Automatic Detection and Visualization of Garment Color in Western Portrait Paintings

Cihan Sarı[1], Albert Ali Salah[2,3], and Alkım Almıla Akdag Salah[3,4]

[1] Boğaziçi University, Systems and Control Engineering,
cihan.sari@boun.edu.tr

[2] Boğaziçi University, Computer Engineering,

[3] Utrecht University, Information and Computing Sciences
a.a.salah@uu.nl

[4] Istanbul Şehir University, College of Communications
a.a.akdag@uu.nl

Abstract: Paintings give us important clues about how males and females were perceived over centuries in the Western culture. In this paper, we describe a system that allows scholars to automatically visualize how the clothing colors of male and female subjects changed over time. Our system analyzes a large database of paintings, locates portraits, automatically classifies each portrait's subject as either male or female, segments the clothing areas and finds their dominant color. An interactive, web-based visualization is proposed to allow further exploration of the results. To test the accuracy of our system, we manually annotate a portion of the Rijksmuseum collection, and use state-of-the-art image processing and computer vision algorithms to process the paintings. We use a deep neural network based style transfer approach to improve gender recognition (or more correctly, sex recognition) of the sitters of portraits. The annotations and the code of the approach are made available.

1. Introduction

The history of color is full of instances of how and why certain colors became associated with certain concepts, ideas, political views, connotations of status and power. Sometimes these occurred arbitrarily, such as the instance when pink was assigned to baby girls, and blue started to be associated with baby boys at the turn of 19th Century (Paoletti, 1987). Sometimes, color associations had very tangible reasons, such as in the case of Marian blue. Because of the scarcity and great value of lapis lazuli, from which the blue pigments were extracted, this color was reserved only for painting Virgin Mary for centuries. Individual colors have convoluted and contested histories, since they have been attached to many symbols at any given time. John Gage, an art historian who has devoted 30 years of research on the topic of color, explains the conundrum of what he terms as "politics of color" in a simple fashion: "The same colors, or combinations of colors can, for example, be shown to have held quite antithetical connotations in different periods and cultures, and even at the same time and in the same place." (Gage, 1990).

The purpose of the present study is to introduce a system and an approach for automatically extracting the main color of clothing worn by people in Western artworks. The study is part of a larger research program to translate Laqueur's (1990) thesis on how sex is understood in the Western culture.

Laqueur, a well-known historian and sexologist, argued that sex, like gender, is a constructed concept, at least in the Western world. This argument is based on his meticulous study of Western Medicine from Ancient Greeks to the 20th Century. According to Laqueur, there is a one-sex model, in which the woman and the man are described to be essentially the same. Sexual differences, in this model, are perceived along a spectrum, with women at one end of, and men at the other end. Around the 19th Century, according to Laqueur this understanding shifted to a newer model, which Laqueur terms as the two-sex model, where man and woman were perceived as distinctly different, in fact as two separate entities. As evidence, Laqueur explored medical imagery, but extended his claims from medical discourse and scientific treatises to examples drawn from social and cultural norms and beliefs. Today, with the help of algorithms and the plethora of digital archives from earlier centuries, it is possible to test Laqueurian theory in order to provide additional evidence of this shift from same-sex to two-sex in visual culture more broadly.

Why do we choose color to trace the visual reflections of Laqueur's theory? In order to properly analyze all the visual material at hand via image analysis, we need algorithms that relate the visual to its context. Current image retrieval and analysis algorithms are not at that stage of development yet. It is possible to search through millions of images for a certain concept or object if the algorithms are specifically trained for it. However, when it comes to 'see' what the object stands for, the image analysis fails to deliver. What we have at hand are algorithms that can easily detect low-level features of images ("low-level" translates to edges, color, contrast, luminosity, saturation, etc.). Among these features, color is the foremost when it comes to "bearing" connotations on its own.

The portrait paintings in Western Culture were undoubtedly associated with social status, wealth and power. With the rise of petit bourgeois to power, not only the nobility, but also the commoners could afford portraits as a way to leave a trace of themselves behind. Even then, the resulting artworks were not only simple ways of capturing immortality, but they served as reflections of the sitters' position in society. From the chosen garb, to the ornaments worn and objects depicted in the portrait, every detail had a distinct function. Among the possible objects to detect and dissect (ornaments, swords, regalia, books, etc.), garments of sitters were chosen because (1) they are easier to detect automatically (2) they usually contain rich color information and (3) after the sitters' faces, they contain the most interesting visual information, and hence capture the attention of the audience.

Visualizing clothing colors over time for cross-referencing with historical data can reveal changes in how particular colors were used in a given time period and culture. For performing such an analysis over large painting collections, automatic image analysis and visualization tools are necessary. While computer vision research has progressed very rapidly in recent years in processing photographs and videos, finding people and objects in paintings is much more difficult, as there is a great variety of styles and less annotated data for training classifiers. In this study, we use and combine several computer vision approaches for the difficult problem of extracting the dominant clothing colors in thousands of paintings automatically. Such a tool will allow the digital humanities scholars to observe color trends in paintings over long periods of time, and test hypotheses about colors of clothing and their changes. We make a case study of clothing worn by women and men, and give an illustrative example visualization to contrast the two. We also explore whether it is algorithmically possible to automatically determine the perceived sex of the sitter of portraits using current computer vision technology, to find women and men in paintings without manual intervention. We develop a simple and intuitive interactive information visualization interface to depict the results of the image processing algorithms, and illustrate the results on a dataset assembled from the Rijksmuseum collection.

The flow of our proposed model is illustrated in Fig. 1. The upper part describes the training of a classifier to decide whether the painting contains a male or a female subject: the portrait sitter. The lower part shows how, for a given painting, the dominant clothing color and the perceived sex are determined, and the result is added to the interactive visualization. Each colored box in this visualization represents one painting, and if the user clicks on one of these, the actual painting and its basic metadata (the title and the name of the artist) is visualized in a small window.

