

Melvin Wevers, Nico Vriend, Alexander de Bruin

What to do with 2.000.000 Historical Press Photos? The Challenges and Opportunities of Applying a Scene Detection Algorithm to a Digitised Press Photo Collection

Abstract

In 1962, Dutch celebrity Ria Kuyken was attacked by a circus bear. Cees de Boer captured this moment, for which he was awarded both a World Press Photo and the Silver Camera (Zilveren Camera). Though this photo popularised Fotopersbureau De Boer, which Cees had founded in 1945, the importance of the collection lies in its scale. Approximately 2,000,000 photos taken of about 250,000 events in sixty years, accompanied by extensive metadata. Not only major nationwide events are represented, but also subjects of small scale, human interest, such as the shopkeeper around the corner. Our aim is not only the digitisation and publication of all 2,000,000 photo negatives of Fotopersbureau De Boer but also to explore how artificial intelligence can enrich this collection, benefiting both users of the archive and cultural historians studying historical photographs. One of our efforts focuses on scene detection, a method to detect the ‘scene’ represented in an image (Zhou et al, 2018). We will rely on transfer learning to adapt existing computer vision models to our collection and the needs of our users. Existing models can generate labels with high accuracy, however, these labels are ahistorical and more often than not irrelevant to our collection. We will label subsets of the images via crowdsourcing to train and improve existing models. As such, we can add labels relevant to our collection to the model, which are absent in existing models. In this paper, we will highlight the opportunities and challenges of applying artificial intelligence to a collection of historical photographs.

Keywords

press photography, computer vision, scene detection, digital history, digitisation

Introduction

In recent decades, historians have started to emphasise the importance of photographs as historical sources of evidence. Even though historians commonly draw from textual sources or oral testimony, Peter Burke argues that we should start taking images seriously as a means to learn about the past.¹ Historical photographs and what is represented in them can be used to study specific world views, cultural-historical phenomena, or expressions of identity.² Edwards and Hart examined the photograph's materiality and traced the material forms in which images appeared.³ Others took a more contextual approach and examined how institutional forces determined how images were produced, distributed, and presented.⁴ A specific focus is on press photography and how this has shaped how we view images and read the news.⁵

The emphasis on visual material dovetailed with the wave of digitisation that started in the late twentieth century. Many archival sources, including photographs, have been digitised and made accessible. The availability of these sources in digitised form makes it possible to apply computation to them, offering new possibilities for research. These possibilities range from searching for specific sources to large-scale analysis of cultural datasets, in what Lev Manovich dubbed cultural analytics.⁶ In the case of textual sources, we have seen how users can apply full-text search to find instances of specific word use among millions of documents.⁷ Moreover, through text mining techniques, scholars can now also analyse linguistic trends at scale, in what Franco Moretti's has dubbed *distant reading*. Moretti explains that distance is a 'condition of knowledge ... [that] allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems.'⁸ As such, the text disappears.

Technological innovations in computational image analysis and the available digitised visual material allow scholars to apply similar methods, in what Lauren Tilton and Taylor Arnold have aptly titled *distant viewing*. The goal here is to extract 'semantic metadata from images.'⁹ These metadata can range from descriptions of colour use, the materiality of the image, the descriptions of camera angles, the detection of specific objects, or the labelling of images. These metadata can then guide exploratory data analysis or more quantitative approaches. The computational methods that underpin *distant viewing* are also known as computer vision: a subdomain of Artificial Intelligence aimed at gaining a high-level understanding of images. After making headway in its use for self-driving cars, surveillance, and the analysis of social media posts, the technology is now gaining

ground in historical research and the cultural heritage sector.¹⁰ Computation can be used to detect duplicate or near-duplicate images, colour use through a collection, or it can classify images according to medium, for example, illustrations or photographs.¹¹ Computer vision can study not only form and medium but also the content of images. Algorithms can guide queries for images that share visual similarities, i.e., images that look alike or images containing visually similar objects.¹² In addition, we can use computer vision to classify an image (image classification) or to detect objects and their position in images (object detection). The former adds metadata to images that can aid search and enable scholars to select images based on their classification. The latter allows for more fine-grained analysis and can be used to examine the composition and position of certain elements in images, which can support semiotic content analysis. For example, one could measure the position of people in an image.¹³

The ability to analyse visual objects at scale, allows us to make inferences about trends and compare collections. As such, we can add robustness to existing studies of visual (historical) culture that often rely on relatively small sets of images.¹⁴ Furthermore, having knowledge on trends adds context to outliers and helps us underline the uniqueness of an image. At the same time, it can also highlight diversity or bias in the representation of certain cultural-historical phenomena.¹⁵

However, the application of Computer Vision methods to historical photo collections is no straightforward endeavour. Although the potential might be clear, when these methods are applied to historical image collections, it becomes apparent that some constraints and challenges need to be addressed. Thus, this article describes the possibilities and challenges when applying computer vision to a specific press photo collection, namely the recently digitised Fotopersbureau De Boer (photo press agency De Boer) press photo collection. Moreover, we focus on one specific computer vision task: *scene detection*. This image classification task describes the scene depicted on an image. We discuss how the metadata generated through this task can benefit both users of the archive and cultural historians. We also describe possible pitfalls related to this type of metadata. While this concerns one dataset and a specific computer vision task, this case study offers clear insights into broader conceptual and practical issues related to computer vision.

Firstly, we briefly explain computer vision and argue how it can be applied to historical visual material, specifically press photography. We examine this question from the perspective of historians and heritage institutes. The goals of these two groups do not necessarily overlap. However, both parties are instrumental in adapting this technology to new domains and making it accessible to the

community of scholars and the public. Secondly, we describe our dataset and explain its relevance. Thirdly, we present a use case to show how we have annotated our material and applied the scene detection algorithm. Finally, we offer some concluding remarks and possible avenues for future work.

Computer Vision for Historical Research

While we have seen a shift towards visual material in (digital) history, the focus is still predominantly on textual sources. In the last decade, many textual sources have been digitised and processed with Optical Character Recognition, which turns the text into a machine-readable format, making it possible for scholars to apply computational methods to process historical texts. In the last few years, the analysis of visual material has been made more feasible with the increasing accuracy of computer vision models and the availability of frameworks, such as Keras and Fast.AI.¹⁶ These frameworks make it easier to train and apply models to data.

One technological innovation increased the accuracy of computer vision, namely Convolutional Neural Networks (CNNs). CNNs are part of what is called Deep Learning. These networks consist of multiple layers of artificial ‘neurons’ that can extract features from images, which the algorithm uses to classify images. The lower layers in this deep structure can identify simple features, such as edges, while higher levels can extract shapes resembling ears or wheels. The neural network relies on features extracted in the lower layers to determine features in the higher levels. For an algorithm to learn this information, labelled datasets are processed through the neural network, from which the algorithm then *learns* the visual features. The result is a model equipped to make specific inferences, for example, the detection of cats or dogs in an image.¹⁷ In summary, neural networks can be trained to learn what groups of pixels (features) best describe the visual elements capturing specific content or stylistic information.

