28 } \
29 https://api.clarifai.com/v2/models

```

Listing 3.7: antwoord van de API na het creëren van het model

```

1 {
2   "status": {
3     "code": 10000,
4     "description": "Ok",
5     "req_id": "66be8c1a7bbc4cca9398d02c7c15a097"
6   },
7   "model": {
8     "id": "HvA_theme",
9     "name": "HvA_theme",
10    "created_at": "2019-07-13T14:15:49.966900528Z",
11    "app_id": "1b9b01460b944923a9eb1578668d4ea6",
12    "output_info": {
13      "output_config": {
14        "concepts_mutually_exclusive": true,
15        "closed_environment": true,
16        "max_concepts": 0,
17        "min_value": 0,
18        "test_split_percent": 10
19      },
20      "message": "Show output_info with: GET /models/{",
21        "model_id}/output_info",
22      "type": "concept",
23      "type_ext": "concept"
24    },
25    "model_version": {
26      "id": "68b9e3c20ae04abbbdc2e3fb974b9a0f",
27      "created_at": "2019-07-13T14:15:49.994520471Z",
28      "status": {
29        "code": 21102,
30        "description": "Model not yet trained"
31      },
32      "active_concept_count": 5
33    }
34 }

```

```
33 }
34 }
```

Het model trainen

Telkens er nieuwe trainingsbeelden toegevoegd zijn, moet het model getraind worden. Door het trainen van het model gaat het systeem naar alle beelden met de concepten kijken en daarvan leren (Clarifai, n.d.-a).

Listing 3.8: bash commando om het model te trainen

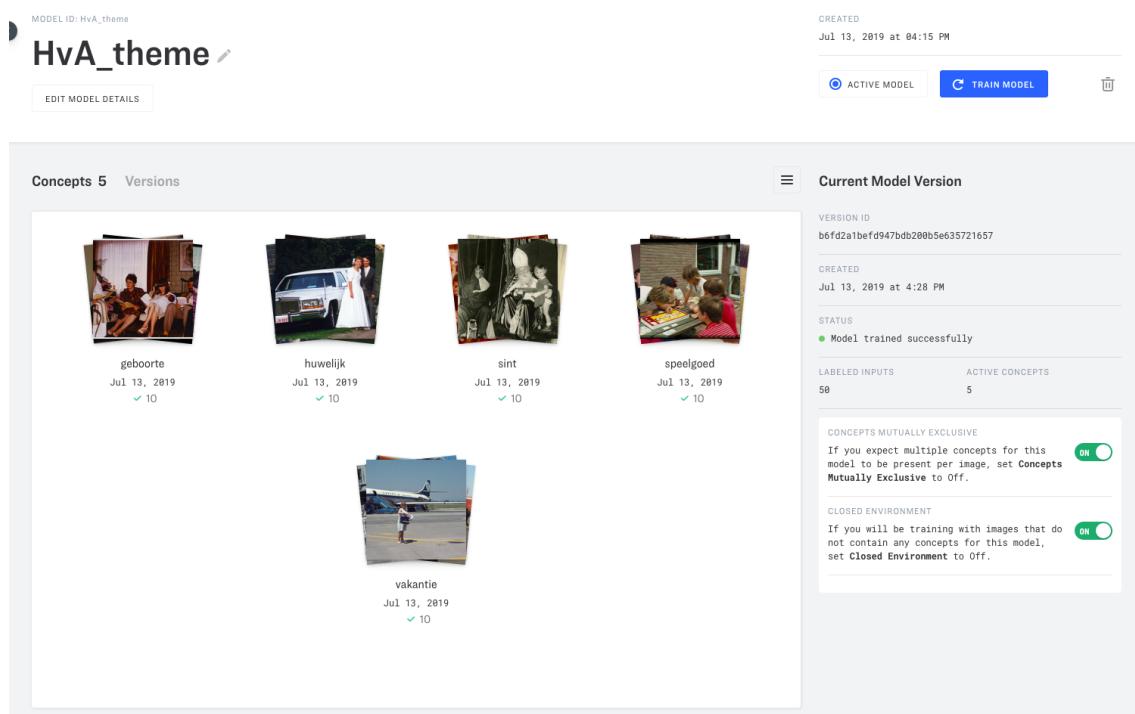
```
1 #!/bin/bash
2 curl -X POST \
3   -H 'Authorization: Key "${API_key}"' \
4   -H 'Content-Type: application/json' \
5   "https://api.clarifai.com/v2/models/${model_id}/
       versions"
```

Listing 3.9: antwoord van de API nadat het model getraind werd

```
1 {
2   "status": {
3     "code": 10000,
4     "description": "Ok",
5     "req_id": "2f513295f4074ec39115e3e5e500d652"
6   },
7   "model": {
8     "id": "HvA_theme",
9     "name": "HvA_theme",
10    "created_at": "2019-07-13T14:15:49.966900Z",
11    "app_id": "1b9b01460b944923a9eb1578668d4ea6",
12    "output_info": {
13      "output_config": {
14        "concepts_mutually_exclusive": true,
15        "closed_environment": true,
16        "max_concepts": 0,
17        "min_value": 0,
18        "test_split_percent": 10
19      },
20      "message": "Show output_info with: GET /models/{
21        model_id}/output_info",
22      "type": "concept",
23      "type_ext": "concept"
24    },
25    "model_version": {
26      "id": "b6fd2a1befd947bdb200b5e635721657",
27      "created_at": "2019-07-13T14:28:25.060240688Z",
28    }
29  }
30 }
```

```

27     "status": {
28         "code": 21103,
29         "description": "Custom model is currently in
30             queue for training, waiting on inputs to
31             process."
32     },
33 }
34 }
```



Figuur 3.6: Het model kan via de webinterface van Clarifai getraind worden.

Na het trainen krijgt het model een versiecode. Met deze code is het mogelijk om bij het voorspellen een bepaalde versie van het model te gebruiken. Als blijkt dat het model na iteratie 3 beter is dan dat van iteratie 4, dan kan je met de versiecode het model van iteratie 3 gebruiken.

Het model valideren

Nadat het model getraind was, werd het uitgetest op de validatieset. Hiervoor werd workflow 1 uit 3.6.1 doorlopen met de validatiesets voor de twee modellen.

Om het model te beoordelen werd gebruik gemaakt van de gemiddelde rappel, precisie en F_1 -score (Lievens, 2017; Koehrsen, 2018):

A	D	E	F	G	H	I	J	K
beeld	origineel	iteratie 1	iteratie 1 (negatieveven)	iteratie 2	iteratie 3	iteratie 4	iteratie 5	iteratie 5 (negatieveven)
	speelgoed	speelgoed	speelgoed	speelgoed	speelgoed	speelgoed	speelgoed	speelgoed
	speelgoed	speelgoed	speelgoed	huwelijk	huwelijk	speelgoed	huwelijk	huwelijk
	vakantie	vakantie	vakantie	vakantie	vakantie	vakantie	vakantie	vakantie
	vakantie	huwelijk	huwelijk	huwelijk	huwelijk	huwelijk	huwelijk	huwelijk

Figuur 3.7: Validatie van het custom model in Google Sheets.

- De rappel is het percentage van de positieve trainingsbeelden die ook effectief als positief werd gelabeld door het model (verhouding tussen het aantal true positives en het totaal aantal positieve voorbeelden van het concept);
- De precisie is het percentage van de beelden die als een bepaald concept herkend werden daadwerkelijk dat concept bevatten (verhouding tussen het aantal true positives en het totaal aantal herkende beelden voor dat concept);
- De precisie en de rappel kunnen gecombineerd worden in één enkele score, de F_1 -score. Om een grote F_1 -score te behalen moeten zowel de precisie als de rappel het goed doen.

Hoe groter deze scores, hoe performanter het model. Clarifai geeft meerdere concepten aan een model. Er werd enkel rekening gehouden met het concept dat de hoogste waarschijnlijkheidsscore gekregen heeft. Hierbij werd geen drempelwaarde ingesteld. Dit betekent dat als aan een beeld een correct concept toegekend werd met een waarschijnlijkheid van 10%, dat deze ook als een true positive bestempeld werd.

De hoogste F_1 -score voor het Themamodel werd behaald in de vierde iteratie met een score van 92% (tabel 3.2). In de vijfde iteratie daalt de score licht tot 91%. Na de eerste iteratie (tien trainingsbeelden voor ieder thema) werd er al een F_1 -score van 78% behaald. Het toevoegen van veertig negatieve voorbeelden na de eerste iteratie zorgde voor een lichte stijging tot 80%. Als de vierde iteratie van naderbij bekijken wordt (tabel 3.3), dan valt het op dat het model erg goed is in het classificeren van de Sinterklaasfoto's (score van 100%). Het minst scoorde op de speelgoed- en geboortefoto's met een F_1 -score van respectievelijk 84% en 85%. Bij de speelgoefoto's werd een lagere rappel behaald (80%), maar was de precisie hoger (89%).

Bij het Periodemodel, met de onevenwichtige en beperkte trainingset, werd op de valida-

	aantal positieven	aantal negatieven	F_1 -score (%)
iteratie 1/A	10	0	78
iteratie 1/B	10	40	80
iteratie 2	20	40	86
iteratie 3	30	40	88
iteratie 4	40	40	92
iteratie 5	50	40	91

Tabel 3.2: De F_1 -scores voor de verschillende iteraties van het Themamodel met vermelding van het aantal postieve en negatieve voorbeelden per thema na iedere iteratie.

	true positives	false negatives	false positives	rappel	precisie	F_1 -score
geboorte	17	3	3	85%	85%	85%
huwelijk	20	0	2	100%	91%	95%
Sint	20	0	0	100%	100%	100%
speelgoed	16	4	2	80%	88,9%	84,2%
vakantie	19	1	1	95%	95%	95%
gemiddelde	18,4	1,6	1,6	92%	92%	91,9%

Tabel 3.3: De resultaten (rappel, precisie en F_1 -score) van de vierde iteratie bij het trainen van het Themamodel.

	aantal positieven	aantal negatieven	F_1 -score (%)
iteratie 1/A	21,5	0	39,9
iteratie 1/B	21,5	193,5	39,9
iteratie 2	30,2	193,5	48
iteratie 3	36,2	193,5	47,3
iteratie 4	39,9	193,5	43,9

Tabel 3.4: De F_1 -scores voor de verschillende iteraties van het Periodemodel met vermelding van het gemiddeld aantal postieve en negatieve voorbeelden per thema na iedere iteratie. Zie figuur om een beeld te hebben van de volledige spreiding van de trainingset

	true positives	false negatives	false positives	rappel	precisie	F_1 -score
00s	0	2	0	0%	0%	0%
10s	1	2	0	33%	100%	50%
20s	4	3	5	57%	44,4%	50%
30s	5	5	12	50%	29,4%	37%
40s	8	4	13	67%	38,1%	48,5%
50s	15	23	15	39%	50%	44%
60s	13	21	13	48%	59,4%	52,8%
70s	12	9	12	55%	47,8%	51,2%
80s	5	5	5	58%	58,3%	58,3%
90s	2	3	2	25%	33,3%	28,6%
gemiddelde	7,1	7,7	7,7	43%	46,1%	42,1%

Tabel 3.5: De resultaten (rappel, precisie en F_1 -score) van de tweede iteratie bij het trainen van het Periodemodel.

tieset slechtst een maximale F_1 -score van 48% behaald na iteratie 2. Het valt op dat vanaf iteratie 3 de score weer daalt. Een mogelijke verklaring is dat het aantal trainingsbeelden per periode te verschillend is. De rappel en de precisie stijgen voor een beperkt aantal periodes, maar dalen voor alle andere periodes. Het toevoegen van negatieve voorbeelden na iteratie 1 brengt geen verandering in de rappel, precisie en F_1 -score. Iteratie 1 en iteratie 1 met negatieven hebben beiden een F-score van 40%. Iteratie 2 behaalde de grootste rappel voor beelden uit de jaren 40 (67%) en de grootste precisie op beelden uit de jaren 10 (100%). De grootste globale score werd behaald op beelden uit de jaren 80 met een F-score van 58%. Het model scoorde het slechtst op foto's uit het eerste decennium de jaren 1900 en de jaren 90, met een F_1 -score van respectievelijk 0% en 29%. Dit zijn periodes met weinig trainingsbeelden (respectievelijk vier en tien). Het is opmerkelijk dat het model bij foto's van de jaren 10 toch een F_1 -score van 50% behaalt, terwijl deze periode ook maar acht trainingsbeelden heeft.