2. **Background**

Within computer vision and multimedia retrieval, computer based analysis of artworks is receiving increasing attention (Spratt and Elgammal, 2014; Sartori et al., 2015; Işıkdoğan et al., 2016). The research focused on creating automatic programs that, given an artwork, can identify the artist, the style, or the production date, as well as on searching for stylistically similar artworks in a collection (Stork, 2009). While some of this research followed reductionist perspectives and was heavily criticized for losing sight of critical content, the fact remains that computer vision can provide art historians with tools that can be used in locating visual materials with certain aspects in large collections. For instance Crowley and Zisserman's (2014) recent retrieval system allows one to search for simple concepts (e.g. “train”, “dog”) in painting databases, without requiring human metadata annotations for these concepts. Thus, it becomes possible to retrieve and visualize paintings of a particular period that show a certain visual quality, or contain a certain object or feature.

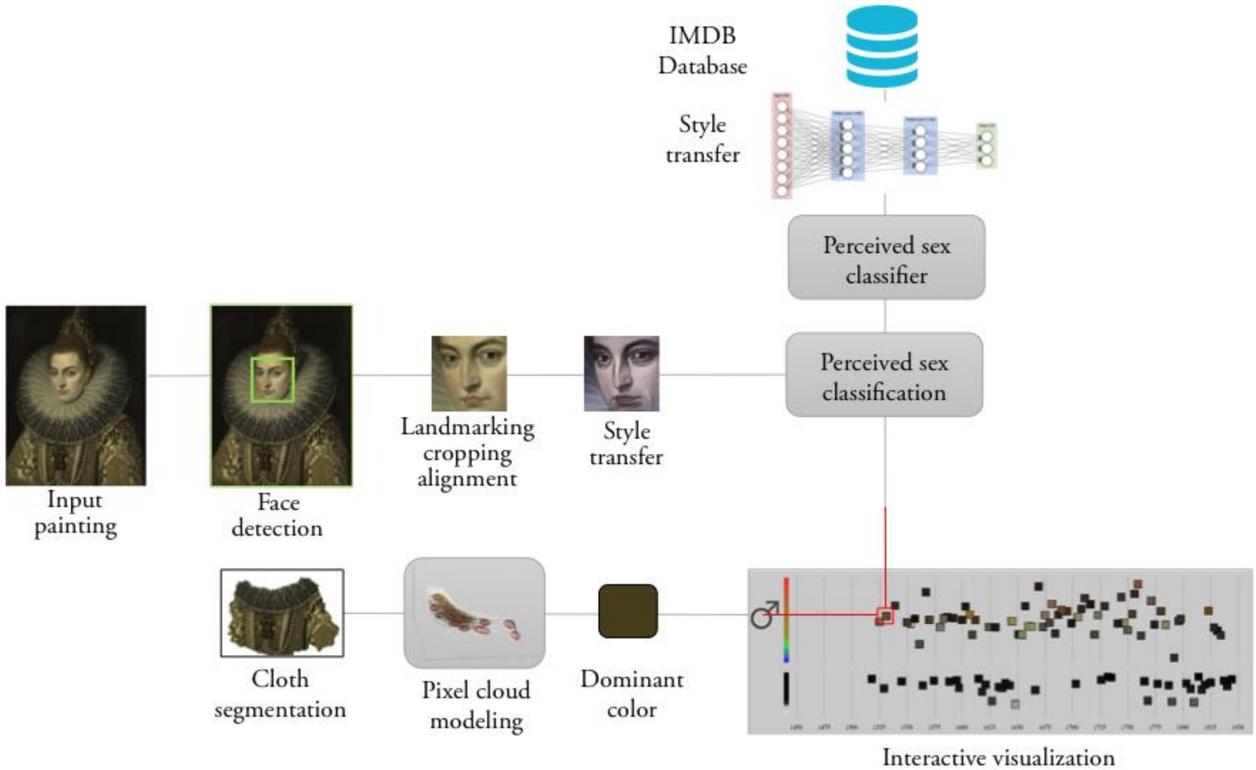

Figure 1. The flow of the proposed approach. The upper part depicts the offline training, and the lower part is the online processing and visualization of a single painting

Today, it is possible to automatically detect faces in paintings, or to attribute style to paintings via various statistical techniques (Kim et al., 2014). With deep neural networks, we can transform the style of a painting to another style, within certain limitations (Gatys et al., 2016). For all these recent developments, the importance of color is apparent. In this study we apply various available image analysis algorithms to observe how the depiction of women and men has changed over time in Western art. If Laqueur's theory about the shift to two-sex theory is valid, we would expect portraits of male and female subjects to become increasingly different after the turn of the 19th century. To test whether portraits and clothing colors provide supporting evidence, we use a cultural analytics approach (Manovich, 2016), where hundreds of thousands of paintings can be automatically processed to find out whether the sitter is a man or a woman, to determine what the dominant clothing color is, and to visualize findings in an interactive interface that allows further exploration. This tool makes it possible to investigate ideas and claims regarding women's and men's clothing in large painting collections.

3. Methodology

In this section, we describe the individual components that make up the system, starting with the database. We first describe the automatic classification of the portrait sitter into male and female classes. Then we describe how a representative color is selected and visualized for the clothing of the sitter.

3.1 Database

Online cultural heritage collections are usually enriched with meticulously tagged metadata about the origins of each artwork. However, such metadata do not necessarily contain the sex of the sitter, or the colors of clothing worn by the sitter. To test our system, we have chosen the Rijksmuseum's database published as part of a challenge, with 112,039 high-resolution images containing extended metadata (Mensink and van Gemert, 2014). The Rijksmuseum is a Dutch national museum dedicated to arts and history in Amsterdam. 650,340 digitized works of art were available from the museum collection at the time of this study [Note 1]. Throughout the digitization process, we added our annotations to the records when possible, using structured vocabularies (Dijkshoorn, 2014). The images from the database were saved at 300 dpi quality, with file sizes ranging from 2 to 5 megabytes in JPEG format, with a corresponding XML file that contains the available metadata. Information in the metadata can include title, dates, painter information, if known, in Flemish (see Tab. 1), but without any reference to what these artworks contain in terms of people and objects (Mensink and van Gemert, 2014). We need additional algorithmic tools to determine these, which we attempt to provide in this paper.

filename	format	date	type	title	subject	contributor	relation
009071	brons	1475	beeldhouwwerk	De gerechtigheid...			
008450	brons	1476	graffiguur	Pseudo Philips van...	41D2	Bruikleen...	BK-AM-33
008550	polychromie	1480	beeldhouwwerk	Bewening van...	73D7211 (+5)	Aankoop...	
104894	boekdruk	1706	inhoudsopgave	Byvoegsel tot de...			RP-P-OB-83.133
091654	polychromie	1735	schaal (objectnaam)	Hartvormige schaal...	25G3	Schenking...	
060116	etsen	1764	prent	Portret van P. Born	31A5331		
069114	hout	1780	werfmodel	Halfmodel van een...			