There are two critical questions related to the use of computer vision on historical material. Firstly, can we use existing models off the shelf? Existing models can make predictions with high accuracy; however, these models were trained on contemporary data labelled with specific use cases in mind. Consequently, the objects that existing models can detect in images are limited (often around 40-50 types). More importantly, the labels are primarily ahistorical and often irrelevant to historical collections, e.g., parking metre and frisbee. In the case of similar labels, such as ‘telephone’, the visual representations of the objects often differ between modern and historical photographs. Also, the medium itself has often changed, granting contemporary photographs a different

materiality than historical ones. Colour was introduced, and technological innovations such as zoom lenses changed how photographs look. Models trained on millions of contemporary images are not sensitive to deal with the colour schemes and object representations in older images.

To adapt existing models to the use cases relevant to the heritage sector and historical research, we need to train models with historical material and rethink which labels to use and annotate images with such labels. Manual annotation for computer vision entails either adding a label to the entire image (image classification) or drawing boxes around specific objects and labelling them (object recognition). The algorithms can then use these annotations as training material to produce better models. Luckily, we do not have to label millions of images; we can rely on the ‘knowledge’ already stored in existing models. There are two main ways to build upon the existing models. One approach to adapt the models to the historical domain is *fine-tuning* their performance by *feeding* them historical images. This method builds upon the categories existing in modern data sets. For example, the model learns to recognize cars from the 1970s after processing images labelled as a car from this period. While this improves accuracy when working with historical material, we are still working with the same labels.

Another approach is *transfer learning*, which replaces the existing labels with an updated annotation scheme. The existing model is further optimised using the historical material annotated with these new categories.¹⁸ Because the models have already been pre-trained on millions of images, we often only require a small number of images to learn a new category. Of course, this depends on the visual complexity of the category and the diversity of the category in the training data. In other words, an object that always looks the same is easier to learn than one that shares visual aspects but also differs considerably.

The second key question related to computer vision and historical material is methodological. Why would one even want to apply these methods to our historical visual material? Firstly, the enrichments added to images through computer vision can aid search. Currently, it is difficult or often impossible to search for images in archives. When images have been correctly processed using OCR, we can search for text in images. Also, we can search based on metadata that the archive linked to visual material. However, the level of detail in metadata varies considerably from extensive, specific descriptions to brief descriptions of the type of images or often bearing no metadata on the content. Computer vision can help add metadata to images at scale; these metadata could range from a description of colour use, medium type, or subject matter. Existing research that applies computer vision methods to historical photographs, investigates automatically the dating of images, the finding

of matches between images taken from different viewpoints,¹⁹ photogrammetric analysis of images,²⁰ and the classification of historical buildings.²¹ These applications make it possible for users to browse through collections based on the metadata they generate.²² In the case of photo agencies like Fotopersbureau De Boer, the market partly dictated the descriptions and the types of index cards in a collection. These descriptions were made to increase the findability of images for a specific purpose, namely, to sell these images.²³ These descriptions, however, are not necessarily suitable for historical research or to increase findability for users of a historical photo collection. While descriptions of images can guide users to specific images, more generic metadata can help users find sets of images on a specific topic. The latter is helpful when studying diachronic and synchronic expressions of visual culture. Also, models trained on one collection could enrich other collections of images, for which metadata is lacking. This makes it possible to search for images using a predefined list of classes.

Secondly, next to facilitating search, the added metadata can also support quantitative analysis of visual culture. One could, for example, calculate how many people are depicted on images that estimate crowd sizes for historical events. Alternatively, information on form could be of use to study stylistic trends in visual collections.

Thirdly, applying computer vision to historical photographs contributes to the development of computer vision technology that is more sensitive to the historical nature of both the content and medium. Drawing from the expertise of historians and professionals in the heritage sector, we can work towards models that capture the historical dimensions of visual culture.²⁴ In this vein, this article offers a case study that functions as a blueprint for further collaboration and exploration of the possibilities of computer vision of the historical study of visual culture.

The Data: The Fotopersbureau De Boer collection

In 2013, the regional Dutch archive Noord-Hollands Archief acquired the De Boer press photo collection. The collections consist of approximately two million 35mm photo negatives displaying around 250,000 events between 1945 and 2005. In 2020, the archive decided to digitise all photo negatives on high resolution, using a new digitisation method.²⁵ During digitisation, photo negative sheets are placed on a table that can be moved along its x and y-axis. A camera above the table takes a picture each time specific coordinates are reached. This method can scan up to fifteen thousand negatives per day, digitising all 35mm negatives within six months.²⁶

Even though most photos produced by Fotopersbureau De Boer were sold to local newspapers in the Haarlem area, the agency was of national importance, especially in the 1950s, 1960s, and 1970s.²⁷ In 1962, founder Cees de Boer (1918-1985) had his moment of glory, both nationally and internationally, when he won *De Zilveren Camera* and a World Press Photo award in the category 'News' with a photo of Dutch celebrity Ria Kuyken being attacked by a bear (see Figure 1). Winning a World Press Photo award boosted the agency's commissions and economic growth. The number of photos taken by the agency more than doubled in one year.²⁸ At its height, in the 1970s, 1980s, and 1990s, the agency employed five to six photographers in 24/7 shifts, some of whom also won prestigious prizes.²⁹ Fotopersbureau De Boer increasingly became an agency of regional importance in the last decades of its existence, though national newspapers still belonged to its clientele.

The collection's importance lies in its breadth. We find images of major national events, such as the 1953 North Sea flood in the province of Zeeland or the arrival of Martin Luther King in the Netherlands in 1964, as well as pictures of the only two shows The Beatles gave in the Netherlands, in Blokker—a small village just north of Amsterdam—and at the flower fair *Treslong* in Hillegom.³⁰ In addition to well-known images of major national events, numerous photographs depict regional events, including construction sites, sports events, openings of restaurants, dramatic car accidents, and parades. These images of seemingly mundane events can function as historical sources to study everyday life, fashion, or other forms of cultural expression or identity. Moreover, these images are relevant to the public. They offer a glimpse of their lived history through pictures of their street or their relatives and acquaintances throughout the years.

Next to its scope, the collection is unique, at least in the Netherlands, for its extensive metadata accompanying the press photos. The agency meticulously administered their collection in two forms: logbooks (see Figure 2) and subject index cards (see Figure 3). To find specific photos in the collection, Cees de Boer started writing daily logs of photos taken in June 1952.³¹ The administration in the logbooks gives us clear insight into the agency's business and its clients. The secretariat collected the assignments, sorted them by subject and area, and assigned them to photographers working that day. Unlike the world-renowned Magnum agency, photographers at De Boer added the information in the logs themselves.³² After each assignment, they attached a unique number to each roll of film, which they noted in the first column of the log, followed by a description of the event or the name of a person portrayed. The other columns detail to which newspaper or magazine the photos were sent for

