# The Face of Art: Landmark Detection and Geometric Style in Portraits

JORDAN YANIV, Tel Aviv University & The Interdisciplinary Center, Herzliya
YAEL NEWMAN, Tel Aviv University & The Interdisciplinary Center, Herzliya
ARIEL SHAMIR, The Interdisciplinary Center, Herzliya

Fig. 1. Top: landmark detection results on artistic portraits with different styles allows to define the geometric style of an artist. Bottom: results of style transfer of portraits using various artists' geometric style including Modigliani, Picasso, Keane, Leger and Foujita.
From left to right: Portrait of Bindo Altoviti, 1515 by Raphael courtesy WikiArt [Public Domain] via (http://bit.ly/2HzoPyz), Gypsy Woman with a Baby, 1919 by Amedeo Modigliani courtesy WikiArt [Public Domain] via (http://bit.ly/2EbAWkn), Two Women with Rice Cakes and Swords, 1844-1845 by Utagawa Kunisada courtesy Van Gogh Museum [Public Domain] via (http://bit.ly/2JUuFh1), Little Girl with Doll, 1918 by Tsuguharu Foujita courtesy WikiArt [Public Domain US] via (http://bit.ly/2Q82xbf), Portrait of the Composer Anton von Webern, 1914 by Oskar Kokoschka courtesy WikiArt [Public Domain US] via (http://bit.ly/30AuxsS), Woman with Peanuts, 1962 ©Estate of Roy Lichtenstein courtesy Image-Duplicator [Fair Use] via (http://bit.ly/2HA3DIF). Natural face images from [Minear and Park 2004], used with permission.

Facial Landmark detection in natural images is a very active research domain. Impressive progress has been made in recent years, with the rise of neural-network based methods and large-scale datasets. However, it is still a challenging and largely unexplored problem in the artistic portraits domain. Compared to natural face images, artistic portraits are much more diverse. They contain a much wider style variation in both geometry and texture and are more complex to analyze. Moreover, datasets that are necessary to train neural networks are unavailable.

We propose a method for artistic augmentation of natural face images that enables training deep neural networks for landmark detection in artistic portraits. We utilize conventional facial landmarks datasets, and transform their content from natural images into "artistic face" images. In addition, we use a feature-based landmark correction step, to reduce the dependency between the different facial features, which is necessary due to position and shape variations of facial landmarks in artworks. To evaluate our landmark detection framework, we created an "Artistic-Faces" dataset, containing 160 artworks of various art genres, artists and styles, with a large variation in both geometry and texture. Using our method, we can detect facial features in artistic portraits and analyze their geometric style. This allows the definition of signatures for artistic styles of artworks and artists, that encode both the geometry and the texture style. It also allows us to present a geometric-aware style transfer method for portraits.

CCS Concepts: • **Computing methodologies** → *Image processing*; Image representations; *Non-photorealistic rendering*; *Neural networks.*

Additional Key Words and Phrases: facial landmark detection , neural networks, artistic image augmentation, geometry aware style transfer

**ACM Reference Format:**
Jordan Yaniv, Yael Newman, and Ariel Shamir. 2019. The Face of Art: Landmark Detection and Geometric Style in Portraits. *ACM Trans. Graph.* 38, 4, Article 60 (July 2019), 15 pages. https://doi.org/10.1145/3306346.3322984

Authors' addresses: Jordan Yaniv, Tel Aviv University & The Interdisciplinary Center, Herzliya, jordanya@mail.tau.ac.il; Yael Newman, Tel Aviv University & The Interdisciplinary Center, Herzliya, yaelnewman@mail.tau.ac.il; Ariel Shamir, The Interdisciplinary Center, Herzliya, arik@idc.ac.il.


0730-0301/2019/7-ART60 $15.00
https://doi.org/10.1145/3306346.3322984

## 1 INTRODUCTION

Portraiture has been an important part of art going back as far as 5000 years ago to ancient Egypt. Before the invention of photography, a painted, sculpted, or drawn portrait was the only way

to record the appearance of a person. However, portraits have always been more than just a recording of a face, they are the artist's interpretation of the subject: they can be realistic, abstract, representational etc. (Figure 1) - they convey the style of the artist.