Table 1. Sample from Rijksmuseum metadata

3.2 Classification of perceived sex

This process is commonly called "gender recognition" in computer vision and pattern recognition literature (Cottrell and Metcalfe, 1991). However, what is meant by this term seldom has anything to do with what "gender" means today. "Perceived sex recognition" would describe more accurately what is being done, namely, to determine the apparent sex of the person from the facial image. Several publications have appeared in recent years with the aim of automatic recognition of sex from faces. The survey by Ng et al. (2012) describes a variety of approaches on sex recognition in natural images. We have followed a multi-step approach for estimation of perceived sex, including dataset preparation, face detection and registration, feature extraction, and classification into male/female classes, each detailed in a subsection.

3.2.1 Face detection and registration

Automatic processing of people starts with detecting them in paintings, and the detection of faces is the first step for this purpose. Face detection and analysis tools are well-developed and suitable for paintings. There have been focused studies to address face recognition tasks on artistic images (Srinivasan et al., 2015). For the purposes of face detection, mainstream algorithms perform sufficiently well on paintings that are of interest for this study. In order to separate the portraits within the Rijksmuseum dataset from other artifacts, we used a face detection algorithm (Viola and Jones, 2001). This method uses a series of very fast comparisons to enable rapid and robust face detection. We do not detail this approach here, as it is very well known, and available as a function in the OpenCV computer vision library [Note 2].

Once the portraits are determined, and the faces of the sitters are localized, we use a face-based approach to classify the sex of the sitter. This is not a completely solved problem, but deep learning based classifiers have been shown to reach over 93% accuracy on photographs (Ranjan et al., 2017). However, in order to improve the processing of the faces, we first align them with a general model. It is known that good alignment, also called registration, is a key for robust facial processing (Salah et al., 2007).

Registration is the determination of a geometrical transformation that aligns points in one picture with the corresponding points in another picture. The anchor points used in facial registration are called “landmarks". These are typically points like mouth and eye corners, the tip of the nose, and points regularly spaced along the boundary of the face. Once such landmarks are located in an image, the image can be transformed, rigidly or non-rigidly, to a target shape, represented as a set of landmark points.

Given a set of shapes, it is possible to generate a mean shape that will serve as alignment target automatically, via Generalized Procrustes Analysis (GPA) (Gower, 1975). Furthermore, if it is used on the training set (i.e. the set of samples used for developing the algorithms) to extract this mean shape, GPA also produces the alignment of each training sample. Further samples can similarly be aligned to the mean shape, which is simply a set of 2D points. After this step, we can classify portraits into male and female classes.

3.2.2 Dataset preparation with style transfer

Training an automatic supervised classifier requires images with male/female labels, to serve as examples for the algorithm. We used a custom dataset we created by crawling 5,603 actor and 5,262 actress images from the Internet Movie Database (IMDb) by using known artist names. Face detection is performed on these images, followed by a manual correction to eliminate faces with low resolution and incorrect detections. Our system is trained with this database, and adding other data resources, such as 10k US Adult Faces (Bainbridge et al., 2013), and Labeled Faces in the Wild (LFW) (Huang et al., 2007) datasets, did not improve results. Both LFW and 10K US Adult Faces databases contain samples from natural poses, pictures taken in daily life and similar causal conditions. However, most IMDb images were deliberately posed. In our experiments, we discovered that using IMDb images alone for training gives the best result, as portrait paintings and IMDb pictures are more similar in terms of pose and expression. Guessing sex directly from painting names is not a reliable approach, but it can help in cases where the title is very straightforward.

Once the images are prepared, we can train a classifier to separate males and females. However, our labeled samples are photographs, and our target domain is paintings. In order to implement a system that would perform better on paintings, we use an approach called "style transfer" to convert our photographs into painting-like images. Gatys et al. published a deep neural network model that achieves style transfer for a number of paintings selected as target styles (Gatys et al., 2016). We have tested a number of images, and empirically selected one (*Giorgio de Chirico, Horses on the seashore)* as the target style. This style created marked facial contours, and provided a good common space where images can be matched. Examples of style transfer can be seen on Fig. 2. The matching in the sex classifier is done between the style transferred versions of both IMDb faces and Rijksmuseum faces, which now look more comparable.

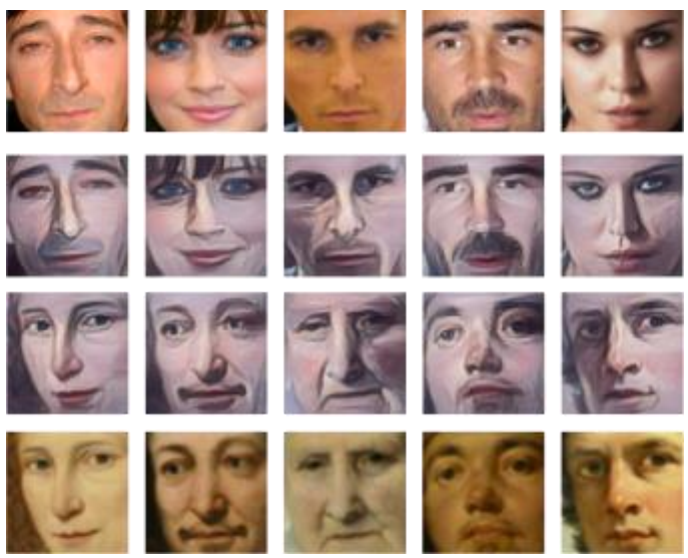

Figure 2. Aligned face images of the IMDb dataset (top row), their style transferred versions (2nd row), samples from the Rijksmuseum dataset (bottom row), and their style transferred versions (3rd row). Classifiers are trained with images on the second row, and tested with images on the third row.