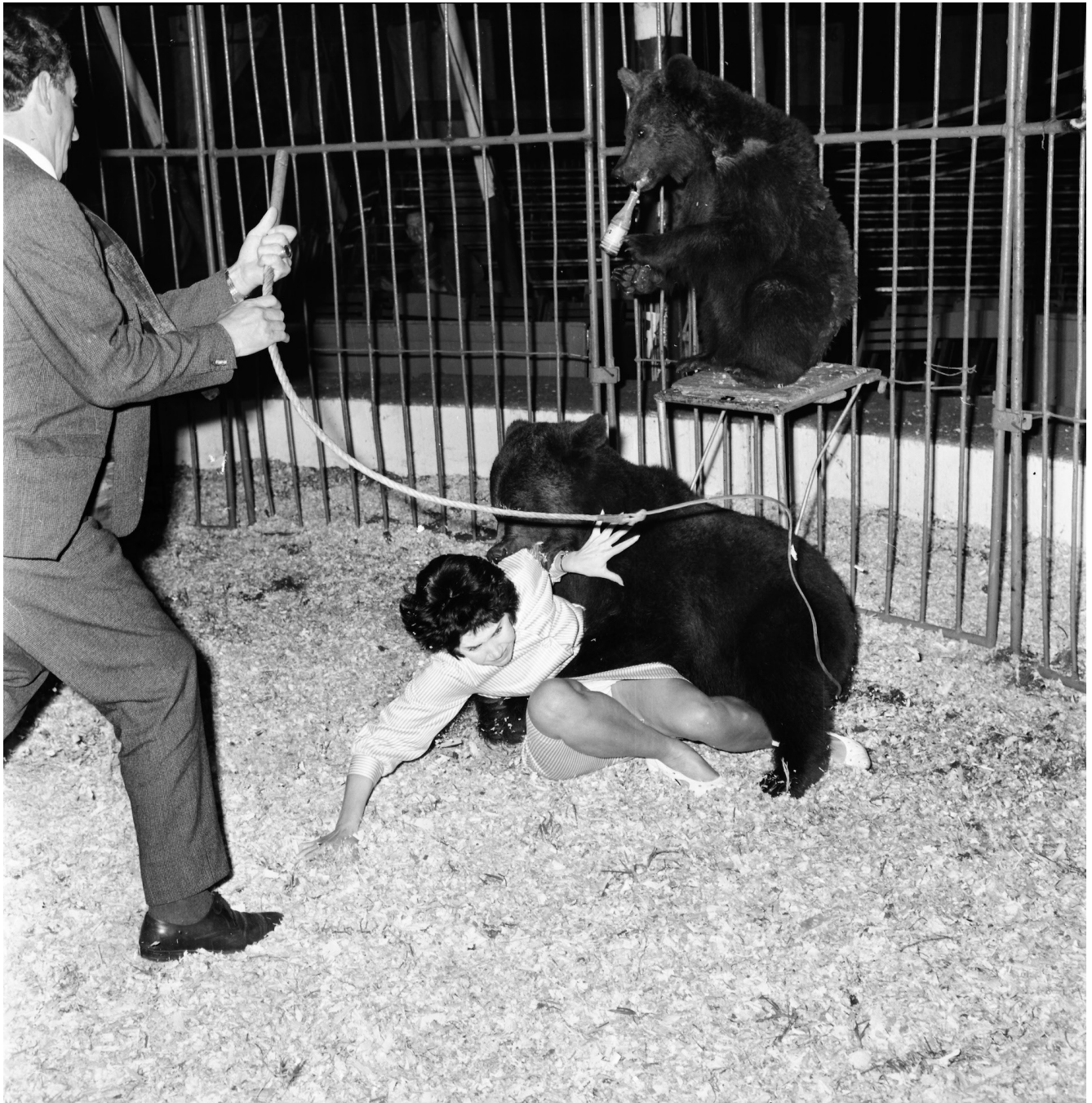


Figure 1. Ria Kuyken attacked by a circus bear (1962). Photo taken from De Boer Collection (collection Fotopersbureau De Boer, 01806A00_08, Noord-Hollands Archief).

publication. Based on these records, we know that the agency supplied photos to all major Dutch newspapers and magazines. The logs show the range of publications that featured the agency's photos and indicate how frequently photos were used. The agency captured between one and five

	1	2	3	4	5	6
	negatief	km + aantal	HD UC	Damiatie	Bijlage	overig
Vervolg Dinsdag 30-10-2001						
ru. dingshuis, H'lem	49918 k			✓		
verkeerscontrole hanting, H'lem	49919 k			✓		
schouw Bulderkes, Vrijhuizen	49920 k			✓		
Koepelverandering, H'lem	49921 k			✓		
MIS Haarlem	49922 k			✓		
Bezoek Haarlem	49923 k			✓		
ANANDA Haarlem	49924 k			✓		
Marinierssteeg, Haarlem	49925 k			✓		
Cobelen, Haarlem	49926 k			✓		
Simon Levett, Haarlem	49927 k			✓		
Heerfutsfeer	49928 k			✓		
Ontkleding Placette 2'voert	49929 k			✓		
XL Stiles Heemstede	49930 k			✓		
JoHeco Computer Heemstede	49931 k			✓		
Bilder Bachtveld H'stede	49932 k			✓		
v Limburg Reisburo H'stede	49933 k			✓		
Russischeborrel H'stede	49934 k			✓		
Dhr/Meur Doers Haarlem	49935 k			✓		
tonaalg. Advendo Jmude	49936 k			✓		
Woensdag 31-10-2001						
Werkveldingse Obelisk H'lem	49937 k			✓		
Mataalverwerking R Steiner, H'lem	49938 k			✓		
Dhr. Beerslof, Vrijhuizen	49939 k			✓		
Asbest, Vrijhuizen	49940 k			✓		
Temo de Nooy en Bildersgilde H'lem	49941 k			✓		
Petra Zuidberg, Halfweg	49942 k			✓		
Argand v. Halen, H'lem	49943 k			✓		
columnarium, H'lem	49944 k			✓		
Vader Tijd Jmude en Rijkensdorp H'lem	49945 k			✓		
Che Chede karpersingel 62A H'lem	49946 k			✓		
V. Dam Chrolatone Heemstede	49947 k			✓		
Nicole de Cock Haarlem	49948 k			✓		
C. Smit & W. Mosterdruis H'lem	49949 k			✓		
v Maale - Oudekerke (oud), 2'voert	49950 k			✓		
P. van Meeren & J. poldermarkt, Dordrecht	49951 k			✓		
Funda met peren, H'lem	49952 k			✓		
kola Ur, H'lem	49953 k			✓		
Vrijdag 2-11-2001						
Honden wachter, Halfweg	49954 k			✓		
Ridsfabriek, Vlaanderenpolder	49955 k			✓		
Berenheer dagverblijf	49956 k			✓		
Gevulde Beethoven H'lem	49957 k			✓		
Melges, H'lem	49958 k			✓		
Sessie Wimp H'lem	49959 k			✓		
zeilboten uit water	49960 k			✓		
gortpap eten Heerhuis	49961 k			✓		
P. Maltmaker, H'lem	49962 k			✓		
Dhr. KAL, H'lem	49963 k			✓		
Dhr. Schmidt, Nieuwenhuys	49964 k			✓		
TAM Geene, Dordrecht	49965 k			✓		
Museum Melano, H'lem	49966 k			✓		
Dhr. van der Vliet, Vrijhuizen	49967 k			✓		
open Kunstlijn - zaal	49968 k			✓		
zolder	49969 k			✓		
Gr. Thakht	49970 k			✓		
tafel in bieb, Vrijhuizen	49971 k			✓		
fest. Jethu college, Dordrecht	49972 k			✓		