Given the growing abilities of digital tools and algorithms, portrait analysis, synthesis and stylization are also in high demand. However, most algorithms that work well for natural images fail when applied to more artistic inputs. The differences between artistic portraits and face images span two domains: texture appearance differences, and geometric differences. Many recent works concentrate on capturing appearance style [Jing et al. 2017], but neglect geometric style.

In this work, we try to bridge the gap between machine understanding of face photographs and understanding of artistic portraits. We concentrate on detecting facial landmark in artworks and learning geometric style. Facial landmark detection aims to localize a set of predefined landmarks such as eye corners or mouth corners in a face depiction. It is the basis of any portrait drawing analysis, and a fundamental problem in computer vision. Detecting facial features in art allows us to model the *geometric* style of an artist and use it, for instance, in geometry-aware style transfer (see Figure 1).

Using neural networks, impressive progress has been made on facial landmark detection in recent years on natural face images. However, moving to artistic portraits this becomes much more of a challenge. Compared to natural face images, artistic portraits are much more diverse. They contain a much wider variation in both geometry and texture and are more complex to analyze. Facial features in artworks are often exaggerated in ways that lead to the deviation from the implicit humanly attributes. Moreover, large scale datasets that are necessary to train neural networks are unavailable.

We propose to use *artistic image augmentation* of natural face images. We utilize conventional facial landmarks datasets, and transform their content from natural images into "artistic face" images. This transformation is performed using image warping with various geometric variations along with neural style transfer. Our augmentation method enables training deep neural networks despite the absence of annotated data of artistic portraits.

The network we train in this work is based on the ECT (Estimation-Correction-Tuning) framework for facial landmark detection [Zhang et al. 2018], which combines the advantages of the global robustness of a data-driven method, the outlier correction capability of a model-driven method, and non-parametric optimization of landmark mean-shift. We enhance the frameworks performance on artistic portraits by using artistic augmentation in the training procedure. In addition we add a spatial transformer network (STN) component [Jaderberg et al. 2015] to the network, and use a feature-based correction step. These additions reduce the dependency between the different facial features, which are necessary due to position and shape variations of facial landmarks in artworks.

To evaluate our landmark detection, we create an "artistic-faces" dataset, containing 160 artworks of various art genres, artists and styles, with a large variations in both geometry and texture. To maintain consistency with previous works, the images were annotated with 68 facial landmarks using a semi-automatic procedure. We measure our results on this dataset and publish the data for future research.

We demonstrate several applications for artistic facial feature detection. First, we can use the detected features for style analysis. By extracting shape information from multiple artworks of the same artist, we can learn signature facial features proportions and define the geometric style of the artist. Second, creating a signature for geometric style allows us to compare and classify artworks based on their geometric style and not just texture. Third, we can utilize geometric style for synthesis application such as geometry-aware style-transfer for portraits (see Figure 1 and other figures in this work).

Our main contributions are:

- A method for artistic facial feature detection based on neural networks.
- The collection of an "artistic-faces" dataset for future research.
- A method for analyzing the geometric style of artists and portraits, creating a signature of style.
- A method for geometry-aware style transfer for portraits.

## 2 RELATED WORK

*Facial Landmark Detection.* A common approach for facial landmark detection is to learn a regression model for feature positions [Cao et al. 2014; Lv et al. 2017; Xiong and la Torre 2013]. Many current methods leverage deep CNN to learn facial features and regressors [Dollar et al. 2010; Sun et al. 2013] with a cascade architecture to progressively update the landmark estimation [Zhu et al. 2015, 2016]. Another approach to facial landmark detection takes the advantages of end-to-end training from deep CNN model to learn robust heatmap of facial landmarks [Bulat and Tzimiropoulos 2016, 2017; Yang et al. 2017].

Our work follows Estimation-Correction-Tuning (ECT) framework presented by Zhang et al. [2018]. This framework combines the advantages of the global robustness of data-driven method (deep CNN), outlier correction capability of model-driven method (PDM) [Cootes and Taylor 1992] and non-parametric optimization of regularized landmark Mean-Shift (RLMS) [Saragih et al. 2011]. We adapt the ECT framework for the artistic faces domain by using artistic image augmentation in the training procedure. To reduce the dependency between the different facial features we use a feature-based correction step, and incorporating a Spatial Transformer Network (STN) component [Jaderberg et al. 2015] into the network. STNs have increased the accuracy of a wide range of tasks (e.g. classification), as the network learns invariance to geometric warping. Yu et al. [2016] have used a variation of STN for landmark detection. In their work, the STN predicts the warping parameters on a feature map, but applied directly to the landmarks, while we use it on the feature map itself similar to [Jaderberg et al. 2015].