3.2.3 Feature extraction

We have contrasted two facial feature sets for classifying the sex of the sitters. The first one, local binary patterns (LBP), is a representation that is not affected much from illumination changes and shadow effects, and subsequently used frequently for face analysis (Ahonen et al., 2006; Shan et al., 2009).

The second set of features is based on the recently popularized deep learning method. In this approach, one uses a pre-trained deep neural network, which takes an input image and computes a rich internal representation based on image convolutions and pooling operations, in stages. Each stage is one layer of the deep network, and early stages encode simple image features (such as lines, corners, and color blobs). Later stages accumulate more and more semantic features, until the last layer, which performs identification of the person. We have used the Visual Geometry Group (VGG) model (Parkhi et al., 2015), and since our aim is not identification, we use the pen-ultimate layer as the chosen representation.

3.2.4 Classification

Once the LBP or VGG features are extracted from the cropped, aligned, and style-transferred faces, they can be fed into a classifier. We contrasted support vector machines (SVM) (Hearst et al., 1998; Suyken and Vandewalle, 1999), and random decision forests (RDF) for this purpose (Breiman, 2001). The parameter selection and training conditions for these classifiers are available in the code we share with the paper. We have used grid search and 2-fold cross-validation to optimize the parameters.

To measure the performance of the classifiers with the sex of the sitter and the dominant clothing color we have manually annotated 1,505 paintings. We have obtained the best results with LBP features, in conjunction with SVM classifiers (see Table 2). SVMs use a kernel function to transfer features to a high-dimensional space, where the target classes become linearly separable. Radial basis functions (RBF) are most popular kernel functions used for this purpose, and were used in our approach. We have used Weka library's implementation for SVM with RBF kernel and grid search algorithm (Hall et al., 2009).

	IMDB	IMDB and 10k	IMDB, 10k and LFW
LBP & RDF	76.35%	76.94%	71.96%
LBP & SVM	**80.40%**	76.41%	76.25%
VGG & SVM	78.14%	76.54%	77.54%

Table 2. Perceived sex recognition performance of different features (LBP or VGG), classifiers (RDF or SVM), and different training datasets, without style transfer

Using style transfer improves the results. Detailed results of classification of male and female on annotated paintings after style transfer, with LBP features and SVM classifier are given in Table 3. From a total of 499 female portraits, 280 are classified correctly as female. The success rate is much higher in male sitters; from 1,006 portraits, 932 samples were classified correctly (i.e. 92.64%).

The big difference between male and female classification rates may be due to several factors. The changing hair styles (e.g. comparing modern IMDb images with Renaissance paintings), we believe to be of little importance, as the face area is cropped to exclude the hair. However, exclusion of hair and clothing makes classification of females more difficult for the computer. The annotators, obviously, have access to these cues. We note that there is usually bright makeup on the actress photos crawled from IMDb database. This means that the computer learns to expect such makeup for classifying females. We conjecture that this is the major factor behind the lower classification rate for females. Another important factor is the roughness of the facial lines in paintings, particularly for elder sitters. The fine and polished appearance of (mostly) Hollywood actresses may not serve as adequate models for comparison. Figure 3 illustrates some of the misclassifications of the automatic algorithm.

Annotated as	Classified as		Total Count	Performance
	Female	Male		
Female	280	219	499	56.11%
Male	74	932	1,006	92.64%
Total	354	1,151	1,505	**80.53%**

Table 3. Perceived sex recognition performance of LBP features and SVM classifier with RBF kernel, trained from stylized IMDb images, and tested on stylized Rijksmuseum images

3.2.5 Alternative approaches

It is possible to improve these results with better data curation for the classifiers, and with more data. It will not be incorrect to say that today's machine learning approaches are extremely data hungry. The VGG model we have used was trained with 2.6 million face images (Parkhi et al., 2015), but there are models that were trained with up to 500 million faces (Taigman et al., 2015). Currently, we do not have access to such high numbers of faces with annotation.

Another point of improvement is the style transfer. In the present approach, we bring photographs with ground truth and paintings with unknown sex into a specific style space to compare them. It is also possible to create one classifier for each probed painting by transferring the annotated photographs to that painting's style, and to re-train the classifier, thereby avoiding losses arising from the modification of the probed painting itself. However, this approach is extremely time-consuming.

One possible approach is to use a single, well-trained classifier for changing the style of a painting into the style of a photograph (Tomei et al. 2018). We have investigated this alternative in detail, and trained several CycleGAN models for this job (Zhu et al. 2017). In this adversarial learning approach, there are generator networks that synthesize paintings from photographs and photographs from