Wat doet de scores dalen na een hogere iteratie? Bij het Periodemodel vermoedden we dat dit komt door het onevenwicht in het aantal trainingsbeelden tussen de verschillende periodes, maar voor het Themamodel is dit moeilijker te bepalen omdat de thema's even goed getraind zijn. Misschien komt het door een licht onevenwicht in het aantal periodes in de trainingsbeelden van het themamodel? Bij iteratie 4 liggen de aantallen iets dichter op elkaar (met een uitschieter voor de jaren 50), terwijl in iteratie 5 de uitschieter van de jaren 50 nog groter geworden is. Ook bij het periodemodel zien we de verhouding tussen het aantal thema's in de foto's groter worden na iteratie 2. Dit moet verder onderzocht worden met grotere datasets om hier conclusies uit te trekken. Een andere mogelijkheid is dat we op de grenzen van de CV API gestoten zijn.

4. Resultaten met het ingebouwd model

De 845 beelden van Huis van Alijn werden getagged door het general model van Clarifai. Aan iedere foto werden twintig tags gegeven. De waarschijnlijkheidsscore van de tags bevond zich tussen 100% en 65% en had een gemiddelde van 93%.

De verkregen tags zijn op te delen in een aantal categorieën en gaan over verschillende aspecten van het beeld:

- entiteiten die op de beelden te zien zijn, waaronder *volwassene, speelgoed, bloem, kameel, jurk, meubels* en *strand*;
- activiteiten die uitgevoerd worden op de beelden, zoals *dansen, zitten, rusten, shoppen* en *reizen*;
- emoties of gevoelens die te zien zijn op het beeld of die opgeroepen worden door het beeld, denk aan *plezier, liefde, genot, avontuur, affectie, sexy* en *mooi*;
- maar ook contextuele concepten, zoals *vriendschap, familie, toerisme, vrije tijd* en *het samen zijn*;
- en hoeveelheden: *één, twee, drie, vier, veel, enkele*. Dit waren vaak weinigzeggende termen die opvallend vaak fout waren. Het was ook niet altijd duidelijk waar die hoeveelheden op sloegen. Veel mensen? Veel bloemen?
- tot slot kregen we ook informatie over de foto zelf: *portret, kleur, zwart-wit, sepia, monochroom...*

Tijdens de validatie werd vastgesteld dat Clarifai veel algemene termen geeft, zoals *mensen, volwassene, kind, groep* en *veel*, maar soms waren we ook verwonderd over de relevantie van de tags. Enkele van die relevante tags zijn: *schildersezel, roos, uitstapje, airbus, souvenir, bazaar, grasland, broche* en *bruidsmeisje*.

Na validatie bleken 11.517 van de 16.900 tags correct te zijn (68,15%). Dit resulteerde in



Figuur 4.1: Drie foto's uit de fotocollectie van Huis van Aijn waarvan alle tags correct waren en de foto waar het minst aantal tags correct waren (rechtsonder).

371 unieke termen die de foto's beschreven. Bij 21 foto's waren alle twintig tags correct. 97 beelden (11,5%) hadden minder dan tien correcte tags, waarvan de slechts scorende een foto is van het thema speelgoed. Deze foto had slechts zes juiste tags. Gemiddeld hadden de foto's 13,6 correcte tags.

Vooraf hadden we verwacht dat het model niet goed zou scoren op de Sinterklaasfoto's. Sinterklaas is immers een ritueel dat enkel voorkomt in de lage landen en enkele vroegere koloniën van Nederland. De kans lijkt dus klein dat het model de concepten van Sinterklaas kent. Omdat modellen voornamelijk getraind worden met hedendaagse foto's, vreesden we ook dat de resultaten met de oudere foto's minder goed zouden zijn.

Na validatie bleek het vermoeden rond de Sinterklaasfoto's correct te zijn (zie tabel 4.1). Het valt meteen op dat dit thema het laagste gemiddelde en het hoogst aantal foto's met minder dan tien correcte tags heeft. Aan de beelden van dit thema werden tevens het minst aantal unieke termen gegeven. Het model blijkt ook minder goed op het speelgoedthema te scoren. De resultaten voor de huwelijksfoto's zijn dan weer erg goed. Het gemiddelde voor dit thema is ongeveer 1,5 punten beter dan het gemiddelde van alle beelden.

Wat betreft de periodes is het moeilijker om conclusies te trekken. Uit tabel 4.2 kan afgeleid worden dat de recentere foto's (1960-1999) beter scoren en dat de oudste foto's

	\bar{x} (/20)	max	min	aantal unieke termen	aantal < 10
Geboorte	13,69	20	7	112	7
Huwelijk	15,16	20	8	158	7
Sint	10,03	15	7	51	48
Speelgoed	11,28	17	6	126	23
Vakantie	13,42	20	9	212	12
Alle foto's	13,63	20	6	371	97

Tabel 4.1: een overzicht van het gemiddelde (\bar{x}), maximale en minimale aantal juiste tags van de twintig tags van Clarifai voor de foto's per thema, het aantal unieke termen dat gegeven werd aan ieder thema en het aantal foto's waarvan minder dan tien tags correct waren.

het slechtst scoren. Niettemin moet hier nuance van gemaakt worden. Er zijn immers slechts negen beelden uit de jaren 1900. Bovendien bestaan foto's uit de jaren 50 - en in mindere mate uit de jaren 40 - voor een groot deel uit foto's over Sinterklaas, wat hun slechtere score kan verklaren. Foto's uit de jaren 60 scoren opmerkelijk goed, maar dit valt dan weer te verklaren door de grote aanwezigheid van huwelijksfoto's in die periode (meer dan 75%).²³

4.1 Gedetailleerde resultaten per thema

In dit deel wordt dieper ingegaan op de resultaten per thema. Voor ieder thema wordt de top 30 van meest voorkomende termen, het maximale, minimale en gemiddeld aantal correcte tags, het aantal beelden met minder dan tien juiste tags en het aantal unieke termen toegelicht.

4.1.1 Geboorte

In totaal waren er 96 beelden voor dit thema. De API scoorde goed op deze foto's, in lijn met het totale gemiddelde. Er werden 112 unieke termen gebruikt om de beelden te beschrijven. Gemiddeld waren 13,7 tags correct. Het maximale aantal correcte tags was twintig, het minste zeven. Slechts zeven beelden hadden minder dan tien correcte tags (7,2%).

De dertig meest voorkomende termen verwijzen naar mensen, kinderen en nakomelingen. Het model ziet ook enkele emoties in de beelden: liefde en affectie. Het merendeel van de foto's waren ook portretten van moeder (en soms ook vader) met baby. Soms stonden er ook andere mensen, zoals broers, zussen, familieleden en vrienden op de foto.

Uit de top 30 valt op dat er een groot verschil is tussen het aantal verschijningen van

²³zie 3.2

	\bar{x} (/20)	max	min	aantal < 10
00s	11,44	14	9	1
10s	12,69	16	8	2
20s	12,91	19	9	3
30s	13,26	20	9	3
40s	12,16	20	7	19
50s	12,33	20	6	48
60s	15,01	20	7	7
70s	14,56	20	7	4
80s	14,26	20	7	3
90s	14,48	20	9	1
Onbekend	12,76	20	9	6
Alle foto's	13,63	20	6	97

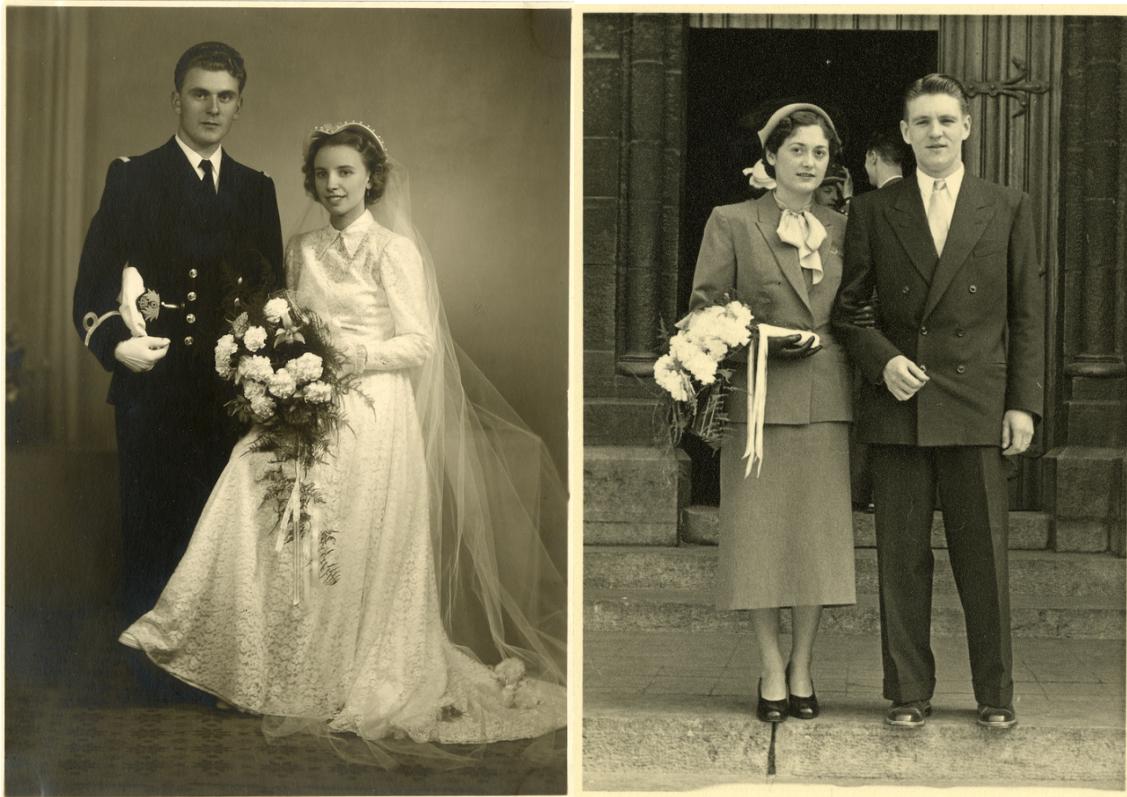
Tabel 4.2: een overzicht van het gemiddelde (\bar{x}), maximale (max) en minimale (min) aantal juiste tags per periode en het aantal foto's waarvan minder dan tien tags (of de helft) correct waren.

mensen (93)	zoon (48)	gezichtsexpressie (22)
volwassene (83)	groep (35)	meisje (19)
kind (82)	twee (32)	affectie (17)
vrouw (76)	sepia (29)	drie (16)
portret (75)	binnenshuis (27)	verschillende (15)
familie (65)	kamer (27)	broer of zus (13)
kledij (62)	meubels (25)	bed (10)
nageslacht (51)	liefde (25)	jongen (8)
man (49)	samenkommen (24)	voertuig (7)
baby (48)	monochroom (23)	stoel (7)

Tabel 4.3: De dertig meest voorkomende tags van geboortefoto's met het ingebouwde Clarifai-model



Figuur 4.2: een best scorende (boven) en slechtst scorende (onder) foto van het thema geboorte uit de fotocollectie van het Huis van Alijn.



Figuur 4.3: een best scorende (links) en slechtst scorende (rechts) foto van het thema huwelijk uit de fotocollectie van het Huis van Alijn.

de meest voorkomende term (93 keer) en die op plaats 30 (7 keer). De top 11 komt in minstens de helft van de beelden voor. Het valt hierbij ook op dat term baby slechts aan de helft van de foto's gegeven is.

Een opvallende vaststelling was dat het model de term *dochter* niet kent; enkel *zoon* werd gegeven. Doordat het voor ons niet mogelijk was om het verschil tussen een mannelijke en een vrouwelijke baby te zien, hebben we dit altijd als correct beschouwd.

Het model vergiste zich soms ook. Man en vrouw met kind werden in zestien van de gevallen door het model als een trouwfoto beschouwd. Ook de tags bruidegom (3 keer) en bruid kwamen voor (2 keer). Dit is vermoedelijk doordat op die foto's de baby gewikkeld was in een sluier, wat verkeerdelijk door het model als een bruidssluiер gepercipieerd werd.