Figure 2. The logs of four days in 2001. In this period, we see many photographed events per day and a strong regional focus with almost all photos sold to the local newspapers Haarlems Dagblad and IJmuider Courant. Picture taken by Paul Maessen (Noord-Hollands Archief).

events per day on film in the early years. In 2005, when the company was sold to United Photos, the number of documented events had risen to twenty a day, with approximately three to twenty photographs taken per event. This increase illustrates the expansion of the photo agency and the growing use and visibility of photojournalism in society.³³

LE PAAL		
9192 K	le paal clubhuis THB	8-2-71
9212 K	le paal sporthal Meerwijk	15-2-71
9223 K	le paal Atheneum Heemstede	18-2-71
9244 K	le paal opv. centrum Poelpolder	24-2-71
9273 K	le paal werkpl. de Tunnell	5-3-71
9312 K	le paal school Schalkwijk	18-3-71
9327 K	le paal Nw Vennepe	23-3-71
9347 K	le paal woningen Spaarndam	29-3-71
9370 K	le paal kerk Poelpolder	7-4-71
9370 K	le paal kerk Poelpolder	
9524 K	le paal clubhuis Bennebroek	22-5-71
9529 K	le paal Sterflat	24-5-71
9537 K	le paal Pen centrale	26-5-71
9990 K	le paal Nw Vennepe	21-10-71
10016 K	le paal zwembad Schalkwijk	28-10-71
10111 K	le paal Bernadottelaan	25-11-71
10133 K	le paal Soc. werkplaats Hoofddorp	1-12-71
10158 K	le paal reval. school	8-12-71
10184 K	le paal Raiffeissenbank	15-12-71
10287 K	le paal huizen Noordwijkerhout	14-1-72
10416 K	le paal woningen door weth Ockeloen	28-2-72
10497 K	le paal bejaardenwoningen Halfweg	27-3-72
10561 K / 10560 K	le paal door Prins Bernhard Oegstgeest	17-4-72
10566 K	le paal LTS Heemskerk	20-4-72

Figure 3. One side of the subject index card 'first post'. This refers to pictures of the first post in the ground construction, illustrating that the agency did not just take photographs of the unique and extraordinary but also of the mundane.

A few years after its establishment, the agency started to organise its photos by subject matter, running parallel to the chronological logs. The agency indexed photos retrospectively on subject index cards and sustained the thematic structure until 1990, resulting in about 135,000 entries spread over 1,500 unique subjects. There are many subjects on these cards that take us from 'excavations' to 'soccer', and from 'nudism' to 'silly things'. The number of photographs per category varies considerably. Where, for instance, 'ghosts' has one card, with only five registrations, a subject such as 'persons' has no less than 141 cards with registrations covering it front and back (see Table 1).

Subject	Number of cards
Persons	141
Soccer	74
Buildings: exterior	64
Portraits	60
Openings, commissioning	59
Schools	49
Music	40
Fire, fire brigade	36
Buildings: interior	35
Construction works	35

Table 1. Top 10 topics (out of ca. 1.500) of the subject index cards of Fotopersbureau De Boer (1945-1990), with the number of cards.

Volunteers transcribed all subject index cards (1945-1990) and logbooks of Fotopersbureau De Boer for the period between 1990 and 2004. After digitising the negatives, volunteers also helped link the images to the metadata.³⁴ In December 2021, the complete collection of Fotopersbureau De Boer appeared online.³⁵

The availability of such an extensive collection of digitised negatives spanning sixty years, raises the question how we are to make sense of all these images. The linking of the index cards and logbooks to the digitised material is vital from an archival point of view. It also allows users to search pictures using the agency's metadata. However, the descriptions on these cards are often very particular and regularly refer to aspects that are not expressed visually. Nonetheless, we can combine the information stored in these external sources with other enrichments we make through computer

vision. Requirements expressed by scholars or general users could inform these enrichments. One such enrichment is a label indicating whether a photograph was taken indoors or outdoors.

To determine what the archive users would find of interest, we conducted fourteen interviews with digital humanities experts, heritage experts and visitors online historical archives, focusing on visitors of the Noord-Holland Archief's historical image repository.³⁶ We discussed several computer vision tasks. The respondents showed the most interest in object and scene detection and a task known as location recognition, which uses an algorithm to infer from the visual input where an image was taken. The respondents also showed an interest in using computer vision to link press photos to their publication in newspapers.³⁷ Respondents found the estimation of group sizes, facial emotion detection, detection of logos, and posture estimation less relevant.

Scene Detection Use Case

In this article, we discuss the task of *scene detection* rather than location recognition or object detection. Scene detection is an image classification task that predicts the scene (as a single label) represented in an image.³⁸ This task tries to describe 'the place in which the objects seat (sic),' rather than classifying the object depicted on an image or locating objects in a picture by drawing bounding boxes around them. A scene is described as 'environment(s) in the world, bounded by spaces where a human body would fit.'³⁹ While humans are adept at recognising a scene's functional and categorical properties, they find it difficult to categorise a scene unequivocally.⁴⁰ Human observers have access to contextual factors that help us describe the scene depicted in an image. Computer vision algorithms rely on a model representing visual features of a scene, but the algorithm processes the image in isolation.

Figure 4 shows the top five predicted labels using our scene detection model. We see an image of a married couple at their wedding sitting at a dining table in a large space. The model labels the image of a dining room with an almost 0.9 probability. The labels 'portrait group', 'marriage', and 'ceremony' also applied, while 'fishery' is obviously wrong. This shows the model's predictive power but also makes clear that it is difficult to describe a scene using a single label. In what follows, we will discuss how we constructed this scene detection model and reflect on its uses and shortcomings.

Scene detection, which offers a description of an entire image, is especially relevant for historical photographs since users of heritage collections often search not just to discover representations of individual objects.⁴¹ The ability to search for images based on the scene

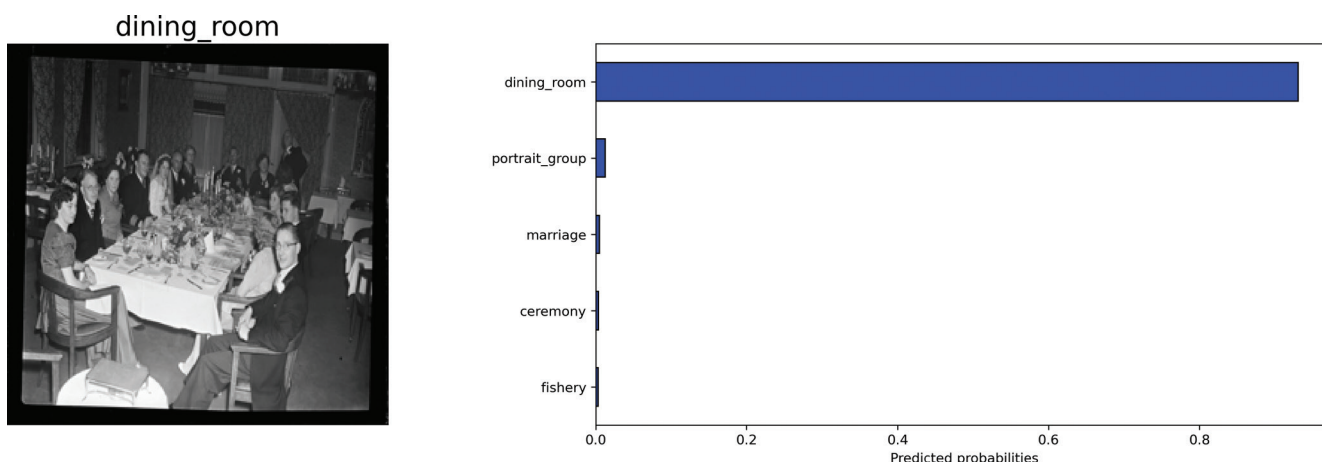


Figure 4. Scene Prediction. The model correctly classified the picture as ‘Dining Room’. The predicted labels ‘portrait_group’ and ‘marriage’ also apply.