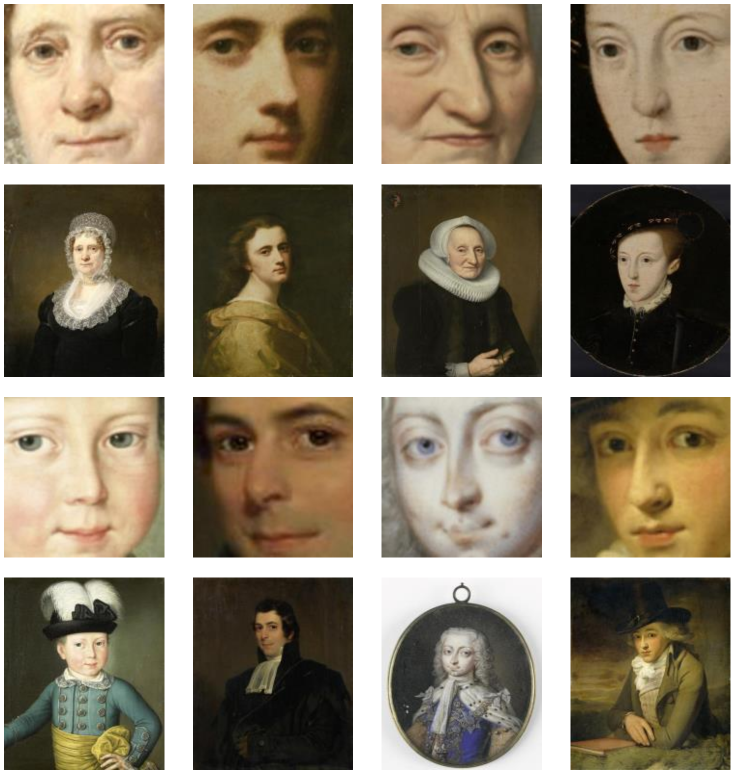

Figure 3. Sitters of portrait paintings assigned to incorrect classes by the automatic approach. Females classified as males (top) and males classified as females (bottom)

Figure 4. Rows are 1) original painting, 2) CycleGAN (trained with Monet paintings), 3) CycleGAN (trained with full Rijksmuseum paintings and IMDb photos), 4) CycleGAN (trained with cropped Rijksmuseum faces, IMDb faces, and movie-cropped faces), 5) Style transfer.

paintings, as well as discriminator networks to discriminate between paintings and photographs, trained jointly. It takes 10-17 days to train these models on a machine with a single GPU. We have used the default parameter setting, and the original implementation of the authors, available from their paper. Figure 4 shows examples synthesized from three such models, which differ mainly in their training set. The first is the original CycleGAN model, trained with 1K Monet paintings and 6K natural images. The second is a CycleGAN we re-trained with 1,505 Rijksmuseum paintings and 8,543 IMDb images. This setting is presumably more realistic, as it incorporates many painting styles. The third is a model trained only with cropped faces from Rijksmuseum and IMDb. This makes the algorithm focus on the faces, as the background is ignored. While visually promising, the gender classification performance of CycleGAN approaches stayed below the proposed model, giving 80.4%, 77.54%, 76.08% accuracy, respectively.

3.3 Cloth area segmentation

Given a painting including a subject, we would like to determine the main color of the clothing for that subject. This is a simpler problem than finding the exact segmentation boundaries of a person, which is a difficult problem that requires large amounts of annotated images for supervised training (Everingham, 2010; Lin et al., 2014). The problem of object segmentation in computer vision corresponds to the problem of finding the exact set of pixels that belong to an object automatically, given the image. The set of pixels is also called a 'mask' (see Figure 5, right). To learn a supervised machine learning model for this problem, a large set of images is required, where the object pixels are annotated. In the PASCAL VOC (Everingham, 2010) and MS COCO (Lin et al., 2014) benchmarks, large amounts of object boundary segmentations were collected from human annotators via Amazon's Mechanical Turk service. This is an expensive process. However, for finding the colors of clothes, a coarse segmentation may be sufficient, where the exact clothing boundary is not required, but we have an area that we are mostly sure that it belongs to the clothing.

There are several prior studies on automatic segmentation of human clothing (Kalantidis et al. 2013; Gallagher and Chen, 2008). Kalantidis et al. introduced a system that used a model image for pose estimation, which is subsequently used for clothing segmentation. This approach was used to retrieve products from online shopping catalogs (Kalantidis et al., 2013). Gallagher et al. (2008) used clothing segmentation to track and identify an individual across different photos taken within a short time interval (Gallagher and Chen, 2008). They have adapted the graph cut algorithm, which is a popular approach in automatic object segmentation, to the problem of segmenting out clothing in the images.

Similar to Gallagher et al. (2008), we have used the graph cut approach; using facial landmark points to initiate the segmentation process of the GrabCut algorithm (Rother et al., 2004). In the GrabCut approach, the image pixels are separated into foreground and background classes. A region of interest is selected as a seed area to grow the foreground and background segments. We generate an initialization mask based on the previously detected facial landmark coordinates. In order to obtain a good position for this mask, we use the distances of the facial landmarks as a yardstick, and define an area below the face, proportional in size to a multiple of the face area (see Fig. 5). This approach worked well for most of the paintings. While it did not correctly segment out all clothing boundaries, it cropped areas that were representative of the clothing colors.

As an alternative to GrabCut, we have experimented with a state of the art deep neural network approach developed for body segmentation in videos (Varol et al., 2017). In this approach, a 3D body generator is used to synthesize many realistic human appearances. The authors superimpose these body images onto different backgrounds to create a large dataset, and train a deep neural network for body segmentation. Unfortunately, this approach was too specialized for photorealistic scenes, and failed on paintings.

Segmentation is affected by objects that are in close proximity with the sitter (such as an armor pieces, or quite frequently, babies held in the arms of the sitter) or by exposed skin from a low-cut dress. Another important problem in segmentation is the fine materials and lacework, which often describe very complex clothing boundaries.

Figure 5. The GrabCut algorithm initialization with the region of interest (left) and the segmented area (right)

3.4 Color representation

Once the segmentation is completed, we need to summarize the colors contained by the garment. The result of the segmentation can be treated as a cloud of pixels, for which a representative color or a small set of representative colors should be determined.

The most obvious difficulty for color representation is the presence of multiple colors in garments. Choosing only the most dominant color is an oversimplification, and perhaps even ill-defined. Consider a bland, regular, white garment, adorned with a small, but bright red ornament. We may argue that the presence of red is highly salient, and has to be taken into account. A frequently encountered case is the presence of a shawl or cape of bright colors (red and blue are used often) over the garment. Are we going to select the color of the garment, or the cape? We return to this problem in the next section, when we discuss the visualization of the results. Other important factors in color representation come from the psychophysics of color. The colors seen by the subjects depend on external light sources (Pinto et al., 2006). While color constancy algorithms can use heuristics to estimate and correct for source colors, the paintings are too diverse to adapt a single set of heuristics, such as natural image statistics (Gijsenij and Gevers, 2007). Furthermore, physical qualities of colors and their effects on the viewer can be quite different, as is well known in visual psychophysics (Birren, 1976).