4.1.2 Huwelijk

Voor dit thema waren vierhonderd beelden beschikbaar. Het general model van Clarifai scoorde erg goed op dit thema. Er werden 158 unieke termen gebruikt om de beelden te beschrijven. Gemiddeld waren 15,2 tags correct, wat ruim boven het gemiddelde is van alle beelden. Het maximale aantal correcte tags was twintig, het minste acht. Slechts zeven beelden hadden minder dan tien correcte tags, wat maar 1,75% van het totaal aantal trouwfoto's is!

mensen (399)	sluier (198)	verschillende (83)
vrouw (387)	groep (160)	smoking (75)
volwassene (377)	jurk (159)	huwelijk (72)
man (354)	ceremonie (147)	bruids (63)
bruiloft (341)	gezichtsexpressie (132)	veel (58)
portret (316)	samenkommen (127)	verbintenis (57)
kledij (311)	liefde (121)	meisje (56)
bruidegom (255)	monochroom (117)	koppel (55)
twee (251)	bloemstuk (101)	kind (53)
bruid (239)	familie (99)	boeket (48)

Tabel 4.4: De dertig meest voorkomende tags van trouwfoto's met het ingebouwde Clarifai-model

De dertig meest voorkomende termen verwijzen naar termen die je verwacht bij trouwfoto's: mensen, een koppel, vrouw, man, bruid, bruidegom, boeket, op een bruiloft, in smoking en (trouw)jurk. Ook in deze foto's wordt liefde gezien.

De top tien van de termen komt in minstens de helft van de beelden voor. We stellen hier tevens een verschil vast tussen de meest (399 keer) en minst (48 keer) voorkomende term. Het valt op dat de term bruiloft 341 keer voorkomt; die van huwelijk 72 keer.

Uit analyse van de termen kan geconcludeerd worden dat het ingebouwde model van de Clarifai API trouwfoto's kan onderscheiden zonder dat verdere training nodig is.

4.1.3 Sinterklaas

Zoals verwacht scoorde het Clarifai model niet goed op het taggen van de foto's over Sinterklaas. Voor dit thema waren 97 beelden beschikbaar. Er werden slechts 51 unieke termen gebruikt om de beelden te beschrijven. Dit is 70 termen minder dan geboorte, dat na Sinterklaas de minst unieke termen heeft. Gemiddeld waren slechts tien tags correct, wat ruim onder het gemiddelde is van alle beelden (13,6). Het maximale aantal correcte tags was vijftien, het minste zeven. 48 beelden hadden minder dan tien correcte tags, wat 49,5% van het totaal aantal Sinterklaasfoto's is.

De reden voor deze lage cijfers komt zeer waarschijnlijk doordat het model de concepten van Sinterklaas niet kent. Het model gaf zeer regelmatig de tag *retro* bij deze beelden, maar die werd steeds als incorrect aangevinkt. Alle beelden van Huis van Alijn zijn immers retro als het met hedendaagse ogen bekeken wordt. Daarnaast werden ook de termen *zoon*, *nakomelingen* en *familie* als incorrect gevalideerd als er geen ouder of volwassene op de foto staat.

De dertig meest voorkomende tags verwijzen naar een portret van twee, drie of vier mensen met volwassene(n) en kind(eren). De 97 beelden waren ook portretten van een of meerdere kinderen, soms vergezeld van volwassenen, met Sinterklaas. Wat opvalt is dat de tags sterk de nadruk op kleding leggen: kledij, sluier, outfit, kostuum, toga, bovenkleding.



Figuur 4.4: een best scorende (links) en slechtst scorende (rechts) foto van het thema Sinterklaas uit de fotocollectie van het Huis van Alijn.

mensen (97)	drie (24)	broer of zus (10)
kledij (97)	gezichtsexpressie (21)	vier (8)
volwassene (95)	samenkommen (20)	bovenkleding (7)
portret (91)	groep (19)	recreatie (7)
kind (90)	kostuum (15)	acteur (6)
man (86)	verschillende (15)	sepia (6)
twee (53)	jongen (14)	uniform (6)
monochroom (30)	vrouw (14)	meisje (5)
outfit (30)	familie (13)	toga (5)
sluier (25)	jas (10)	zitten (5)

Tabel 4.5: De dertig meest voorkomende tags van Sinterklaasfoto's met het ingebouwde Clarifai-model

Ook bij deze beelden is er een groot verschil tussen de meest voorkomende termen en de minst voorkomende termen. Mensen en kledij komen in alle foto's voor, terwijl meisje, toga en zitten voor slechts vijf foto's gebruikt werden. De top 7 van de tags werd aan minstens de helft van de beelden gegeven.

4.1.4 Speelgoed

Er waren 101 foto's voor dit thema. Net als bij Sinterklaas presteert Clarifai voor deze foto's onder het gemiddelde. Gemiddeld waren 11,3 tags correct en hadden 23 beelden minder dan tien correcte tags (22,8%). De slechts scorende foto had 6 juiste tags; voor de best scorende foto's waren dit 17 tags. Voor het beschrijven van deze foto's werden 126 unieke termen gebruikt.

De dertig meest voorkomende termen verwijzen vooral naar kinderen, slechts in mindere



Figuur 4.5: een best scorende (links) en slechtst scorende (rechts) foto van het thema speelgoed uit de fotocollectie van het Huis van Alijn.

mensen (100)	recreatie (29)	speelgoed (18)
kind (96)	twee (29)	broer of zus (17)
kledij (65)	binnenshuis (28)	straat (17)
portret (51)	familie (25)	voertuig (17)
kamer (48)	groep (25)	thuis (16)
jongen (44)	volwassene (22)	vrouw (14)
monochroom (43)	zitten (22)	sepia (12)
meisje (42)	samenkommen (20)	zitplaats (11)
meubels (38)	gezichtsexpressie (18)	stoel (10)
één (34)	plezier (18)	man (9)

Tabel 4.6: De dertig meest voorkomende tags van de speelgoedfoto's met het ingebouwde Clarifai-model

mate naar speelgoed zelf. Het valt bijvoorbeeld op dat het afgebeelde speelgoed vaak niet herkend (of getagged) werd. Wel wordt er plezier en recreatie herkend. De term speelgoed (18 keer) komt ook voor, maar wel in mindere mate, net als vrije tijd (5 keer), spelen (5 keer) en spel (4 keer). Het ingebouwde model is bijgevolg niet voldoende om het merendeel van de speelgoedfoto's eruit te halen.

De grote verschillen tussen de meest (100 keer) en minst (9 keer) voorkomende termen vallen op. Enkel de top 4 van de termen wordt aan minstens de helft van de beelden gegeven. Dit komt vermoedelijk omdat deze foto's een minder vast format hadden dan de foto's van de vorige thema's. Die foto's waren vaak portretten van resp. moeders/ouders met hun pasgeboren kind, bruid en bruidegom op hun trouw en kinderen met Sinterklaas. De speelgoedfoto's hadden een grotere variëteit aan composities: portretten van kinderen die met hun speelgoed poseerden, maar ook kinderen en/of volwassenen die gezamenlijk een spel speelden.

mensen (142)	straat (47)	meisje (23)
volwassene (115)	voertuig (46)	weg (23)
vrouw (98)	portret (35)	landschap (21)
man (83)	veel (32)	water (21)
recreatie (81)	strand (30)	vrije tijd (20)
reizen (79)	monochroom (30)	vakantie (19)
kledij (77)	buitenhuis (30)	familie (18)
groep (75)	verschillende (30)	plezier (18)
kind (50)	stad (26)	kust (18)
samenkommen (48)	twee (25)	gebouw (17)

Tabel 4.7: De dertig meest voorkomende tags van de vakantiefoto's met het ingebouwde Clarifai-model

Naar de reden voor de mindere score voor deze beelden is het gissen. Net zoals bij Sinterklaas werden vaak de termen *retro* of *vintage* gegeven, wat als fout beschouwd werd. Idem voor de termen *zoon*, *nakomelingen* of *familie*, wanneer er geen ouder op de foto staat. Ook vergiste het model zich regelmatig. Kinderen werden vaak als volwassene getagged of kregen een fout gender (bv. een jongen die als meisje gezien werd). Poppen werden dan weer vaak als baby gelabeld. Het valt op dat de tags zich hoofdzakelijk op mensen concentreren. Mogelijk is het model minder goed in objectherkenning, maar dit moet verder onderzocht worden.

4.1.5 Vakantie

Het ingebouwde model van Clarifai scoorde goed op de 151 vakantiefoto's, in lijn met het algemene gemiddelde. Er werden 212 unieke termen gebruikt om de beelden te beschrijven. Gemiddeld waren 13,4 tags correct. Het maximale aantal correcte tags was twintig, het minste negen. Slechts twaalf beelden hadden minder dan tien correcte tags (8%).

De dertig meest voorkomende termen verwijzen naar mensen, vrije tijd, onderweg zijn en het strand. Het model ziet ook plezier in de beelden. Eveneens wordt het thema vakantie in negentien foto's herkend, net zoals toerist (13 keer) en toerisme (8 keer).

Ook deze foto's hebben geen vast format. Het zijn vakantiekiekjes. Dit verklaart de grote verscheidenheid aan termen om de beelden te beschrijven. De top 7 van de termen komt in minstens de helft van de beelden voor. Zoals bij de andere thema's is er een groot verschil tussen het aantal verschijningen van de meest voorkomende term (142 keer) en de minst voorkomende term (17 keer).

4.2 Vergelijking met de beschrijvingen van registratoren

De tags werden vergeleken met de beschrijvingen van de foto's door de menselijke registratoren. Hiervoor hadden we een export uit het collectiebeschrijvingsysteem van het



Figuur 4.6: een best scorende (boven) en slechtst scorende (onder) foto van het thema vakantie uit de fotocollectie van het Huis van Alijn.

museum gekregen van de 496 foto's van de thema's Sinterklaas en Huwelijk.

Beschrijvingen bestonden uit:

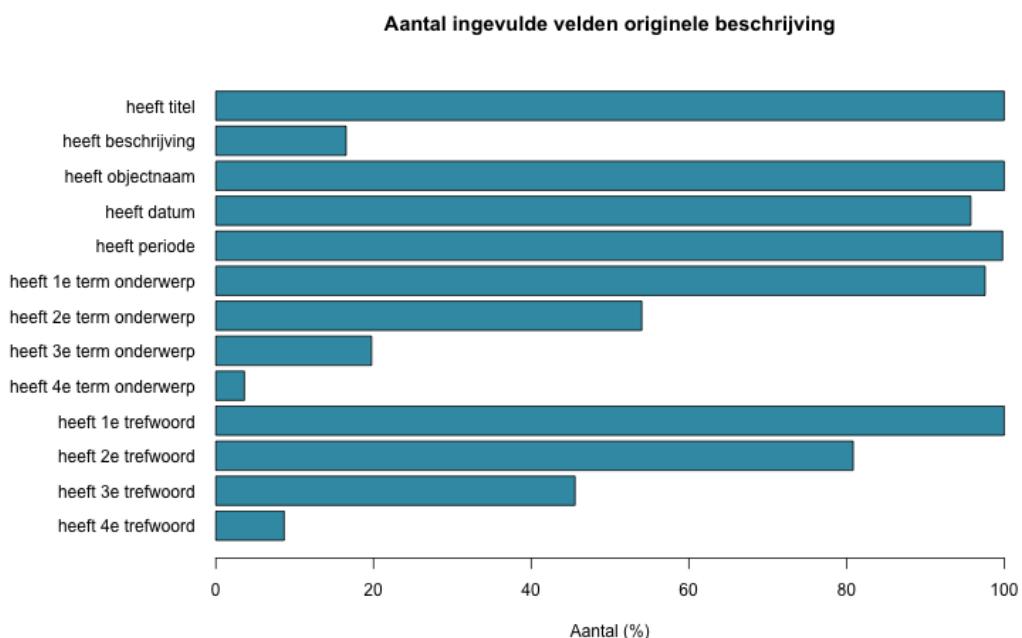
- Titel: de titel bestond meestal uit het afgebeelde onderwerp en personen, de locatie en de datum, bv. *Huwelijk Juliette en Werner, Gent, 1962* of *Meisje bij Sinterklaas, Grand Bazar Gent, 1949*.