represented in them is a valuable feature for heritage institutions, especially since in many photo collections, metadata are sparse, too specific, or not necessarily a description of what is visually expressed.⁴² In the case of press photography, photographs often captured a particular historical event or setting. Pictures of similar events are highly diverse and are generally connected through a series of overlapping similarities, some of which can be expressed visually. Therefore, to capture a scene a more general label is more fitting than a description of individual objects.

We relied on transfer learning to adapt existing computer vision models to the contents of our collection and the needs of our users.⁴³ In transfer learning, we use ‘what has been learned in one setting (...) to improve generalization in another setting.’⁴⁴ Rather than training a model from scratch, we use an existing model to jumpstart detecting scenes in our data using our labels. As a starting point, we used the popular Places-365 model, which can predict 365 different scenes.⁴⁵ As the name implies, this scheme contains 365 categories of scenes, ranging from ‘alcove’ to ‘raft’.⁴⁶ The scenes are very modern and have a clear North American bias, with categories such as rodeo or images of characteristic American kitchens. These scenes are therefore not immediately useful for our data collection. Still, the model has been trained on 1.8 million images and is therefore already sensitive to picking up visual cues that can aid in predicting scenes.

Before the actual annotation, we needed to devise an updated labelling schema. With the assistance of archivists and cultural historians, we removed categories and added additional

categories to the places-365 schema. These categories were more relevant to the contents of the images taken by Fotopersbureau De Boer. The construction of a labelling scheme is always a trade-off between specificity and generalisation. In determining the categories, we kept in mind whether there remained a historical and visual consistency in the categories and whether the category would be of use for users of the collection. Moreover, the specificity also is related to the amount of training material. If we make categories too specific, even though these categories are visually somewhat distinct, we reduce the amount of training data. We must make sure not to reduce the number of images in a category too drastically, as this will severely impact the model's ability to make accurate predictions.

In the case of press photography, we noticed that scenes were either characterised by a particular object ('sculpture'), scenes defined by the location ('park'), or the action ('cycling') performed in the image. Even though scene detection can also learn images with a central object and labelled as such, it might be more fitting to revert to object annotations in these cases. In the latter category, we also find objects, such as trees, flags, automobiles, and bicycles; however, these are not necessarily exclusive to a scene. Even though the presence of particular objects characterised these events, in many cases, they cannot be described by just that single object or person. Moreover, there are instances of sizable inter-class similarity between scenes, for instance, 'library' and 'bookstore'. At the same time, we find significant intra-class variations, in which the depicted scenes that belong to one category are quite diverse, for example, 'theatre'.

Using a subset of approximately 2.5 thousand digitised images in *De Boer* as our pilot dataset, we constructed 115 unique categories.⁴⁷ The distribution of the categories in our training set is heavily skewed and long-tailed (see Figure 5). The categories 'soccer' and 'construction site', for instance, are very well represented, while most others appeared much more infrequently. Even though transfer learning can produce accurate results with few training examples, more training data is preferred, especially in categories with more significant visual variations.

We loaded the pre-trained Places-365 model, which we then tuned to our categories using the deep learning framework Fast.AI.⁴⁸ To account for the unevenly distributed and often small number of training data and to prevent overfitting, we experimented with different data augmentation methods. Overfitting refers to the model adapting too closely to the training data, making it unsuitable for making predictions for images that the model did not see during training. The data augmentations include, among other techniques, the flipping, rotating, and zooming of the source

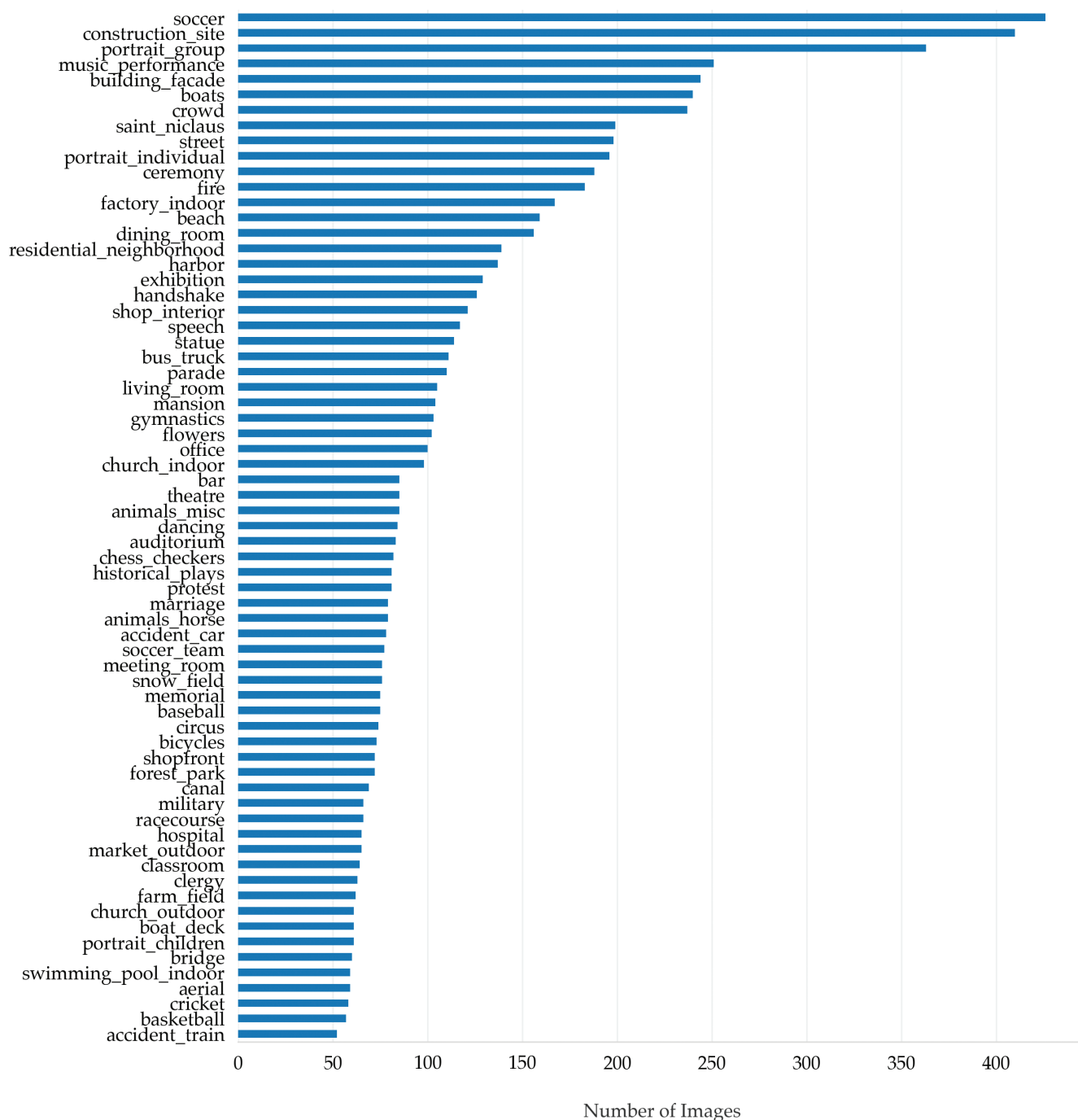


Figure 5. The distribution of scene categories that appeared more than fifty times in the pilot training set. We used this annotated set to re-train the places-365 model.