*Face and Landmark Detection in Art.* Previous research on faces in art focus on specific art genres, and mainly address the problems of face detection and recognition. Several works concentrate on Manga face images [Sun and Kise 2010; Yanagisawa et al. 2014]. For example, the Manga FaceNet proposed by Chu and Li [2017] offers a CNN based architecture for detecting Manga faces. Other works focus on comics characters [Nguyen et al. 2017] or caricatures [Huo et al. 2018]. Jha et al. [2018] annotated caricatures with 15 facial landmarks and trained a network for recognition and landmark detection

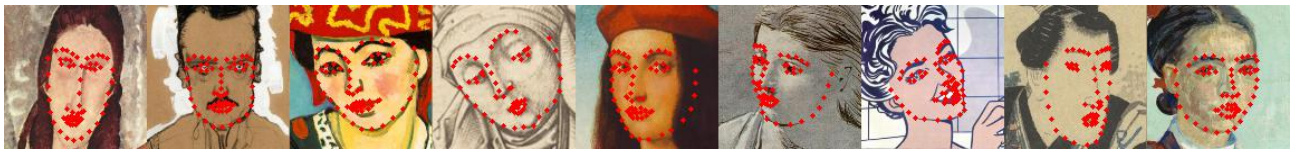


Fig. 2. Samples from the Artistic-Faces Dataset including landmarks.
From left to right: Portrait of Jeanne Hebuterne, 1919 by Amedeo Modigliani courtesy WikiArt [Public Domain] via (http://bit.ly/2WPHkoW), Portrait of Eduard Kosmack, Frontal, with Clasped Hands, 1910 by Egon Schiele courtesy WikiArt [Public Domain] via (http://bit.ly/30sCkss), The Red Madras Headdress, 1907 by Henri Matisse courtesy WikiArt [Public Domain US] via (http://bit.ly/2LPdjo2), Saint Elizabeth of Thuringia, c. 1475/1480 by Israhel van Meckenem courtesy NGA Images [Public Domain] via (http://bit.ly/30sCqAk), Woman in white, 1923 by Pablo Picasso courtesy WikiArt [Public Domain US] via (http://bit.ly/2WQNJjF), Portrait of the Young Pietro Bembo, 1504 by Raphael courtesy WikiArt [Public Domain] via (http://bit.ly/2YB9q7P), Girl in Bath, 1963 ©Estate of Roy Lichtenstein courtesy Image-Duplicator [Fair Use] via (http://bit.ly/2QbnDoR), Actor and Woman on a Riverbank, 1820-1830 by Utagawa Kunisada courtesy Van Gogh Museum [Public Domain] via (http://bit.ly/2VHN3Ri), La Mousme seduta, 1888 by Vincent van Gogh courtesy NGA Images [Public Domain] via (http://bit.ly/2Jtr8GS).

in faces on datasets that combine cartoon faces and human faces. Interestingly, they achieved better performance for the landmark detection by using only real face images in the training set. Stricker at al. [2018] propose a new landmark model for manga, including 60 facial landmarks suitable for capturing the manga facial shape. They annotated 1446 images out of the 109Manga dataset [Ogawa et al. 2018] for training a CNN based on the Deep Alignment Network (DAN) architecture [Kowalski et al. 2017].

As opposed to these previous works, we propose a framework that supports multiple artistic genres and styles. We use a 68 landmarks model [Sagonas et al. 2013], to match current research in the natural faces domain and enable one-to-one comparison, matching and transition between natural and artistic faces. Moreover, a larger number of landmarks enables better analysis and synthesis of the portrait artistic geometric style.

*Style Transfer.* Image stylization has been extensively studied in literature. Inspired by the power of CNN, Gatys et al. [2015] presented an optimization based method for transferring the style of a given artwork to an image. Current neural style transfer methods fit into one of two categories [Jing et al. 2017]: optimization-based methods [Gatys et al. 2015; Li and Wand 2016; Li et al. 2017], and model-based neural methods [Johnson et al. 2016; Ulyanov et al. 2016]. The first category transfers the style by iteratively optimizing an image. The second category optimizes a generative model offline, and produces the stylized image with a single forward pass. Style transfer methods that target portraits specifically achieve better results [Kaur et al. 2017; Seleim et al. 2016]. However, the success of previous style transfer methods remains limited to color and texture, and fail to transfer geometric style (e.g. feature exaggeration in caricatures).