- Beschrijving: geeft wat extra informatie over wat op de foto geportretteerd wordt. Meestal wordt er meer informatie gegeven over de locatie op de foto, zoals *Feestvierders komen aan in feestzaal de Raadskelder in Gent* of *De foto werd genomen in de Grand Bazar in de Gentse Veldstraat*, maar soms wordt er ook meer duiding gegeven over de kledij of de personen op de foto.
- Objectnaam: Dit is steeds *foto* of *dia* voor de beelden in deze case.
- Creatiedatum van het object, steeds in de vorm jaar-maand-dag.
- Periode: het decennium waarin de foto getrokken werd.
- Afgebeelde onderwerp: dit zijn termen die beschrijven wat op de foto afgebeeld is zoals *huwelijk* of *Sinterklaas*.
- Trefwoord: deze trefwoorden worden gebruikt om de collectie te doorzoeken.

Het valt op dat er voor de verschillende beelden slechts een beperkt aantal velden ingevuld werden. Alle beelden hebben een titel, objectnaam en minstens een trefwoord (zie figuur 4.7). De andere velden werden minder ingevuld. Het veld *beschrijving* werd zelfs maar voor 88 beelden ingevuld. Gebrek aan tijd en personeel zijn hier de oorzaak van, zoals reeds aangegeven in Hoofdstuk 1.1. Huis Van Alyn pakt de registratie projectmatig aan. Wanneer er een tentoonstelling plaatsvindt, wordt de registratie van items over dat onderwerp aangepakt. Het museum heeft daarom thema's waarvan de beelden heel precies beschreven zijn, maar ook thema's waarbij dit minder het geval is.

Titel, beschrijving, periode en datum kunnen niet door Clarifai getagged worden, maar het is wel interessant om te kijken of het in staat is om afgebeeld onderwerp en trefwoord aan te duiden. De registratoren gebruikten 107 unieke termen voor deze velden. Dat is minder dan het aantal unieke termen dat door Clarifai gebruikt werd. Clarifai gaf gemiddeld minstens zes correcte tags aan ieder beeld (gemiddeld 13,6). Het valt uit figuur 4.7 op dat de registratoren aanzienlijk minder termen geven aan de foto's: nog geen 50% van de beelden heeft drie trefwoorden en ongeveer 50% slechts twee afgebeeld onderwerpen. Vooral de Sinterklaasfoto's zijn erg beperkt beschreven. Bijna alle beelden van dit thema hebben slechts een trefwoord en afgebeeld onderwerp, namelijk de term *Sinterklaas*.

Elf van de dertig meest voorkomende termen (zie tabel 4.8) zijn plaatsnamen, waarvan Clarifai is niet in staat om deze termen te genereren. Andere termen verwijzen vooral naar entiteiten die op het beeld te zien zijn (bruidspaar, park), informatie over de foto (studioportret en groepsfoto) en activiteiten (dans). Emoties of gevoelens, die Clarifai wel geeft, ontbreken. Het is opvallend dat een registrator slechts tien keer de term *bruidegom* en zestien keer de term *bruid* gebruikt heeft. Clarifai was in staat om die termen in dezelfde foto's respectievelijk 255 keer en 239 keer te herkennen.

De termen zijn doorgaans specifieker dan de termen van Clarifai (bruidspaar, bruidsboeket,



Figuur 4.7: Een staafdiagram die aanduidt hoeveel velden (percentage) ingevuld werden door de registratoren van huis van Alijn. Het gaat enkel om de foto's van de thema's Sinterklaas en Huwelijk.

huiskamer (zie tabel 4.4)), maar het lijkt niet onmogelijk om op basis van de Clarifai-tags deze termen te verkrijgen. Clarifai geeft namelijk wel de termen *huwelijk* en *bloemstuk* of *bloemen* waarvan de term *bruidsboeket* afgeleid kan worden of de termen *bruid* en *bruidegom* waarvan de term *bruidspaar* gemaakt kan worden.

Uit deze analyse concluderen we dat een ingebouwd model van een CV API in staat is om een registrator te ondersteunen bij het werk. Clarifai is sneller en vollediger en kan zo de foto's voorzien van meer termen. Het dient verder onderzocht te worden of de service gebruikt kan worden om termen te genereren uit de termenlijst van Huis van Alijn.

4.3 Toepassen in de praktijk?

Het loslaten van het ingebouwde model van Clarifai op de 845 beelden van Huis van Alijn leverde 11.517 juiste (68,15%) en 5.383 foute (31,85%) tags op. Dit gebeurde in ongeveer 35 minuten. Om dit in een productieomgeving te gebruiken, zijn een aantal strategieën mogelijk om het aantal fouten te verkleinen.

huwelijk (400)	vervoer (34)	Merelbeke (12)
bruidspaar (351)	dans (32)	Zomergem (12)
Sinterklaas (192)	interieur (28)	Gentbrugge (11)
bruidsboeket (189)	park (20)	behang (11)
Gent (165)	huiskamer (17)	bruidegom (10)
feest (94)	bruid (16)	Sint-Amandsberg (9)
studioportret (57)	Sint-Martens-Latem (15)	taart (9)
kerk (54)	Zwijnaarde (15)	Aalst (8)
auto (36)	groepsportret (15)	Beervelde (8)
bloem (plant) (34)	Veldstraat (Gent) (14)	Loppem (8)

Tabel 4.8: De dertig meest voorkomende termen gebruikt door registratoren van Huis van Aijn voor Sinterklaas- en huwelijksfoto's. De term *Sinterklaas* komt 192 keer voor omdat het zowel als afgebeeld concept als als trefwoord gebruikt wordt.

drempelwaarde (%)	aandeel juiste tags (%)	aantal foute tags (%)
95	55	17
90	85	25
85	94,8	29,6
geen	100	31,85

Tabel 4.9: Verhouding tussen drempelwaarde, totaal aantal van de juiste tags en het aantal foute tags

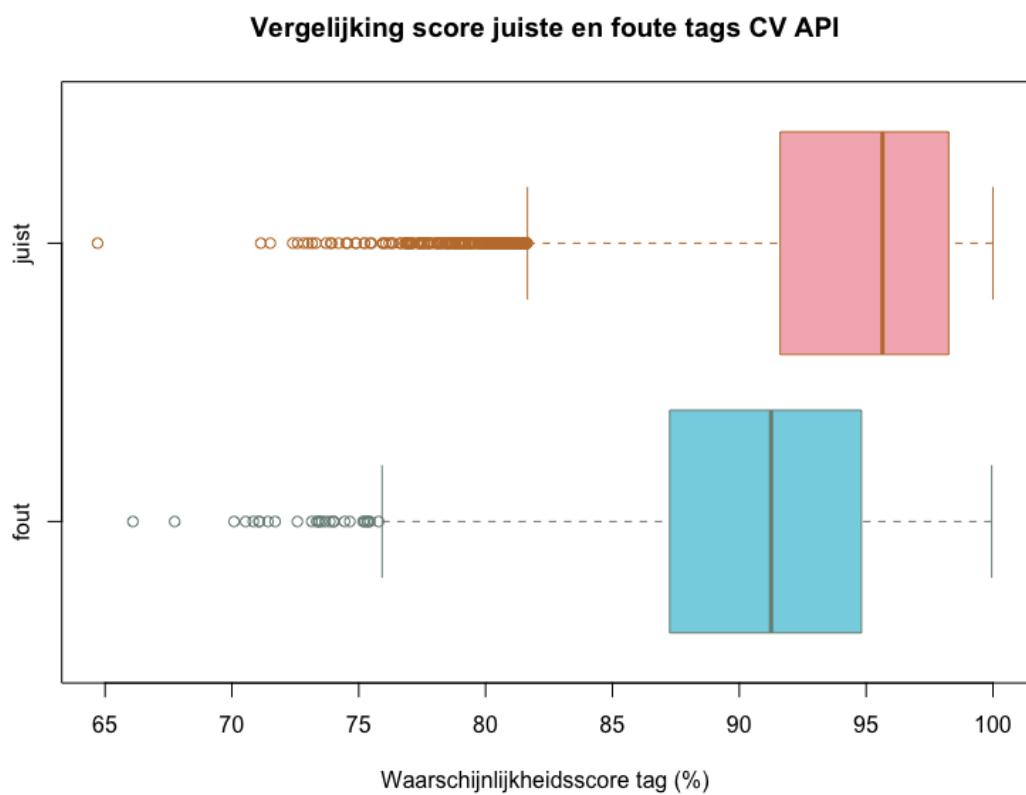
4.3.1 Instellen van een drempelwaarde

Er kan een drempelwaarde ingesteld worden om het aantal foute tags te verkleinen. Bij het instellen van een drempelwaarde worden alleen maar tags aanvaard waarvan de waarschijnlijkheidsscore groter is dan de drempelwaarde. Bij het uitvoeren van de case werd geen drempelwaarde ingesteld. De hoogste waarschijnlijkheidsscore van een correcte tag was 100%, de laagste 64,7%.

Indien het museum de drempelwaarde instelt op 90%, krijgt het 82% van het totaal aantal juiste tags. 25% van de verkregen tags zijn dan fout. De drempelwaarde kan nog hoger gesteld worden, op 95%, maar dan krijgt het slechts 55% van het totaal aantal juiste tags. Slechts 17% van de verkregen tags zijn dan verkeerd. De drempelwaarde kan verlaagd worden naar 85% om een groter aandeel van de juiste tags te hebben, maar dan zijn 29,6% van het aantal tags fout. Nu is dit aantal 31,85% zonder drempelwaarde (tabel 4.9).

4.3.2 Selectie maken van thema's

Een andere strategie om fouten te verkleinen is om een selectie te maken van de foto's die door de CV API gelabeld moeten worden. Het museum zou ervoor kunnen kiezen om thema's waarmee het ingebouwde model niet vertrouwd is, niet te laten labelen, zoals



Figuur 4.8: Twee boxplots die de waarschijnlijkheidsscores vergelijken van tags van Clarifai die als juist en fout beoordeeld worden. Er is een overlap merkbaar tussen beide boxplots waardoor het niet mogelijk is om het grootste deel van de juiste tags te behouden en de fouten te vermijden bij het instellen van een drempelwaarde.

bijvoorbeeld de Sinterklaasfoto's. Deze foto's doen immers het gemiddelde dalen en het aantal foute tags stijgen. Een nadeel aan deze strategie is dat de foto's al op voorhand ingedeeld moeten zijn in thema's.

5. Resultaten met de custom modellen

In ongeveer een kwartier tijd hebben zowel het periodemodel als het themamodel de 845 beelden van Huis van Alijn ingedeeld periodes of thema's. In die 845 beelden zaten ook de beelden die gebruikt werden voor training en validatie. Het was opmerkelijk om vast te stellen dat ook deze beelden soms door de modellen fout geklassificeerd werden. Clarifai gaf meerdere tags per foto. Omdat een foto slechts ingedeeld kan worden in één thema, hielden we enkel rekening met de tag die de hoogste waarschijnlijkheidsscore gekregen heeft.

In dit hoofdstuk worden afzonderlijk de resultaten van beide modellen besproken. Tot slot zal net als in hoofdstuk 4 suggesties gedaan worden voor de toepassing in de praktijk.

5.1 Themamodel

Het Themamodel heeft 845 beelden geanalyseerd en geklassificeerd. 753 beelden (89%) werden correct geklassificeerd. De waarschijnlijkheidsscore van de tags bevond zich tussen 100% en 39% en had een gemiddelde van 94,6%. De gemiddelde rappel voor de vijf thema's is 90%, terwijl de gemiddelde precisie een waarde van 86% heeft.

Het custom model kon de foto's van het thema Sinterklaas het beste indelen (F_1 -score van 97%). Dit thema haalde in hoofdstuk 4 nog de slechtste resultaten, maar heeft bij het Themamodel de grootste precisie en de tweede grootste rappel. Dit is te verklaren doordat de foto's sterk op elkaar lijken. Op iedere foto staat een oudere man met baard en mijter, vergezeld door een of meerdere kinderen die met de oude man op de foto poseren. Opvallend is ook de sterke score van de vakantiefoto's (F_1 -score van 94%), terwijl deze foto's juist gekenmerkt worden door een grote variëteit (strandfoto's, kampeerfoto's,

	true positives	false negatives	false positives	rappel	precisie	F_1 -score
geboorte	81	15	50	84,4%	61,8%	71,4%
huwelijk	346	54	6	86,5%	98,3%	92%
Sint	93	4	1	95,9%	98,9%	97,4%
speelgoed	88	13	21	87,1%	80,7%	83,8%
vakantie	145	6	14	96%	91,2%	93,6%
gemiddelde	150,6	18,4	18,4	90%	86,2%	87,6%

Tabel 5.1: De performantie (rappel, precisie en F_1 -score) van het Themamodel op de volledige dataset van Huis van Alijn.