image. These transformations increase the number of different training images the model sees, which reduces overfitting and improves the model's generalisability.⁴⁹ Training a model always involves

finding a balance between underfitting and overfitting, which is dependent on the size and complexity of the training data. In the cultural heritage domain, we often work with limited sets of annotated training material, making transfer learning a fitting solution. We need to stay aware of overfitting especially when working with these limited data sets. In our use case, we also want the model to predict scenes in unseen data, which requires a model that can generalise, i.e., not only a model that works on Fotopersbureau De Boer.

We checked how well our model performs against annotated data, held out from training. The original places-365 model achieved a 54.74% top-1 accuracy and 85.08% top-5 accuracy on its training material. Top-1 accuracy refers to the percentage of images where the predicted label with the highest probability matches the ground-truth label. Top-5 accuracy calculates whether the ground-truth label is in the top-five predicted labels.⁵⁰ Our model reached a top-1 accuracy of 0.68 and a top-5 accuracy of 0.88. In almost nine out of ten cases, the correct result could be found in the top five results. Notably, our model, which was trained on a relatively small number of images, already achieved good results, which underscores the power of transfer learning and the use of the Places-365 model as a starting point for more specific scene detection tasks.

Errors most frequently occurred in visually diverse categories or images that fit multiple categories.⁵¹ The categories that the model most often confused with each other include: ‘ceremony,’ ‘handshake,’ ‘residential neighbourhood,’ ‘harbour,’ and ‘portrait group.’ The model respectively predicted these categories as: ‘handshake,’ ‘ceremony,’ ‘street,’ ‘boats,’ and ‘ceremony.’ The predictions offered by the model are sensible since they are, for the most part, closely related to the correct labels. For example, the distinction between ‘street’ and ‘residential neighbourhood’ is complex, and it might be more appropriate to attach both labels to the image. This result leads us to conclude that it might be better to annotate images using multiple labels, creating a model that offers multi-label predictions.

The ability to search for specific scenes in a collection such as Fotopersbureau De Boer helps scholars to answer questions related to photography in general or press photography more specifically on a local, regional, and national scale. Rather than locating specific photographs, we can find diachronic and synchronic series of photographs of a particular topic. Researchers can further examine a subset using qualitative methods, but it can also serve as a more homogenous input for other types of visual analysis. Using the information on the depicted scenes, scholars could examine the representation of scenes, such as protests, at scale.

Concluding Remarks

This article demonstrated how computer vision can be applied to historical photo collections, such as the 2,000,000 digitised press photos of Fotopersbureau De Boer (1945-2005) using transfer learning. We used the specific example of scene detection to show how we already reached high accuracy with a limited amount of training data. The development of the annotation schema proved to be one of the more difficult steps in the process. For future work, we recommend a concerted effort by heritage institutes in developing an annotation framework for historical images, like ICONCLASS for art historical material.⁵² Using computer vision, we can add metadata to visual collections, which serve three concrete purposes. Firstly, these metadata enable search that can benefit multiple types of archive users. Secondly, scholars can use enriched collections to study large-scale trends in visual collections. Questions could include, how fashion trends changed over time, when certain brands, such as Coca-Cola, appeared in Dutch streets, or how funeral practices have changed over the years. Thirdly, the development of annotated datasets could benefit the computer vision community. Researchers in this field can develop models using these datasets, which are better attuned to the visual representation in the past.

Currently, we are improving our scene detection model using a combination of computer vision predictions and crowdsourcing. Using the model presented in this article, we make predictions, which are checked through crowdsourcing.⁵³ When annotators indicate that labels are missing, we add those labels to our training data. Moreover, checked predictions will be added to the training data, increasing the amount of training data, and consequently, the model's accuracy. We have made the code, training data, and resulting model publicly available. After completing our crowdsourcing project, we will also make this training data and model available.⁵⁴ Other institutes could then apply this model to their collection or adjust it using transfer learning to their labelling schema.

One final insight this project has offered is that while computer vision models have improved enormously over the last years, the major bottleneck that remains is the lack of annotated historical visual material. Even though applying transfer learning to off-the-shelf models requires relatively small amounts of training data, we still need to manually annotate additional training data using updated labelling schema that require input from domain experts. We hope to offer a blueprint for other heritage institutes and, more specifically, other historical photo collections with this project. This project has been supported by external funding and uses the crowdsourcing platform VeleHanden.⁵⁵ Since these resources might be out of reach for many groups and institutions, we hope

that his blueprint can function as an invitation for collaboration through which training data, best practices, annotation schemes, and models are shared and improved. In addition to reducing costs, such a concerted effort improves the diversity and quality of the labels and training data, which ultimately leads to better computer vision models for historical research.

Acknowledgements

This research is supported by the Mondriaan Fonds (105039500) and the Prins Bernhard Cultuurfonds (40033061/RDB/MIL).

Author Contribution

Melvin Wevers: first draft, manuscript revisions, data preparation, programming and training algorithms, running experiments, analysing data.

Nico Vriend: first draft, manuscript revisions, data preparations, survey.

Alexander de Bruin: first draft on De Boer historical context, personal correspondence with Nico Vriend on De Boer.

Notes

1. Peter Burke, *Eyewitnessing: The Uses of Images as Historical Evidence*, 1st edition (Ithaca: Cornell University Press, 2004).
2. Mark Sealy, "Decolonizing the Camera: Photography in Racial Time" (Doctoral, Durham, Durham University, 2016), <http://etheses.dur.ac.uk/11794/>; Clément Chéroux, *The Perfect Medium: Photography and the Occult* (New Haven: Yale University Press, 2005); Rob Kroes, *Photographic Memories: Private Pictures, Public Images, and American History*, Interfaces : Studies in Visual Culture (Hanover: Dartmouth College Press, 2007); Ellen Tops, "Foto's met gezag: een semiotisch perspectief op priesterbeelden 1930-1990," KDC bronnen & studies 39 (Nijmegen, Vantilt, 2001); Albert van der Zeijden, *Visuele cultuur: fotografie als historische bron en als medium voor etnologisch onderzoek*, Cahiers voor regionale geschiedenis en volkscultuur 3 (Utrecht: Nederlands Centrum voor Volkscultuur, 2004).