Recent works in image caricaturization [Cao et al. 2018; Shi et al. 2018], use Generative Adversarial Networks (GANs) [Goodfellow et al. 2014] to learn both texture and geometric style of human caricatures. These works achieve impressive results compared to texture-only style transfer methods. Nevertheless, they rely on manual facial feature detection of caricatures and do not learn a model for an artist. Our framework can be used to automatically annotate large-scale datasets containing artworks of various artistic styles, and aid such algorithms to learn geometric style models as well as expand beyond the caricature domain to other artistic styles.

*Data Augmentation.* Data augmentation is a standard technique for improving the generalization of deep neural networks. It artificially inflates a dataset by using transformations to derive new examples from the original dataset, gaining invariance to whichever transformations are used. Recent works incorporate image texture style in the augmentation process. Jackson et al. [2018] propose a new augmentation method using style randomization. They utilize a multi-style model-optimized framework [Ghiasi et al. 2017] to learn texture style of images and show performance improvement in the fields of image classification, cross-domain classification and monocular depth estimation. Dong et al. [2018] propose a Style-Aggregated-Network (SAN) to deal with the intrinsic variance of natural image styles for facial landmark detection. For training their generative module, they create 3 stylized versions of each original image using Adobe Photoshop (light, gray and sketch styles). Using the stylized images along with the originals they train face generation models to transfer styles via CycleGAN [Zhu et al. 2017] and show performance improvement over state-of-the-art algorithms.

For our purpose of landmark detection in art, we need to deal with a significantly larger intrinsic variance of image styles. Artistic portraits have a large variation both in texture and in geometry. For this purpose, we propose an Artistic Image Augmentation method for natural face images. Our goal is to narrow the gap between the natural and artistic face domains and train a landmark detection network suitable for various artistic portrait styles.

## 3 ARTISTIC-FACES DATA SET

The Artistic-Faces dataset contains 160 artistic portraits of 16 different artists, which were chosen to be representative of a wide range of artistic styles, both in geometry and texture. It contains artwork styles ranging from High Renaissance through Cubism to Comics (Figure 2). The portraits are annotated with 68 facial landmarks to remain consistent with previous works in facial landmark detection of natural faces.

For each artwork we provide the following metadata : artist name, artwork title, style, date and source. We use this dataset to evaluate our landmark detection framework. The Artistic-Faces dataset will

be publicly available and can be used to evaluate future works in landmark detection and other algorithms[1].

*Data Collection.* Our dataset is mainly drawn from the Painter By Numbers (PBN) dataset, which consists of 103,250 artworks. The images in the PBN dataset were mostly obtained from the WikiArt dataset[2], In addition, our dataset contains images from online art collections provided by: Van Gogh Museum's collection, Image-Duplicator by the Roy Lichtenstein Foundation, National Gallery of Art (NGA) image collection, Tate collection, The Met collection, Nasjonalmuseet collection and Google image search engine.

Out of the portrait dataset containing around 400 artists and 15,000 artworks, we selected 16 artists that represent a wide variety of styles both in texture and geometry. The face area in the selected artworks, were automatically detected using a MultiTask Cascaded Convolutional Network (MTCNN) [Zhang et al. 2016]. As the MTCNN was trained on natural face images, not all input artworks resulted in a detected face. Around 75% of the input artworks resulted in one or more face detected. The output face bounding boxes were used to crop the selected artworks with a margin of 25% of the bounding box size, and then rescaled to 256x256 pixels. From the resulting face crops we removed false positive detections and profile images, resulting in frontal and semi-frontal face crops. For each of the 16 artists we randomly selected 10 images resulting in 160 face crops of artworks.

Initial landmark detection of 68 facial landmarks was obtained using the Ensemble of Regression Trees (ERT) algorithm [Kazemi and Sullivan 2014] provided by DLib [King 2009]. The initial landmarks were then manually corrected using the landmarker.io landmarking tool provided by the Menpo Project [Alabort-i-Medina et al. 2014].