citytrips, wintersport,...).

Het model scoorde het minst op geboortefoto's. Vooral de precisie voor deze foto's zijn ondermaats (62%). Er waren vijftig false positives²⁴. Veel huwelijksfoto's (54 false negatives²⁵) werden immers als geboortefoto herkend. Net als in vorig hoofdstuk waren de resultaten met de speelgoefoto's niet zo goed. Zowel de precisie als de rappel zijn lager dan het gemiddelde.

Lagere scores zijn soms te wijten aan de vaag afgebakende thema's door het museum. Sommige foto's die door het museum ingedeeld werden in het thema Speelgoed zijn foto's van kinderen met een stuk speelgoed op vakantie, terwijl er ook vakantiefoto's zijn waar kinderen met speelgoed op poseren. Dit verklaart waarom het model zeven speelgoefoto's classificeert als vakantiefoto's en vier vakantiefoto's als speelgoefoto. Het model had het ook moeilijk om poppen van baby's te onderscheiden. Negen geboortefoto's werden door het model ingedeeld in het thema Speelgoed, terwijl zes speelgoefoto's als geboortefoto beschouwd werden (zie tabel 5.2).

Zoals in hoofdstuk 4 werd geanalyseerd of het model slechter scoort op foto's uit vroegere periodes. Dit lijkt niet het geval te zijn. Foto's uit de jaren 60 (en 50) hebben een grotere foutenmarge, maar dit valt te verklaren door de grote hoeveelheid huwelijksfoto's uit deze periode.

5.2 Periodemodel

Ook het Periodemodel heeft 845 beelden geanalyseerd en geklassificeerd. Omdat van 42 beelden de periode niet gekend was door het museum en we dus niet konden nagaan of het model ze juist geklassificeerd heeft, werden deze beelden niet meegerekend bij de beoordeling van de resultaten.

²⁴Dit zijn vijftig foto's die foutief het label geboorte kregen.

²⁵Dit zijn 54 huwelijksfoto's die niet als huwelijksfoto herkend werden.



Figuur 5.1: Voorbeeld van foto's van de thema's Vakantie en Speelgoed die gelijkend zijn. De bovenste foto is een foto die door Huis van Aijn ingedeeld werd in het thema Speelgoed, de onderste foto is een vakantiefoto.

		Voorspeld thema				
		geboorte	huwelijk	sint	speelgoed	vakantie
Feitelijk thema	geboorte	81	3	1	9	2
	huwelijk	43	346	0	6	5
	sint	1	1	93	2	0
	speelgoed	6	0	0	88	7
	vakantie	0	2	0	4	145

Tabel 5.2: Confusion matrix voor het Themamodel. Het Themamodel slaagt erin om meestal het juiste thema eruit te halen.

	true positives	false negatives	false positives	rappel	precisie	F_1 -score
00s	3	6	0	33,3%	100%	50%
10s	13	3	8	81,3%	61,9%	70,3%
20s	14	18	15	43,8%	48,3%	45,9%
30s	34	20	45	63%	43%	51,1%
40s	40	21	70	65,6%	36,4%	46,8%
50s	110	99	71	52,6%	60,8%	56,4%
60s	120	106	56	53,1%	68,2%	59,7%
70s	73	41	51	64%	58,9%	61,3%
80s	43	18	24	70,5%	64,2%	67,2%
90s	10	11	3	47,6%	77%	58,8%
gemiddelde	46	34,3	34,3	57,5%	61,9%	56,7%

Tabel 5.3: De performantie (rappel, precisie en F_1 -score) van het Periodemodel op de volledige dataset van Huis van Alijn.

Het model had 460 beelden (57%) correct geklassificeerd. De waarschijnlijkheidsscore van de tags bevond zich tussen 94% en 20% en had een gemiddelde van slechts 55%. Dit kan betekenen dat het model ofwel te weinig trainingsbeelden had om de verschillende concepten te kunnen onderscheiden, of dat de taak te moeilijk is voor een CV API. De gemiddelde rappel voor de tien periodes is maar 57%, terwijl de gemiddelde precisie een waarde van 62% heeft.

De hoogste F_1 -score wordt behaald bij foto's van de jaren 10 (70%). Dit is nochtans een van de periodes met het laagst aantal trainingsbeelden (8 foto's). Vooral de rappel voor deze periode is erg goed (81%). Ook de 80s scoren goed met een F_1 -score van 67%. Over het algemeen scoren de periodes vanaf de jaren 60 het best, met uitzondering van de jaren 10. Het is gissen waarom. Misschien zijn zwart-wit foto's uit de jaren 20 tot en met 60 vergelijkbaar qua kleur?

	Voorspelde periode									
	00s	10s	20s	30s	40s	50s	60s	70s	80s	90s
Feitelijke periode	00s	3	2	1	0	3	0	0	0	0
	10s	0	13	1	1	1	0	0	0	0
	20s	0	2	14	12	0	3	0	0	0
	30s	0	0	4	34	9	4	3	0	0
	40s	0	1	2	4	40	13	0	0	0
	50s	0	3	3	15	43	110	32	3	0
	60s	0	0	3	8	13	42	120	31	9
	70s	0	0	1	3	0	9	19	73	8
	80s	0	0	0	0	0	0	2	14	43
	90s	0	0	0	1	0	0	0	3	7

Tabel 5.4: Confusion matrix voor het Periodemodel. Het Periodemodel slaagt erin om voor de meeste periodes overwegend het juiste periode te voorspellen, met uitzondering van drie periodes.

Twee periodes behalen een lage rappel, maar een erg hoge precisie. Foto's uit de jaren 00 hebben een rappel van slechts 33%, maar een precisie van 100%; voor de jaren 90 gaat het om een rappel van 48% en een precisie van 77%. Deze periodes scoorden tevens minder goed in de validatieset en behoren tot de groep met het minst aantal trainingsbeelden (respectievelijk vier en tien beelden).

De slechtste scores werden behaald met foto's uit de jaren 20 en 40, met een respectievelijke F_1 -score van 46% en 47%. Bij de jaren 40 is de rappel goed (66%), maar is de precisie erg laag (36%). Het heeft namelijk zeventig false positives. Dit komt doordat veel foto's uit de jaren 50 (43 foto's), en in mindere mate jaren 60 (13 foto's) en 30 (9 foto's), herkend wordt als een foto uit de jaren 40. Als het model fout is, dan vergist het zich meestal met de omliggende periodes (zie tabel 5.4).

5.3 Toepassen in de praktijk?

Het Periodemodel beschouwen we niet voldoende performant om in de praktijk toe te passen. De foutenpercentage is te hoog en de waarschijnlijkheidsscores zijn te laag om goede conclusies te kunnen trekken. In deze sectie wordt daarom enkel gefocust op het Themamodel.

Het is moeilijker dan in 4.3 om duidelijke conclusies te trekken omdat we slechts 845 resultaten hebben om te beoordelen. Het lijkt uit de resultaten dat een CV API vooral goed te trainen is op beelden die een duidelijke weergave en thematiek hebben, zoals de Sinterklaasfoto's.

drempelwaarde (%)	aandeel juiste classificaties (%)	aantal foute classificaties (%)
95	87	3,7
90	91	4,6
80	95	6,6
55	100	11

Tabel 5.5: Verhouding tussen drempelwaarde, totaal aantal van de juiste classificaties en het aantal foute classificaties

Wat betreft waarschijnlijkheidsscores en drempelwaarden zijn er twee strategieën die gevuld kunnen worden:

1. er kan een drempelwaarde ingesteld worden;
2. men kan rekening houden met de waarschijnlijkheidsscore van het concept met de tweede hoogste score.

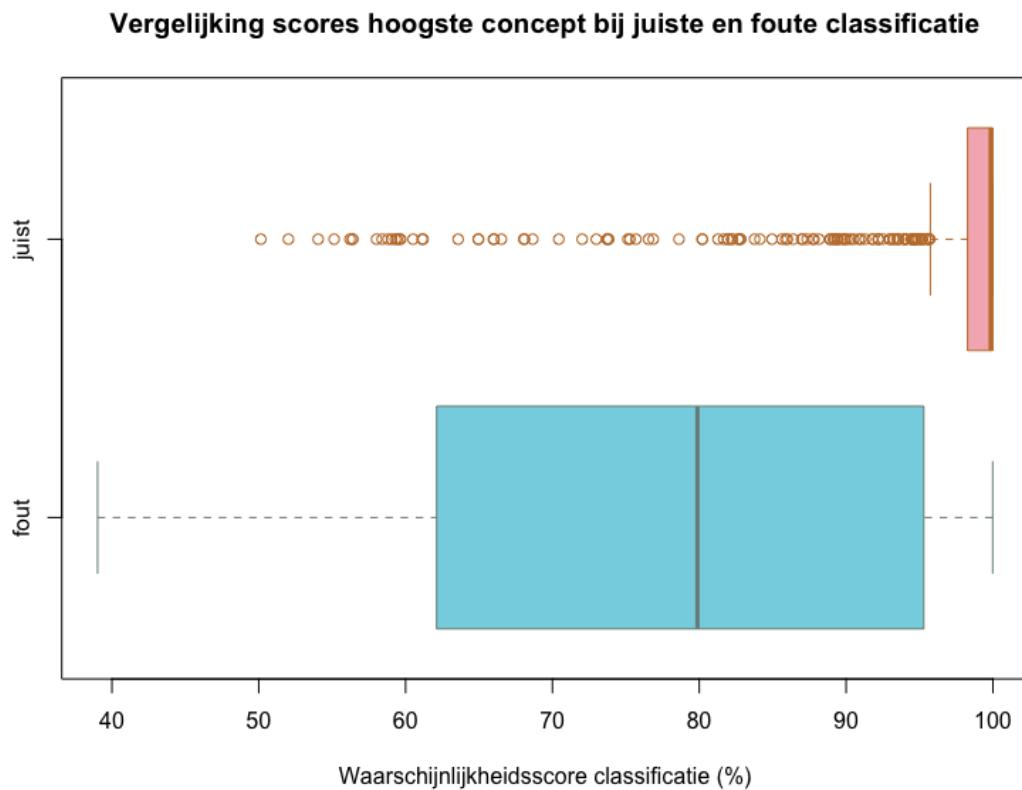
5.3.1 Instellen van een drempelwaarde

Zoals reeds vermeld is de gemiddelde waarschijnlijkheidsscore 94,6% voor zowel juist als fout geïdentificeerde foto's. De gemiddelde waarschijnlijkheidsscore voor de juiste classificaties ligt echter hoger (97%), terwijl die voor de foute classificaties lager ligt (77%). De mediaan²⁶ is respectievelijk 100% en 80%. Ook uit de boxplot in figuur 5.2 valt duidelijk te zien dat het grootste aandeel van de waarschijnlijkheidsscores van de juiste tags buiten het grootste aandeel waarschijnlijkheidsscores van de foute tags ligt. Het is dus mogelijk om een drempelwaarde in te stellen om zoveel mogelijk juiste classificaties te behouden en foute classificaties te weren (zie tabel 5.5).

5.3.2 Rekening houden met de waarschijnlijkheidsscore van de tweede tag

Anderzijds kan er ook gekeken worden naar de waarschijnlijkheidsscore van het tweede hoogste concept dat door het model gegeven wordt. Als het model een foto correct geïdentificeerd heeft, dan krijgt het tweede concept in 64% van de gevallen een waarschijnlijkheidsscore van (afgerond) 0%. Wanneer het model een foto fout geïdentificeerd heeft, dan ligt de mediaan van de waarschijnlijkheidsscore voor het tweede concept op 17% en het gemiddelde op 19%. Ook uit figuur 5.3 blijkt dat er weinig overlap is tussen de waarschijnlijkheidsscores van de tweede tag als de classificatie juist of fout is. Door de tags te weerhouden waarvan het tweede concept een te hoge score heeft, kunnen eveneens fouten vermeden worden. In tabel 5.6 worden hiervoor ook enkele voorstellen gedaan.