We did not have any prior assumptions about the shape of the point cloud of color pixels, or how many modes this cloud may contain. Subsequently, we used an unsupervised approach to find the dominant forms in the pixel cloud and to represent the main colors parsimoniously. For this challenge, we tried different clustering algorithms to group colored pixels, where each cluster center would represent the colors falling inside the region described by the cluster. The well-known k-means clustering approach (Hartigan and Wong, 1979) and the more advanced incremental mixtures of factor analyzers (IMoFA) algorithm (Salah and Alpaydın, 2004) were contrasted for this purpose. While k-means assumes spherical and identical distributions in the clusters, IMoFA allows arbitrary ellipsoidal shapes in the color space (see Fig.6).

Before describing the clustering approach, we need to stop for a moment to consider how to represent the colors. Ideally, we should operate in a color space that represents the human perception of color similarity. Although human perception of color is sophisticated and contextual, it is measurable to some extent via image analysis. The human visual system computes color in several stages and achieves independence of spectral variations in illumination, and color constancy (Pinker, 1984). Subsequently, simple pixel-based evaluation of color is a simplification of how colors are perceived in paintings. As a painting is digitized, the sensor properties of the camera, as well as the illumination of the painting will have an influence on the pixel values. Finally, paintings will change colors as they age. When an old painting is removed from its frame, the parts that remain hidden from the damaging light inside the frame stay truer to the original colors, and painting restoration can use this information to correct for colors (Barni et al., 2000).

We have experimented at this stage with different color spaces, such as RGB and HSI. The HSI representation is either based on hue (chromatic colors) or intensity (achromatic colors). For saturation values close to zero ($R \approx G \approx B$), hue is undefined. While HSI is a more perceptually suitable representation, it is not ideal for clustering. Hue values are represented in a circular system, and neither the Euclidean distance (used by k-means), nor the Mahalanobis distance (used by IMoFA) correctly operate in the hue space. One simple approximation is to replace the hue value with its two projections on the unit circle. This replaces the manifold distance (i.e. distance obtained by traveling on the hue circle) with a linear distance that cuts across the circle, but it allows computation of distance measures on the color space. The alternative is using the RGB space, where both Euclidean and Mahalanobis distances are defined.

The contrasted clustering algorithms are both iterative approaches, and differ in two aspects. The k-means approach requires the number of clusters to be known in advance, and uses Euclidean distance. This means that it will fit *k* representative spheres to the cloud of pixels, and the sphere that contains the highest number of pixels will be selected to represent the dominant color of the painting by its center value. On the other hand, IMoFA determines the number of clusters automatically, and allows arbitrary covariance shapes through the use of the Mahalanobis distance. We can visualize the IMoFA clusters as ellipsoids, where each point on the surface is equidistant from the cluster center. Figure 6 illustrates the IMoFA cluster model we have used on a cloud of pixels obtained from a painting. Only the three largest clusters are shown, and the two densest clusters correspond to the white and red parts of the garment.

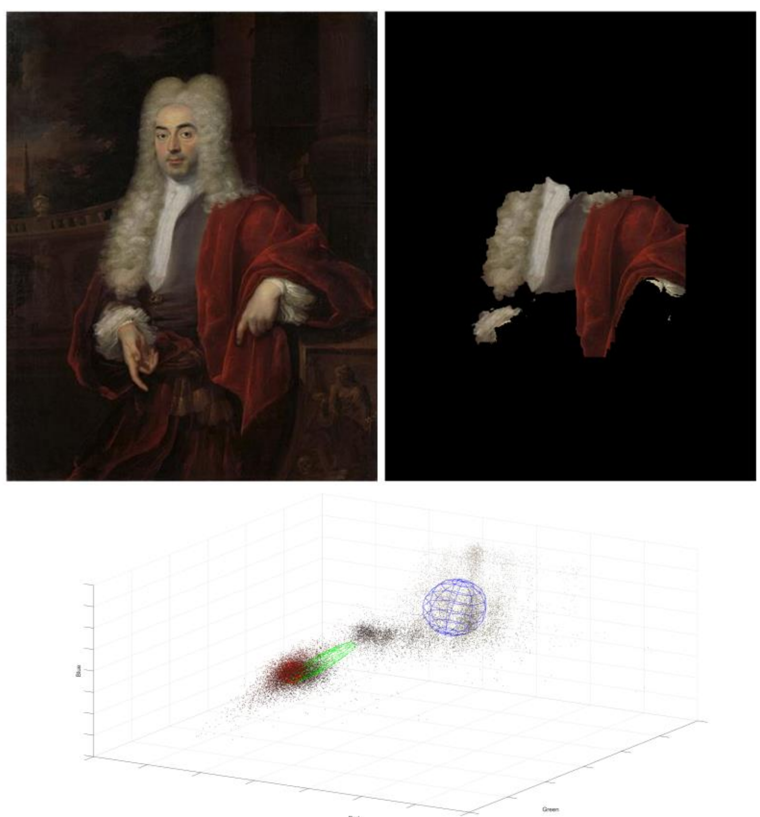

Figure 6. IMoFA cluster model on a cloud of pixels obtained from the detected garment

3.5 Visualization

The last stage of the processing pipeline is to represent paintings with their dominant colors and to produce an interactive visualization of all paintings along a time axis. The plot consists of all the available paintings (or a suitably filtered subset), placed according to their descriptive feature values and years (see Fig.7). These plots are either in RGB vs. time, or saturation vs. time; depending on the selected color definitions. Four plots are prepared and shown together in order to analyze the effects of the perceived sex and its relation with respect to the clothing color and era. The upper two depict paintings of females, and the lower two depict paintings of males. If the hue value cannot be calculated, i.e. if the dominant color is close to white, gray, or black, it is shown in the second part of the graph, sorted according to intensity. On the left, a legend shows the dominant hue or the intensity for the corresponding position in the graph. For this visualization, we use male/female labels checked by the painting expert, who also took the title information into account.