3. Elizabeth Edwards and Janice Hart, *Photographs Objects Histories: On the Materiality of Images*, Material Cultures (London: Routledge, 2004).
4. Martin Lister and Liz Wells, "Seeing beyond Belief: Cultural Studies as an Approach to Analysing the Visual," in *Handbook of Visual Analysis*, eds. Theo Van Leeuwen and Carey Jewitt (London: SAGE, 2001); Mary Panzer, *Things as They Are* (London: Boot, 2005).
5. Hanno Hardt, "Sites of Reality: Constructing Press Photography in Weimar Germany, 1928–33," *The Communication Review* 1, no. 3 (January 1, 1996): 373–402, <https://doi.org/10.1080/10714429609388270>; Thierry Gervais and Gaëlle Morel, *The Making of Visual News: A History of Photography in the Press* (London: Routledge, 2017); Panzer, *Things as They Are*.; Nadya Bair, *The Decisive Network: Magnum Photos and the Postwar Image Market* (Oakland: University of California Press, 2020).
6. Lev Manovich, *Cultural Analytics* (Cambridge, MA, USA: MIT Press, 2020).
7. Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *The American Historical Review* 121, no. 2 (April 1, 2016): 377–402, <https://doi.org/10.1093/ahr/121.2.377>; Jo Guldi, "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora," *Journal of Cultural Analytics*, December 20, 2018.
8. Franco Moretti, "Conjectures on World Literature," *New Left Review*, no. 1 (February 2000): 54–68.
9. Moretti, "Conjectures on World Literature," 54–68.; Taylor Arnold and Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora," *Digital Scholarship in the Humanities* 34, no. Supplement_1 (December 1, 2019): 13–16, <https://doi.org/10.1093/llc/fqz013>.
10. Melvin Wevers and Thomas Smits, "The Visual Digital Turn: Using Neural Networks to Study Historical Images," *Digital Scholarship in the Humanities* 35, no. 1 (April 1, 2020): 194–207, <https://doi.org/10.1093/llc/fqy085>; Eef Masson et al., "Exploring Digitised Moving Image Collections: The SEMIA Project, Visual Analysis and the Turn to Abstraction," *DHQ: Digital Humanities Quarterly* 14, no. 4 (2020); Samarth Bhargav, Nanne van Noord, and Jaap Kamps, "Deep Learning as a Tool for Early Cinema Analysis," in *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia HeritAge Contents*, SUMAC '19 (Nice, France: Association for Computing Machinery, 2019), 61–68, <https://doi.org/10.1145/3347317.3357240>.; Benjamin Lee, "Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset," *Digital Humanities Quarterly* 15, no. 4 (2020).; Matthew Lincoln et al., "CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative," October 9, 2020, <https://doi.org/10.1184/R1/12791807.v2>.

11. Lincoln et al., “CAMPI.”; Wevers and Smits, “The Visual Digital Turn.”; Isabella di Lenardo, Benoit Seguin, and Frédéric Kaplan, “Visual Patterns Discovery in Large Databases of Paintings,” in *Proceedings of ADHO Digital Humanities Conference* (Digital Humanities 2016, Krakow, Poland, 2016); Paul Fyfe and Qian Ge, “Image Analytics and the Nineteenth-Century Illustrated Newspaper,” *Journal of Cultural Analytics* 1, no. 2 (October 15, 2018): 11032, <https://doi.org/10.22148/16.026>.
12. Di Lenardo, Seguin, and Kaplan, “Visual Patterns Discovery in Large Databases of Paintings;” “Yale DHLab - Every Pixel on the Sunset Strip: Machine Vision and Photographs,” accessed February 8, 2022, <https://dhlabs.yale.edu/projects/every-pixel/>; Masson et al., “Exploring Digitised Moving Image Collections.”
13. Melvin Wevers and Thomas Smits, “Detecting Faces, Visual Medium Types, and Gender in Historical Advertisements, 1950–1995,” in *European Conference on Computer Vision* (Springer, 2020), 77–91; Prathmesh Madhu et al., “Understanding Compositional Structures in Art Historical Images Using Pose and Gaze Priors,” in *European Conference on Computer Vision* (Springer, 2020), 109–25.
14. Erving Goffman, *Gender Advertisements* (London: Macmillan, 1985).; Mee-Eun Kang, “The Portrayal of Women’s Images in Magazine Advertisements: Goffman’s Gender Analysis Revisited,” *Sex Roles* 37, no. 11 (December 1, 1997): 979, <https://doi.org/10.1007/BF02936350>.; Philip Bell and Marko Milic, “Goffman’s Gender Advertisements Revisited: Combining Content Analysis with Semiotic Analysis,” *Visual Communication* 1, no. 2 (June 1, 2002): 203–22, <https://doi.org/10.1177/147035720200100205>.
15. “Detecting Faces, Visual Medium Types, and Gender in Historical Advertisements, 1950–1995 | SpringerLink,” accessed February 23, 2021, https://link.springer.com/chapter/10.1007%2F978-3-030-66096-3_7.; Leonardo Impett and Franco Moretti, “Totentanz: Operationalizing Aby Warburg’s Pathosformeln” (Stanford Literary Lab, 2017).; Prathmesh Madhu et al., “Understanding Compositional Structures in Art Historical Images Using Pose and Gaze Priors,” in *European Conference on Computer Vision* (Springer, 2020), 109–25.
16. Wevers and Smits, “The Visual Digital Turn.”; Taylor Arnold and Lauren Tilton, “Distant Viewing: Analyzing Large Visual Corpora,” *Digital Scholarship in the Humanities* 34, no. Supplement_1 (December 1, 2019): 13–16, <https://doi.org/10.1093/llc/fqz013>.
17. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012, 1097–1105.
18. Ian Goodfellow et al., *Deep Learning*, vol. 1, 2 (Cambridge, MA: MIT Press, 2016).

19. Ferdinand Maiwald, "Generation of a Benchmark Dataset Using Historical Photographs for an Automated Evaluation of Different Feature Matching Methods," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2019, 87–94.
20. Ferdinand Maiwald et al., "Photogrammetric Analysis of Historical Image Repositories for Virtual Reconstruction in the Field of Digital Humanities," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2017): 447.
21. Jose Llamas et al., "Classification of Architectural Heritage Images Using Deep Learning Techniques," *Applied Sciences* 7, no. 10 (October 2017): 992, <https://doi.org/10.3390/app7100992>.
22. The Oxford Visual Geometry Group also offers many projects working on the application of Computer Vision to historical images. <https://www.robots.ox.ac.uk/~vgg/research/>.
23. Estelle Blaschke, "From the Picture Archive to the Image Bank," *Études Photographiques*, no. 24 (November 9, 2009), <https://journals.openedition.org/etudesphotographiques/3435>.
24. Eun Seo Jo and Timnit Gebru, "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning," *ArXiv:1912.10389 [Cs]*, December 22, 2019, <https://doi.org/10.1145/3351095.3372829>.
25. Nico Vriend, "Hoe digitaliseer je 2.000.000 fotonegatieven? - Noord-Hollands Archief," NHALab, accessed March 12, 2021, <https://noord-hollandsarchief.nl/ontdekken/nhalab/835-digitalisering-de-boer>.
26. The Dutch company LOTS Imaging invented the X|Y scanning tables. The remaining 6,000 glass negatives, 65,000 6x6 roll films, and 40,000 sheet films have also been digitised. In the process, the Noord-Hollands Archief was advised by photographic material expert and historian Michiel Kort, and the late Onno Zaman, then CEO of digitisation firm Picturae.
27. The agency's national importance consolidated when Cees de Boer became a founding partner of United Photos (1951–1975) together with Ben van Meerendonk (Amsterdam), Tiemen van der Reijken (Delft), and Joep Friezer (The Hague).
28. Cees' son Poppe stated, although the firm was an economic success, the owners enjoyed life as well: 'Money came in with trucks but left in container ships.' Oral History project Noord-Hollands Archief, interviews with former photographers of Fotopersbureau De Boer, 2016: Poppe de Boer, Marisa Beretta, Dorethe Biersteker (secretary), Olaf Kraak, Frans van der Linde, Ron Pichel, Koen Suyk and Paul Vreeker.
29. Poppe de Boer was awarded 'Best Dutch Photojournalist of the Year' in 1969. Frans van der Linde won first prize in the 'Sports' category of *De Zilveren Camera*, 1974, and Marisa Beretta was awarded the second prize in the 'Arts' category of *De Zilveren Camera*, 2002. After their careers at Fotopersbureau De Boer, photographers like Olaf Kraak, Jerry Lampen and Paul Vreeker also received awards.