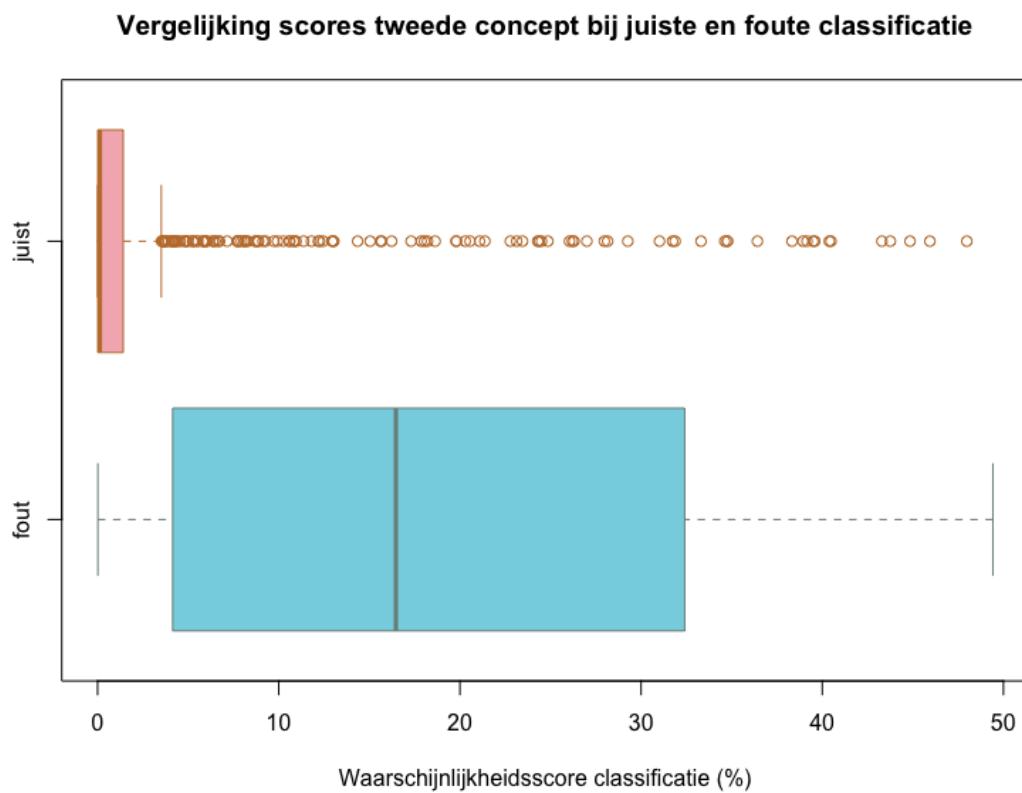
²⁶De mediaan is het midden van een verzameling van gegevens. 50% heeft bijgevolg een waarde dat lager is dan de mediaan; 50% een waarde hoger dan de mediaan.



Figuur 5.2: Twee boxplots die de waarschijnlijkheidsscores vergelijken van het Periodemodel bij juiste en foute classificatie. In tegenstelling tot in figuur 4.8 is er geen overlap tussen de interkwartielafstand van de foute classificaties en de boxplot van de juiste classificaties.

drempelwaarde (%)	aandeel juiste classificaties (%)	aantal foute classificaties (%)
5	87,5	3,3
10	91,6	4,5
15	93,6	5,6
48	100	12,1

Tabel 5.6: Verhouding tussen een drempelwaarde voor het tweede hoogste concept, totaal aandeel van de juiste tags en het aantal foute tags



Figuur 5.3: Twee boxplots die de waarschijnlijkheidsscores vergelijken van het tweede concept van het Periodemodel bij juiste en foute classificatie. Net als in figuur 5.2 is er geen overlap tussen de interkwartielafstand van de boxplot van de foute classificatie en de boxplot van de juiste classificatie.

6. Conclusie

Dit zijn vlug neergeschreven ideeën. Zie ik conclusies over het hoofd?

Het registreren van beelden vraagt veel werk. Musea slagen er niet in om die beelden goed doorzoekbaar te maken. Ook zijn de wensen van het publiek veranderd en willen ze makkelijk kunnen zoeken doorheen de collectie op basis van trefwoorden en termen. CV API's kunnen door musea gebruikt worden om de collectie te laten taggen en ze doorzoekbaar te maken. Uit de literatuurstudie blijkt dat een aantal musea dit al doen. De vraag is echter hoe deze services scoren op historische beelden. Ze zijn immers getraind met hedendaagse beelden. We stelden vast dat CV API iets slechter scoren op historische beelden, maar in het algemeen zijn de scores goed. Zeker als we ze vergelijken met de reeds bestaande beschrijvingen, dan zijn ze nuttig als aanvulling. De CV API kunnen de registrator ondersteunen om meer trefwoorden en termen toe te voegen aan beelden. Ze zijn sneller en bovendien vollediger. Het is wel nodig om een drempelwaarde in te stellen om het aantal fouten te verkleinen. Ook zijn de CV API niet geschikt voor thema's die niet universeel gekend zijn, zoals hier de Sinterklaasfoto's. De CV API scoren vooral goed op foto's waar mensen staan. Het ingebouwde model is wel niet voldoende om de beelden te classificeren volgens de thema's van Huis van Aijn.

Er is ook verschillend academisch onderzoek waarbij neurale netwerken getraind worden om de registrator te helpen bij het classificeren van beelden, per kunstenaar, per kunststroming, per periode, etc. Doordat het aantal beschikbare datasets voor erfgoedbeelden beperkt is, verkennen deze onderzoekers het veld van Transfer Learning. Ze bouwen verder op reeds beschikbare netwerken die getraind zijn op hedendaagse beelden en trainen ze verder met historische beelden. De resultaten hiermee zijn goed. Wij hebben dit ook geprobeerd met de CV API. We zijn verder gegaan met de reeds bestaande infrastructuur van de CV API en hebben een eigen custom model gecreëerd om de beelden van het Huis

van Aijn te classificeren volgens thema en periode. De resultaten met het Themamodel waren goed; die van het Periodemodel minder. Het zou kunnen dat deze uitdaging te moeilijk is voor de CV API, maar onze dataset was ook niet voldoende. Namelijk te klein. Ook hier kunnen drempelwaarden ingesteld worden om het aantal fouten te verkleinen. Trainen is relatief eenvoudig. Doordat er een webinterface is, kunnen ook mensen zonder programmeerervaring het gebruiken. De webinterface is toegankelijk en intuïtief. Uitdaging is om een goede dataset te voorzien waarin alle concepten goed getraind kunnen worden.

Verder onderzoek

- Het is nodig om het Themamodel uit te testen met een grotere dataset om echt uitspraken te kunnen doen. Ook op een dataset die nog niet ingedeeld is in thema's en die gevalideerd wordt door een museummedewerker
- Voor het Periodemodel moet er een betere trainingsset beschikbaar zijn om uit te zoeken of de CV API deze uitdaging aan kan.
- het matchen van tags van Clarifai met de termenlijst van Huis van Aijn → kunnen we deze zo instellen dat een combinatie van Clarifai-tags om te zetten is naar een term uit de termenlijst van Huis van Aijn (bv. bruidsboeket door de combi van de tags huwelijk en bloemen)
- CV API uitstellen met andere soort beelden, bv. beelden van collectieobjecten, schilderijen, standbeelden, etc. Foto's sluiten het dichtste aan bij de beelden waarmee de CV API getraind is.

A. Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

A.1 Introductie

De Vlaamse Overheid investeert in het Vlaamse cultureel erfgoed door collectiebeherende instellingen te ondersteunen en kwaliteitslabels uit te reiken. In ruil voor die steun verwacht de Vlaamse Overheid dat die instellingen een aantal taken uitvoeren, waaronder het registreren, inventariseren en metadateren van cultureel-erfgoedobjecten op een gestandaardiseerde manier, het onderzoeken (en faciliteren van onderzoek) en het presenteren van de collectie. (Agentschap Kunsten en Ergoed, 2014; Gatz, 2016)

De collectiebeherende instellingen lijden aan een historische achterstand m.b.t. de registratie van de eigen collectie (Gatz, 2016). In 2018 werd daarom een nieuwe subsidielijn opgestart om de digitale collectieregistratie weg te werken. Deze subsidielijn werd opgestart vanuit de vaststelling dat de competenties en strategieën ontbreken om een inhaalbeweging te realiseren (Departement Cultuur, Jeugd en Media, 2018).

In de bachelorproef willen we onderzoeken in welke mate Computer Vision API's (vanaf nu afgekort als CVA), zoals Google Cloud Vision²⁷ of Microsoft Computer Vision API²⁸, ingezet kunnen worden om dit registratieproces te versnellen en als strategie gebruikt kunnen worden om een inhaalbeweging te realiseren. Momenteel gebeuren registraties

²⁷<https://cloud.google.com/vision/>

²⁸<https://azure.microsoft.com/nl-nl/services/cognitive-services/computer-vision/>

door domeinexperten. Dit is tijdrovend werk. Met behulp van artificiële intelligentie (AI) kan dit proces deels geautomatiseerd worden. Dit geeft de collectieregistrator de mogelijkheid om zich met minder basaal werk bezig te houden en geeft de musea de kans hun collectie sneller te ontsluiten.

Het VKC Datahub Dashboard²⁹ geeft een goed beeld van de registratieachterstand in Vlaanderen. Dit dashboard analyseert de collectieregistratie van de Vlaamse musea voor Schone (en in de toekomst ook Hedendaagse) Kunsten die aangesloten zijn bij de VlaamseKunstCollectie (VKC)³⁰, waaronder het aantal records ingevuld volgens de minimale registratie en het aantal records ingevuld volgens de basisregistratie.³¹ Dit zijn cijfers van de grootste kunstmusea van Vlaanderen. Uit de cijfers blijkt dat vooral formele en administratieve gegevens geregistreerd worden; inhoudelijke informatie, zoals afgebeeld persoon of afgebeeld concept, die interessant is voor ontsluiting en onderzoek, ontbreekt hoofdzakelijk.

In het verleden werd reeds (theoretisch) onderzoek verricht naar het gebruik van Computer Vision voor cultureel erfgoed en kunst.³² De resultaten van dit onderzoek waren veelbelovend. Nieuwe ontwikkelingen zorgen ervoor dat Computer Vision steeds meer accuraat wordt en steeds eenvoudiger in gebruik, waardoor het meer en meer een instrument wordt waarmee ook developers aan de slag kunnen (Hindle, 2017).

A.2 Stand van zaken

A.2.1 Computer Vision en (cultureel) erfgoed

Een aantal instellingen maken al gebruik van Computer Vision voor de ontsluiting van hun collectie:

- **The Museum of Modern Art (MoMA)**³³ gebruikte diensten van Google via *Google Arts Culture Lab* om aan de historische foto's van afgelopen tentoonstellingen in MoMA foto's uit de kunstcollectie te koppelen. Het algoritme analyseerde hiervoor alle foto's van tentoonstellingszichten. Als het een kunstwerk op de foto's kon herkennen, dan werd er een koppeling gelegd met de afbeelding van dit kunstwerk in de collectie van MoMA.³⁴ MoMA stelde hierbij vast dat het logaritme vooral goed scoort op het vlak van tweedimensionele, statische afbeeldingen (bv. een afbeelding van een kunstwerk), maar dat het minder goede resultaten geeft met afbeeldingen

²⁹<https://dashboard.vlaamsekunstcollectie.be>

³⁰<http://vlaamsekunstcollectie.be/>

³¹De minimale registratie omvat de elementen en velden die in overeenstemming met de ICOM Code of Ethics minimaal gedocumenteerd moeten worden wanneer een object een museum binnenkomt: d.i. bewaarinstelling, objectnummer, titel, korte beschrijving, objectnaam, verwervingsmethode, verwervingsbron en verwervingsdatum. De velden voor de basisregistratie bevatten de acht velden voor minimale registratie en voegen hier nog een tiental andere velden aan toe. Voor meer info, zie het Invulboek Objecten op CEST

³²Zie infra, in het deel *Stand van zaken*.

³³<https://www.moma.org/>

³⁴Bekijk een voorbeeld over een tentoonstelling over Cézanne, Gauguin, Seurat en Van Gogh uit 1929

van 3D-objecten (bv. een standbeeld) of bewegend beeld (MOMA, 2018).