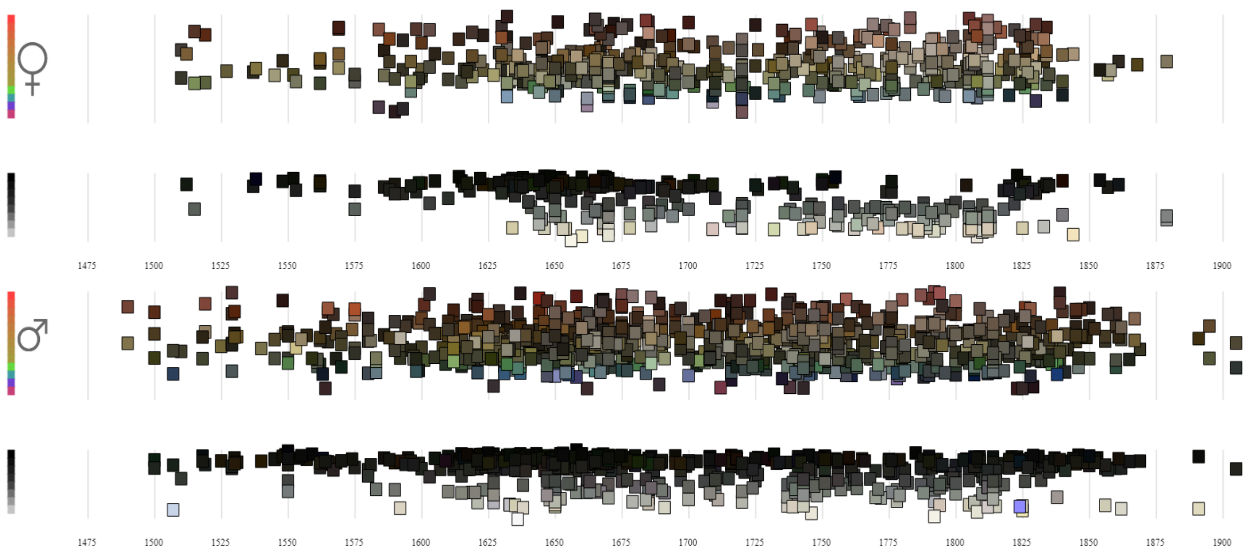

Figure 7. Interactive visualization of a subset of the painting database according to the dominant colors

One issue is the representation of garments with multiple dominant or salient colors. It is possible to represent the garment with multiple points in the visualization, using one point for each color important enough to be represented. The main problem with this approach is that a multi-colored garment would have a much stronger representation than a single-colored garment in the visualization, and mislead the reader. Using size normalization for these points and showing thumbnails proportional to color dominance could be a potential solution, but points from a single image will not be connected in the visualization. We believe the simple, single-color patches give a more intuitive visualization, but other approaches are easy to incorporate once the basic algorithms are in place. In a previous study (Sari et al., 2017), we have reported an approach where we have aggregated garment colors of male and female sitters over a 50-

year period. By pooling dominant colors, we observed that females wear lighter colors compared to males and show higher variance over the years, which can also be observed in Fig. 7. While this does not refute Laqueur's theory, it suggests that the color-based segregation happened over a longer period, starting from mid-17th Century, but more research is needed for clear conclusions.

The visualization is prepared as an HTML page, the color schemes and font styles are defined using a CSS style sheet. Data acquisition, layout population, displays and actions are performed by JavaScript algorithms. When the client-side algorithm of the page is initialized, it retrieves the list of available paintings from the server. Through this list, the meta-data corresponding to the requested painting can be accessed, including title, segmentation result, dominant color, and a link to the painting image file. The collected information is used to populate the rest of the page on the client-side [Note 3].

4. Conclusions

Today, with the increases in computing power and with new capabilities enabled by deep learning algorithms, computer vision has a lot to offer to art historical studies. In this work, we have combined various off-the-shelf algorithms in image analysis to detect the garment colors of female/male sitters in Western portrait paintings from 15th to 19th Century. Our cultural analytics approach provides us with new tools for testing theories of incremental change over time, particularly those that may be too minor to be perceived by looking at a small number of artworks, but discernible on a large database of images when processed with powerful computing methods. Our focus was on establishing and evaluating an approach that would be scalable for the analysis and visualization of thousands of painting in real time. This way, humanities scholars can devise methods to test theories such as Laqueur's transition from one- to two-sex model. Here, we would like to highlight the challenges we have faced, and the solutions we have developed.

The automatic face detection in paintings is a problem already addressed and largely solved in the literature with portraits. We have applied a classic face detection algorithm to extract portrait paintings from a dataset that is specifically generated to challenge style and genre detection. Out of 112,039 items in the dataset only 2,462 were found suitable for the purpose of this study, i.e. portraits with only one sitter. Out of these, we have selected 1,505 samples that were representative of their period. These portraits were annotated for their representative garment colors to generate a reference for further study. This database is made available at the time of the publication as a contribution to digital art history research.

Finding the sex of the sitters automatically was challenging; expert knowledge can be used to avoid errors at this stage. Our automatic approach obtained prominent results in detecting males, but misclassified many female sitters as males. This misclassification can be interpreted as a positive sign of

Laqueurian theory of the transformation from one to the two sex model. In earlier eras, the male and female sitters were apparently more similar in their appearances than the sitters in our training set (i.e. actors and actresses of the IMDb database). The second challenge was to devise an algorithm to extract the most salient color of the detected garment. We have applied different approaches for color extraction, and decided that it is best to choose only one color from a given garment for displaying the results in an interactive visualization interface. This approach gives an overview of the salient colors and their usage distribution over time, while at the same time, it is easy for the user to select and gain extra information for any given point in the visualization.

Notes

1. Data obtained from https://www.rijksmuseum.nl/, on 28 November 2018.
2. https://opencv.org/
3. All the source files are placed on a GitHub repository, and can be freely downloaded from the following address: https: //github.com/CihanSari/ClothingWeb. The interface is available at: https://clothingweb.cihansari.com/showpaintings.