30. Alexander de Bruin, “De Collectie van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2013), 211–15; Alexander de Bruin, “The Fast and the Furious. Sportfotografie van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2014), 180–86; Alexander de Bruin, “Herstelwerkzaamheden in de Haven van IJmuiden Na de Tweede Wereldoorlog. Opnames van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2015), 173–80; Alexander de Bruin, “Brand in Haarlem. Opnames van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2016), 169–76; Alexander de Bruin, “De Motor van Bomans. Opnames van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2017), 166–74; Alexander de Bruin, “Haarlemse Schrijversportretten. Opnames van Fotopersbureau De Boer,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2018), 178–86; Alexander de Bruin, “Fotografen van de Bevrijding van Haarlem. Cees de Boer En Flip Delemarre,” in *Jaarboek Haerlem* (Haarlem: Historische Vereniging Haerlem, 2019), 170–76.
31. Personal correspondence with Poppe de Boer.
32. Bair, *The Decisive Network*, 200–207.
33. Thomas Smits, “Een Explosie van Beeld, 1840–1930,” in *De Krant*, eds. Huub Wijfjes and Frank Harbers (Amsterdam: Boom, 2019), 165–204.
34. Nico Vriend, “Fotografisch Geheugen: De Reportages van Fotopersbureau De Boer,” Vele Handen, December 3, 2021, https://velehanden.nl/projecten/bekijk/details/project/ranh_tagselection_deboer.
35. <https://noord-hollandsarchief.nl/beelden/beeldbankdeboer/>.
36. Due to the COVID-19 pandemic, we could only use video calling to interview a small number of people.
37. Together with other parties, the Noord-Hollands Archief experiments with this task in the project *Krant en Foto's*: <https://krant-en-fotos.nl/>.
38. Bolei Zhou et al., “Places: A 10 Million Image Database for Scene Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, no. 6 (June 1, 2018): 1452–64, <https://doi.org/10.1109/TPAMI.2017.2723009>.
39. Zhou, “Places,” 1452–64.
40. Aude Oliva and Antonio Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope,” *International Journal of Computer Vision* 42, no. 3 (2001): 145–75.
41. Vivien Petras et al., “Europeana– a Search Engine for Digitised Cultural Heritage Material,” *Datenbank-Spektrum* 17, no. 1 (2017): 41–46.; Paul Clough et al., “Europeana: What Users Search for and Why,” in *Research and Advanced Technology for Digital Libraries*, ed. Jaap Kamps et al., *Lecture Notes in Computer Science* (Cham: Springer International Publishing, 2017), 207–19, https://doi.org/10.1007/978-3-319-67008-9_17.

42. While the collection of De Boer has elaborate descriptions, these are often too detailed and not directly linked to the represented scene. Nonetheless, these descriptions are useful to guide the annotation process of the images.
43. Waseem Rawat and Zenghui Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” *Neural Computation* 29, no. 9 (June 9, 2017): 2352–2449, https://doi.org/10.1162/neco_a_00990.
44. Goodfellow, *Deep Learning*, 526.
45. Zhou, “Places.”
46. This paper describes the experiment in more detail, Melvin Wevers, “Scene Detection in De Boer Historical Photo Collection,” in *Proceedings of the 13th International Conference on Agents and Artificial Intelligence* (ICAART, Vienna, 2021). Given that our data set consists primarily of black and white images, it is worth noting that this type of image was excluded from the places-365 data set.
47. These images were made available on the archive’s image repository prior to the large-scale digitisation of the entire collection and show the breadth of the collection.
48. <http://www.fast.ai>.
49. Christian Szegedy et al., “Going Deeper with Convolutions,” *ArXiv:1409.4842 [Cs]*, September 16, 2014, <http://arxiv.org/abs/1409.4842>.
50. Because of the ambiguity between scene categories, a top-5 accuracy is generally seen as a more suitable criterion than a top-1 accuracy.
51. For a more extensive error analysis see: Wevers, “Scene Detection.”
52. “Home — Iconclass,” accessed March 12, 2021, <http://www.iconclass.nl/>.
53. https://velehanden.nl/projecten/bekijk/details/project/ranh_tagcorrection_deboer.
54. Melvin Wevers, (2020). HisVis: Scene Detection Pilot Training Set [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3888006>; Code can be found here: <https://github.com/melvinwevers/hisvis>.
55. <https://velehanden.nl/>.

Biographies

Melvin Wevers is an Assistant Professor of Digital Methods at the University of Amsterdam. His research interests include the study of cultural–historical phenomena using computational means with a specific interest in the formation and evolution of ideas and concepts in public discourse.

Nico Vriend studied history at the Vrije Universiteit Amsterdam and archival studies at the University of Leiden. He leads the project *Fotografisch Geheugen* (2020-2022) at the Noord-Hollands Archief in Haarlem, the Netherlands. The project's aim is to digitise and publish the complete collection of photo press agency *Fotopersbureau De Boer* (1945-2005). Computer Vision tasks are applied to enhance the accessibility of the 2 million press photos of this collection.

Alexander de Bruin, curator of the image collections at the Noord-Hollands Archief, studied art history at the Vrije Universiteit Amsterdam and archival science at the University of Amsterdam. In 2012/2013 he was closely involved in the acquisition of the photo archive of photo press agency *Fotopersbureau De Boer* and co-responsible for all the plans and execution regarding material care, digitisation and accessibility afterwards. From 2014 onward, he annually writes about one of the many photographic subjects from the De Boer collection in the *Jaarboek Haarlem*.

TMG Journal for Media History

Volume 25 No (1)/2022

DOI

<https://dx.doi.org/10.18146/tmg.815>

PUBLISHER

Netherlands Institute for Sound & Vision

COPYRIGHT

Each article is copyrighted © by its author(s) and is published under license from the author(s). When a paper is accepted for publication, authors will be requested to agree with the Creative Commons Attribution 4.0 International License.