- Het **Smithsonian**³⁵ gebruikt AI-technologie om stalen van planten te classificeren. Het Smithsonian is gestart met het systematisch digitaliseren van de collectie voor wetenschappers en in functie van online ontsluiting. De AI-technologie slaagde erin om via deze afbeeldingen twee gelijkaardige planten te onderscheiden met een succesgraad van meer dan 90% (Smith, 2017).
- Het **Nasjonalmuseet**³⁶ was het onderwerp van het *Principal Components* onderzoek. Hierbij werd met het deep learning framework Caffe³⁷ gezocht naar compositieën gelijkenissen tussen kunstwerken en werden ze geklassificeerd op basis van de Iconclass-termen³⁸. Dit resulteerde in een vernieuwende publiekstoegang tot de collectie waarbij de kunstwerken op basis van gelijkenissen gevisualiseerd werden. Hoe meer gelijkenissen, hoe dichter de kunstwerken bij elkaar staan.³⁹ (Nasjonalmuseet, 2017; Westvang, 2017)

A.2.2 Theoretisch onderzoek

Daarnaast is eerder theoretisch onderzoek gedaan naar het gebruik van Computer Vision binnen een erfgoedcontext. In onderstaande lijst worden de onderzoeken vermeld die zich focussen op museumcollecties:

- **Using Machine Learning for Identification of Art Paintings (2013):** In dit onderzoek werd machine learning gebruikt voor het classificeren van kunstwerken van zeven kunstenaars: Cézanne, Dali, Dürer, Monet, Picasso, Rembrandt en Van Gogh. Per kunstenaar werden er tweehonderd kunstwerken gezocht. In 87,13% van de gevallen was de computer correct. De onderzoekers vermoedden dat dit resultaat nog beter kan zijn als de training set groter is. (Blessings & Wen, 2013)
- **The Rijksmuseum Challenge: Museum-Centered Visual Recognition (2014):** Beelden die publiek beschikbaar zijn via de API van Het Rijksmuseum⁴⁰ werd gebruikt voor het oplossen van vier challenges:
 - voorspel de kunstenaar van het afgebeelde object;
 - voorspel het materiaal dat gebruikt werd voor het afgebeelde object;
 - voorspel het jaar waarin het object gemaakt werd;
 - voorspel het soort object (schilderij, tekening, standbeeld...) dat afgebeeld wordt.Het ging om objecten die afkomstig waren van de oudheid tot de late 19e eeuw en om een veelheid aan objecten: schilderijen, foto's, keramiek, meubels, etc. Ook in dit onderzoek waren de resultaten veelbelovend. (Mensink & van Gemert, 2014)
- **INSIGHT (2017-2020):** Hier wil men onderzoeken hoe AI kan gebruikt worden om collecties uit de cultureel-erfgoedsector van beschrijvende metadata te voorzien.

³⁵<https://www.smithsonianmag.com/>

³⁶<http://www.nasjonalmuseet.no/en/>

³⁷Voor meer info, zie: <http://caffe.berkeleyvision.org/>

³⁸Iconclass is een gespecialiseerd kunsthistorisch classificatiesysteem, <https://nl.wikipedia.org/wiki/Iconclass>

³⁹Bekijk bijvoorbeeld schilderijen op basis van hun motief.

⁴⁰<https://www.rijksmuseum.nl/en/api>

De collecties van de Koninklijke Musea voor Schone Kunsten van België en de Koninklijke Musea voor Kunst & Geschiedenis worden als testcase gebruikt. De focus ligt op het vrijgeven van die data als open datasets. (Universiteit Antwerpen, 2017) Uit de paper *Deep Transfer Learning for Art Classification Problems* van Sabatteli e.a. (2018) kwamen veelbelovende resultaten naar boven voor het voorspellen van materiaal, objecttype en kunstenaar.

- **Automated Image Analysis with IIIF (2017):** Dit onderzoek is eerder praktisch gericht en werd uitgevoerd buiten de academische wereld. Het werd uitgevoerd door CogApp, een bedrijf dat software ontwikkelt voor online archieven en musea.⁴¹ Zij hebben verschillende testen gedaan met machine learning. Het meest interessant voor dit voorstel is het onderzoek waarbij drie Computer Vision API's (Google Cloud Vision, Microsoft Computer Vision en Clarifai⁴²) getest werden om beelden te voorzien van extra tags om de doorzoekbaarheid te verbeteren. Naast de grappige resultaten (Roddis, 2018), leidde dit tot goede resultaten waarmee de beeldencollectie op een interessante manier doorzocht kan worden: Je kan er een selectie van beelden verkrijgen door te filteren op verschillende tags, bv. *ik wil een kunstwerk uit de Renaissance van iemand met een snor en een cape*.⁴³ De onderzoekers concludeerden dat de CVA accurate beschrijvingen geven en dat Computer Vision steeds eenvoudiger in gebruik wordt. De foutieve beschrijvingen, zoals in Figuur 2, zijn vermoedelijk een gevolg van het trainen van CVA met hedendaagse beelden die gemaakt werden met een smartphone. (Hindle, 2017)

A.2.3 Computer Vision API's

Tot slot lichten we kort CVA toe. CVA, ook wel image/visual recognition API's genoemd, zijn API's die afbeeldingen automatisch kunnen taggen (metadateren), organiseren en zoeken via machine learning. Het biedt de mogelijkheid om machine learning toe te passen zonder dat je hier zelf een expert in moet zijn. Je kan er zelf modellen mee creëren om de API's te trainen naar je eigen behoefté. Zo leerde Matt Fraser Google Cloud Vision verschillende spinnen te herkennen door de dienst te trainen met honderd afbeeldingen per spin (Fraser, 2018). De bekendste diensten zijn Clarifai, IBM Visual Recognition⁴⁴, Microsoft Computer Vision en Google Cloud Vision. De diensten zijn zo toegankelijk dat iedere ontwikkelaar zonder machine learning kennis modellen kan bouwen, maar dat ook, volgens TechCrunch, iedereen zonder programmeerervaring de diensten kan gebruiken om afbeeldingen te taggen en categoriseren (Lardinois, 2018). Ook uit het onderzoek van CogApp bleken de diensten eenvoudig in gebruik en precies te trainen (Hindle, 2017).

⁴¹<https://www.cogapp.com/about/>

⁴²<https://www.clarifai.com/>

⁴³Probeer het zelf op: <http://labs.cogapp.com/iiif-ml/>

⁴⁴<https://www.ibm.com/watson/services/visual-recognition/>

A.3 Methodologie

In dit onderzoek zal bestudeerd worden of CVA helpen bij het inhoudelijk ontsluiten van erfgoedcollecties. Kunnen de ingebouwde modellen van de API's gebruikt worden, of is er nood aan (doorgedreven) training? Dit wordt onderzocht aan de hand van een prototype waarmee verschillende beelden geanalyseerd worden. Hiervoor wordt een CVA gekozen die het toelaat om eigen modellen te trainen via een webinterface, voorzien is van goede documentatie en tutorials en beschikt over API clients in een programmeertaal die gekend is door de onderzoeker (o.a. Java, Javascript, C#). Achtereenvolgens wordt een vergelijking gemaakt tussen een set beelden die manueel beschreven werden, een set beelden die beschreven werden door de ongetrainde CVA en een set beelden die beschreven werden door een getrainde CVA.

We gebruiken hiervoor beelden van het Huis van Alijn. Het Huis van Alijn is het museum van het dagelijkse leven⁴⁵ Het museum beschikt over een grote collectie beelden die het dagelijkse leven uit de 20e eeuw documenteren. In 2011 organiseerde het museum een crowdsourcingproject waarbij foto's uit de collectie 'Anonieme snapshots' getagged konden worden om ze te beschrijven en beter toegankelijk te maken (Wiericx, 2011). Het zou interessant zijn om deze beelden te vergelijken met de resultaten van de CVA. Collectiemedewerkers van Huis van Alijn zullen mee de resultaten van de proefopstellingen analyseren en bepalen welke set van beelden voor hen over de meest bruikbare tags of metadata beschikken: de manueel beschreven beelden, de beelden getagged door de ongetrainde CVA of de beelden getagged door de getrainde CVA.

A.4 Verwachte resultaten

Van dit onderzoek worden drie resultaten verwachtten.

1. een methodologie die door musea gebruikt kan worden om CVA in te zetten voor collectiebeschrijving. In een rapport zal beschreven worden hoe een museum CVA kan gebruiken en hoe modellen getraind kunnen worden via de webinterface en API clients. Als de CVA zo eenvoudig in gebruik zijn (Lardinois, 2018), dan zouden collectieregistratoren zelf in staat moeten zijn om de software te trainen om hen bij te staan in het ontsluiten van de collectie. Het rapport van de methodiek zal zo opgesteld worden dat de musea er zelf mee aan de slag kunnen.
2. concrete use cases die aansluiten bij de registratiepraktijk en behoeftes van een museum.
3. musea worden bewust gemaakt over het bestaan van computer vision en hoe die technologie kan bijdragen in hun werking.

Afgaand op de resultaten uit reeds eerder gevoerd onderzoek, vermoeden we dat de CVA goed scoren op het herkennen van materiaal, het type object en de voorgestelde objecten en figuren op het object. De kwaliteit van de resultaten zal afhankelijk zijn van de specificiteit

⁴⁵<http://www.huisvanalijn.be>

die verwacht wordt. Hoe specifieker de resultaten van de CVA moeten zijn, hoe meer de modellen getraind moeten worden.

A.5 Verwachte conclusies

CVA kan de metadata van een erfgoedobject verrijken met inhoudelijke informatie die nu niet voorkomt in de basisregistratie, zoals de sfeer en het gevoel dat op een kunstwerk weergegeven wordt en de gebruikte kleuren. Mogelijk is het zelfs in staat om bekende personen op de afbeelding te herkennen, zoals bij Hindle (2017). Deze informatie wordt nu niet opgenomen in collectiebeheersystemen, maar is wel interessant voor onderzoekers en het publiek. We verwachten dat de meerwaarde van de CVA vooral ligt in het bijstaan van de registrator bij het inhoudelijk beschrijven. Voor collecties van het museum die momenteel niet beschreven worden door een gebrek aan personeel, tijd of budget kunnen CVA een eenvoudige manier om deze collecties toch van inhoudelijke metadata te voorzien en ze doorzoekbaar te maken.