Acknowledgments

We thank the anonymous reviewers and Dr. Zoe Borovsky for their constructive comments.

References

Ahonen, T., Hadid, A. and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(12), pp.2037-2041.

Bainbridge, W.A., Isola, P. and Oliva, A., (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), p.1323.

Barni, M., Bartolini, F. and Cappellini, V. (2000). Image processing for virtual restoration of artworks. *IEEE multimedia*, *7*(2), pp.34-37.

Birren, F. (1976). Color perception: Beyond the eye into the brain. *Leonardo*, pp.105-110.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), pp.5-32.

Cottrell, G.W. and Metcalfe, J. (1991). EMPATH: Face, emotion, and gender recognition using holons. In *Advances in neural information processing systems,* pp. 564-571.

Crowley, E.J. and Zisserman, A. (2014). In Search of Art. In *ECCV Workshops (1),* pp. 54-70.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A. (2010). The Pascal visual object classes (VOC) challenge. *International journal of computer vision*, *88*(2), pp.303-338.

Gage, J. (1990). Color in Western art: An issue? *The Art Bulletin*, *72*(4), pp.518-541.

Gallagher, A.C. and Chen, T. (2008). Clothing cosegmentation for recognizing people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Gatys, L.A., Ecker, A.S. and Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414-2423.

Gijsenij, A., and Gevers, T. (2011). Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 687-698.

Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*(1), pp.33-51.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), pp.10-18.

Hartigan, J.A. and Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), pp.100-108.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. and Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, *13*(4), pp.18-28.

Huang, G.B., Ramesh, M., Berg, T. and Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Vol. 1, No. 2, p. 3). Technical Report 07-49, University of Massachusetts, Amherst.

Isikdogan, F., Adıyaman, I., Akdag Salah, A.A. and Salah, A.A. (2016). A New Database and Protocol for Image Reuse Detection. In *European Conference on Computer Vision,* pp. 903-916. Springer International Publishing.

Kalantidis, Y., Kennedy, L. and Li, L.J. (2013). Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval,* pp. 105-112. ACM.

Kim, D., Son, S.W. and Jeong, H. (2014). Large-scale quantitative analysis of painting arts. *Scientific reports*, *4*, p.7370.

Laqueur, T.W. (1990). *Making sex: Body and gender from the Greeks to Freud*. Harvard University Press.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. (2014). Microsoft COCO: Common objects in context. In *European conference on computer vision,*pp. 740-755. Springer, Cham.

Manovich, L. (2016). The science of culture? Social computing, digital humanities and cultural analytics. *CA: Journal of Cultural Analytics*, *1*(1).

Mensink, T. and van Gemert, J. (2014). The Rijksmuseum Challenge: Museum-centered visual recognition. In *Proceedings of International Conference on Multimedia Retrieval,* p. 451. ACM.

Ng, C.B., Tay, Y.H. and Goi, B.M. (2012). Recognizing human gender in computer vision: a survey. In *Pacific Rim International Conference on Artificial Intelligence,* pp. 335-346. Springer, Berlin, Heidelberg.

Paoletti, J.B. (1987). Clothing and gender in America: Children's fashions, 1890-1920. *Signs: Journal of Women in Culture and Society*, *13*(1), pp.136-143.

Parkhi, O.M., Vedaldi, A. and Zisserman, A. (2015). Deep Face Recognition. In *BMVC,* 1(3), p. 6.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*(1), pp.1-63.

Pinto, P.D., Linhares, J.M., Carvalhal, J.A. and Nascimento, S.M. (2006). Psychophysical estimation of the best illumination for appreciation of Renaissance paintings. *Visual Neuroscience*, *23*(3-4), pp.669-674.

Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2017). An all-in-one convolutional neural network for face analysis. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on* (pp. 17-24). IEEE.

Rother, C., Kolmogorov, V. and Blake, A. (2004). GrabCut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics,* 23(3), pp. 309-314. ACM.

Salah, A.A. and Alpaydin, E. (2004). Incremental mixtures of factor analysers. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 1), pp. 276-279. IEEE.

Salah, A.A., Cinar, H., Akarun, L. and Sankur, B. (2007). Robust facial landmarking for registration. In *Annales des Télécommunications* (Vol. 62, No. 1-2), pp. 83-108. Springer-Verlag.

Sarı, C., Salah, A.A. and Akdag Salah, A.A. (2017). Tracing the Colors of Clothing in Paintings with Image Analysis. In *Proc. Digital Humanities*.

Sartori, A., Şenyazar, B., Akdag Salah,A.A., Salah, A.A. and Sebe, N. (2015). Emotions in Abstract Art: Does Texture Matter?. In *International Conference on Image Analysis and Processing*, pp. 671-682. Springer, Cham.

Shan, C., Gong, S. and McOwan, P.W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, *27*(6), pp.803-816.

Spratt, E.L. and Elgammal, A. (2014). Computational beauty: Aesthetic judgment at the intersection of art and science. *arXiv preprint* arXiv:1410.2488.

Srinivasan, R., Rudolph, C. and Roy-Chowdhury, A.K. (2015). Computerized Face Recognition in Renaissance Portrait Art: A quantitative measure for identifying uncertain subjects in ancient portraits. *IEEE Signal Processing Magazine*, *32*(4), pp.85-94.

Stork, D.G. (2009). Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. *Computer Analysis of Images and Patterns*, Springer, pp.9-24.

Suykens, J.A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), pp.293-300.

Taigman, Y., Yang, M., Ranzato, M.A. and Wolf, L. (2015). Web-scale training for face identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2746-2754.

Tomei, M., Baraldi, L., Cornia, M., & Cucchiara, R. (2018). What was Monet seeing while painting? Translating artworks to photo-realistic images. In *European Conference on Computer Vision Workshops.*

Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M., Laptev, I. and Schmid, C. (2017). Learning from synthetic humans. *arXiv preprint arXiv:1701.01370*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision.*