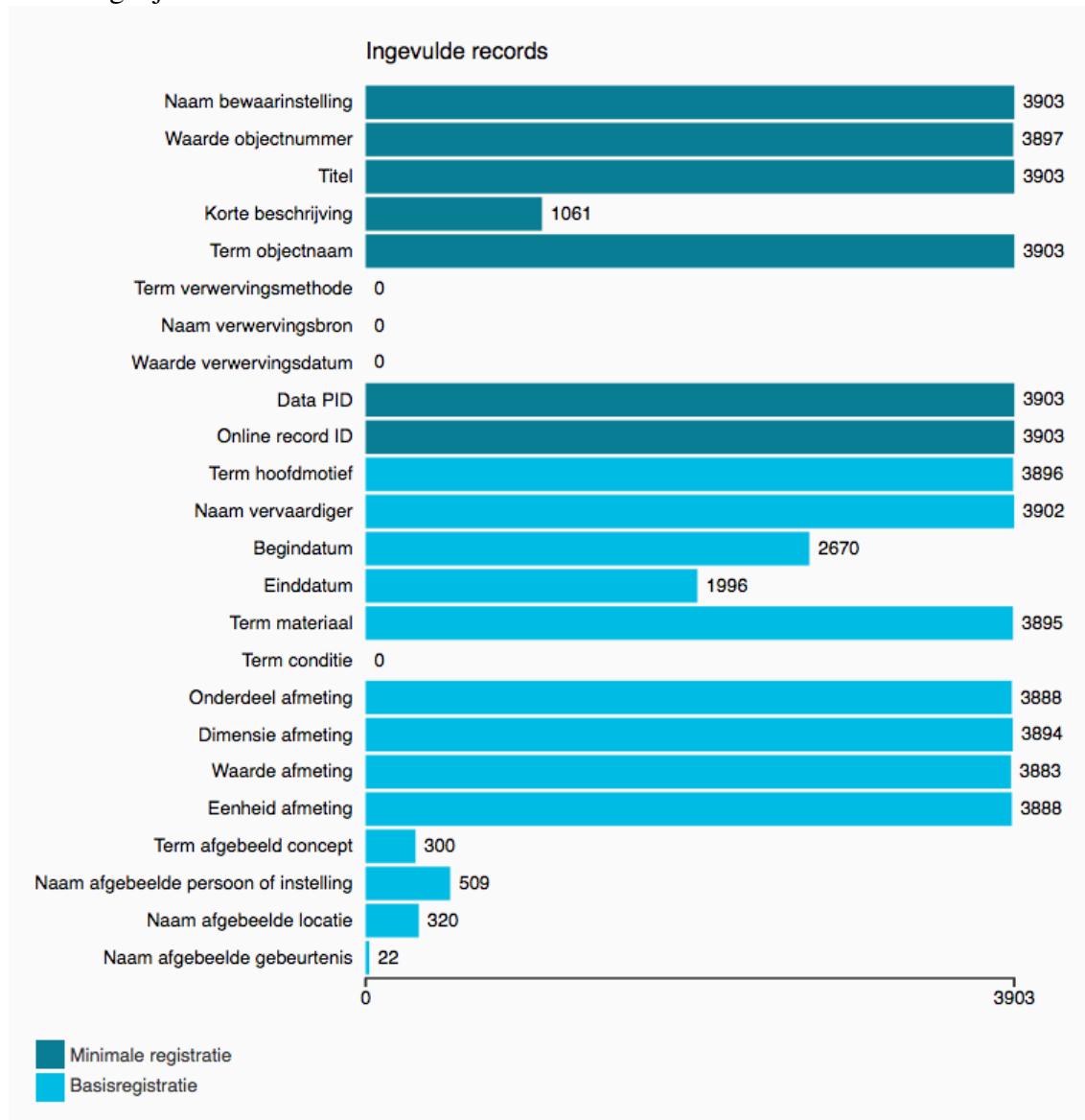
CVA kunnen een hulp zijn bij het realiseren van een inhaalbeweging voor registratieachterstand. Er wordt verwacht dat ze de registrator kunnen ondersteunen in hun taak en informatie kunnen bezorgen die zowel voor het museum, de onderzoeker als de erfgoedgeïnteresseerde interessant is. De CVA zal de registrator niet vervangen, maar kan een meerwaarde bieden aan diens werk en de registrator bijstaan bij het beschrijven van de collectie. In dit onderzoek willen we daarom aantonen wat de meerwaarde van CVA voor collectieregistratoren zijn en dat de technologie een hulpmiddel kan zijn voor collecties die niet beschreven kunnen worden.

Deze technologie zal meer en meer ingebouwd worden in DAM-systemen⁴⁶. DAM-software met Computer Vision is al op de markt: Gezichtsherkenning is ingebouwd in verschillende DAM-systemen.⁴⁷. In dit onderzoek willen we aantonen voor welke use cases deze software gebruikt kan worden in musea.

⁴⁶Digital Asset Management Systemen, een systeem voor het opslaan, beheren en distribueren van digitale assets, zoals afbeeldingen.

⁴⁷Zoals bij ResourceSpace: <https://www.resourcespace.com/knowledge-base/user/facial-recognition>

Figuur A.1: Het overzicht van de velden van de basisregistratie en het aantal keer dat ze aanwezig zijn in de collectiedata van MSK Gent.



Figuur A.2: Vrouw en man met boek worden door Microsoft Computer Vision geïdentificeerd als man en vrouw die een selfie nemen (Roddis, 2018).



A man and a woman taking a selfie

Bibliografie

- Agentschap Kunsten en Ergoed (Red.). (2014). *Handleiding bij het Cultureel-Erfgoeddecreet: Het Kwaliteitslabel*.
- Blessings, A. & Wen, K. (2013). *Using Machine Learning for Identification of Art Paintings*. Stanford University.
- Brownlee, J. (2017, juli 14). What is the Difference Between Test and Validation Datasets? Verkregen 21 maart 2019, van <https://machinelearningmastery.com/difference-test-validation-datasets/>
- Brownlee, J. (2019a, maart 19). A Gentle Introduction to Computer Vision. Verkregen 21 maart 2019, van <https://machinelearningmastery.com/what-is-computer-vision/>
- Brownlee, J. (2019b, mei 1). A Gentle Introduction to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Verkregen 21 juni 2019, van <https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/>
- Clairbot. (2019). General FAQ | Clarifai Help Center. Verkregen 24 juni 2019, van <http://help.clarifai.com/articles/687603-general-faq>
- Clarifai. (n.d.-a). API Guide. Verkregen 2 juli 2019, van <http://developer-dev.clarifai.com/developer/guide/>
- Clarifai. (n.d.-b). General Model | Clarifai. Verkregen 2 juli 2019, van <https://www.clarifai.com/models/general-image-recognition-model-aaa03c23b3724a16a56b629203edc62c#documentation>
- Departement Cultuur, Jeugd en Media. (2018, juli 12). Inhaalbeweging voor digitale collectieregistratie. Verkregen 8 december 2018, van <http://www.kunstenenerfgoed.be/nl/nieuws/inhaalbeweging-voor-digitale-collectieregistratie>
- Derveaux, A. (2019). Beeldherkenning in de registratiepraktijk. Verkregen 22 juni 2019, van https://www.fotomuseum.be/collectie/onderzoek0/projecten/Beeldherkenning_in_de_registratiepraktijk.html

- Dickson, A. (2018, juli 20). A.I. and the Art of Spotting Fakes. Verkregen 1 april 2019, van <https://medium.com/s/story/a-i-and-the-art-of-spotting-fakes-6a674b0bdcef>
- Dickson, A. (2019, maart 29). A.I. Will Enhance - Not End - Human Art. Verkregen 1 april 2019, van <https://onezero.medium.com/a-i-will-enhance-not-end-human-art-f575e9ff9325>
- Elgammal, A. (2017, november 16). Picasso, Matisse, or a Fake? A.I. for Attribution and Authentication of Art at the Stroke Level. Verkregen van https://medium.com/@ahmed_elgammal/picasso-matisse-or-a-fake-a-i-for-attribution-and-authentication-of-art-at-the-stroke-level-f4ec329c8c26
- Elgammal, A., Mazzone, M., Liu, B., Kim, D. & Elhoseiny, M. (2018). *The Shape of Art History in the Eyes of the Machine*. Rutgers University en College of Charlston.
- Fraser, M. (2018, maart 14). Using Google Cloud AutoML to classify poisonous Australian spiders. Verkregen 9 december 2018, van <https://shinesolutions.com/2018/03/14/using-google-cloud-automl-vision-to-classify-poisonous-australian-spiders/>
- Gatz, S. (2016, maart). Conceptnota aan de Vlaamse Regering. Naar een duurzame cultureel-erfgoedwerking in Vlaanderen. Een langetermijnvisie voor cultureel erfgoed en cultureel-erfgoedwerking in Vlaanderen. Vlaamse Regering.
- Gong, K. (2017, november 13). Best Practices for Custom Models in Watson Visual Recognition. Verkregen 5 maart 2019, van <https://medium.com/ibm-watson/best-practices-for-custom-classifiers-in-watson-visual-recognition-1015a273f75d>
- Haskiya, D. (2019, april 9). How to set up a generous interface prototype in less than a day. Verkregen 1 april 2019, van <https://pro.europeana.eu/post/how-to-set-up-a-generous-interface-prototype-in-less-than-a-day>
- Heemkunde Vlaanderen. (2011, november 7). Huis van Alijn zoekt taggers (m/v) voor beschrijven van anonieme snapshots. Verkregen 11 augustus 2019, van <https://www.heemkunde-vlaanderen.be/huis-van-alijn-zoekt-taggers-mv-voor-beschrijven-van-anonieme-snapshots/>
- Hindle, A. (2017, juni 20). Automated image analysis with IIIF: Using Artificial Intelligence for bulk image analysis. Verkregen 8 december 2018, van <https://blog.cogapp.com/automated-image-analysis-with-iiif-6594ff5b2b32>
- ING. (2016). The Next Rembrandt. Verkregen van <https://www.nextrembrandt.com/>
- Koehrsen, W. (2018, maart 3). Beyond Accuracy: Precision and Recall. Verkregen 6 augustus 2019, van <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>
- Lardinois, F. (2018, januari 17). Google's AutoML lets you train custom machine learning models without having to code. Verkregen 17 juni 2019, van <https://techcrunch.com/2018/01/17/googles-automl-lets-you-train-custom-machine-learning-models-without-having-to-code>
- Lievens, S. (2017). *Artificiële Intelligentie: Lesnota's*. Ongepubliceerde cursus.
- Mensink, T. E. J. & van Gemert, J. C. (2014). The Rijksmuseum Challenge: Museum-Centered Visual Recognition. In *ACM International Conference on Multimedia Retrieval*.
- MOMA. (2018). Identifying art through machine learning: A project with Google Arts & Culture Lab. Verkregen 8 december 2018, van <https://www.moma.org/calendar/exhibitions/history/identifying-art>

- Moriarty, A. (2018, juli 23). AI and Museum Collections. Verkregen van <https://medium.com/@adamrmor/ai-and-museum-collections-c74bdb724c07>
- Nasjonalmuseet. (2017). Project: "Principal Components". Verkregen 8 december 2018, van http://www.nasjonalmuseet.no/en/collections_and_research/collection_management/digital_collection_management/Project%3A+%C2%ABPrincipal+Components%C2%BB.b7C_wJjU4L.xls
- Oberoi, G. (2016, juli 11). Comparing the Top Five Computer Vision APIs. Verkregen 9 december 2018, van <https://goberoi.com/comparing-the-top-five-computer-vision-apis-98e3e3d7c647>
- Pim, H. (2018, oktober 17). Exploring Wellcome Collection with computer vision. Verkregen van <https://stacks.wellcomecollection.org/exploring-wellcome-collection-with-computer-vision-7513dff8126d>
- Pokharna, H. (2016, juli 26). For Dummies - The Introduction to Neural Networks we all need ! (Part 1). *TechnologyMadeEasy*. Verkregen 16 juni 2019, van <https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb>
- Roddis, T. (2018, mei 16). When automated analysis goes wrong. Verkregen 8 december 2018, van <https://www.slideshare.net/Europeana/when-automated-analysis-goes-wrong-by-tristan-roddis-europeanatech-conference-2018>
- Rowe, P. (2017, oktober 31). Looking at Sarjeant Gallery's collection through robot eyes. Verkregen van https://medium.com/@armchair_caver/looking-at-sarjeant-gallerys-collection-through-robot-eyes-c7fd0281814e
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fie, L. (2014). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision - September 2014*.
- Sabatteli, M., Kestemont, M., Daelemans, W. & Geurts, P. (2018). *Deep Transfer Learning for Art Classification Problems*. Université de Liège en Universiteit Antwerpen.
- Smith, R. P. (2017, november 3). How Artificial Intelligence Could Revolutionize Archival Museum Research. Verkregen 8 december 2018, van <https://www.smithsonianmag.com/smithsonian-institution/how-artificial-intelligence-could-revolutionize-museum-research-180967065/>
- Tsang, S.-H. (2018, augustus 19). Review: ZFNet - Winner of ILSVRC 2013 (Image Classification). Verkregen 2 juli 2019, van <https://medium.com/coinmonks/paper-review-of-zfnet-the-winner-of-ilsvrc-2013-image-classification-d1a5a0c45103>
- INSIGHT. (2017). Verkregen 8 december 2018, van <http://uahost.uantwerpen.be/insight/index.php/about/>
- Vanstappen, H. (2019). VR4CH: Visual recognition voor erfgoedcollecties. *Eindrapport*.
- Westvang, E. (2017). Principal Components: Machine learning in search of the uncanny. Verkregen 8 december 2018, van <http://bengler.no/principalcomponents>
- Wiericx, B. (2011, augustus 9). Crowdsourcing in het Huis van Alijn. Verkregen van <https://faro.be/blogs/bram-wiericx/crowdsourcing-in-het-huis-van-alijn>
- Wikipedia contributors. (2019a). Computer vision — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Computer_vision&oldid=896283892.
- Wikipedia contributors. (2019b). ImageNet — Wikipedia, The Free Encyclopedia. <https://en.wikipedia.org/w/index.php?title=ImageNet&oldid=900080629>.