## 4 ARTISTIC FACE VARIATIONS

To motivate and plan the adaptation of existing algorithms to detect landmarks of artistic portraits, we need to understand the key differences between natural face images and artistic portraits. The differences between the two domains are revealed by two main aspects: geometric and textural.

### 4.1 Texture variations

The variation in color and texture of natural face images is significantly smaller than the variation within the artistic portraits category. Natural face images are usually homogeneously textured, piecewise smooth, and contain a limited color palette. Faces in art have a wide range of colors and textures, caused both by the medium used (oil, acrylic, charcoal, digital, etc.) and by the style of the artwork (see examples in Figure 2).

To model the appearance style of each portrait, we follow the approach proposed by Gatys et al. [2015] and use Gram-based correlations matrices to model textures. In Figure 3 we embed in 2D the appearance styles of a set of natural faces (NF) and a set of artistic faces (AF) taken from artworks (Section 3) using the T-SNE method. It is clear that there is little overlap between NF (gray points) and AF (red points) data in terms of texture.

[1]http://www.faculty.idc.ac.il/arik/site/faceofArt.asp
[2]https://www.wikiart.org

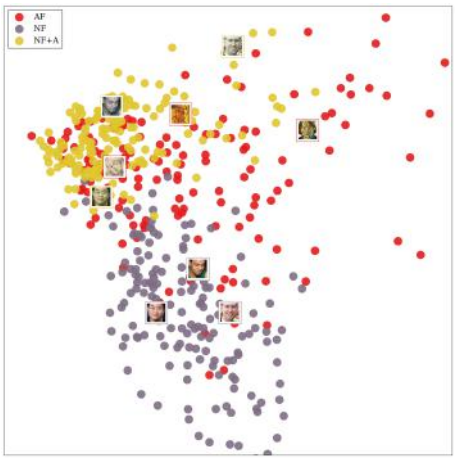



Fig. 3. T-SNE visualization of the vgg Gram matrix style embedding (conv_4_3). Gray points represent natural faces (NF), red points represent artistic faces (AF), and yellow points represent the same natural face images with artistic texture augmentation (NF+A) as presented in our work. Some sample thumbnail images are shown for illustration.

### 4.2 Geometric variations

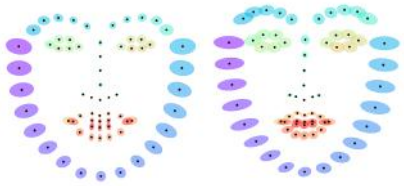
We define the geometry of a face using the following attributes: the shape of the face boundary, the shape of individual facial features (eyes, mouth, nose), the size and proportions of the features and their relative locations. In natural faces, there are typical proportions and standard shapes for each facial feature, and the relative location of the features is almost constant. In a sense, the human face is very "regular" as a canonical shape.

In contrast, in artistic faces there are larger variations in the features' aspect-ratio and relative scale, and in the features' relative locations. Such shape variation can be found in a wide variety of artistic styles, from Primitivism through Expressionism to Comics and Caricatures (see Figure 2).

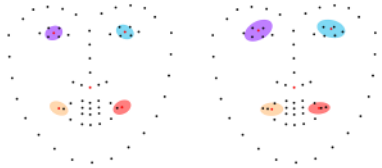
To illustrate one aspect of the differences in the geometric variation of faces, in Figure 4 we compare the distribution of positions of landmark points on two sets: the 300W training set of natural faces (Section 6), and our artistic faces set (Section 3), and on two models: the 68 landmark points of [Sagonas et al. 2013], and a 5 landmark points model. In both models, the variation in positions of the landmark points is larger in artistic faces, demonstrating higher geometric variation.

## 5 METHOD

The base of our landmark detector is the ECT approach of Zhang et al. [2018]. Given an input face image, the landmark detection result is obtained in three steps of estimation, correction and tuning. The Estimation Step aims to compute a global localization of each landmark based on the peak response points in the response maps, which are learned using a fully convolutional network (see supplemental material for architecture details). In the correction step, a more accurate initial shape is computed by correcting outlier landmarks using a pre-trained point distribution model (PDM).



(a) 68 landmark points



(b) 5 landmark points (red points)

Fig. 4. Comparison between natural faces (left) and artistic faces (right) in terms of the landmark point distributions. Mean position and the ellipse of one standard deviation are shown. The variability in artistic faces is larger.

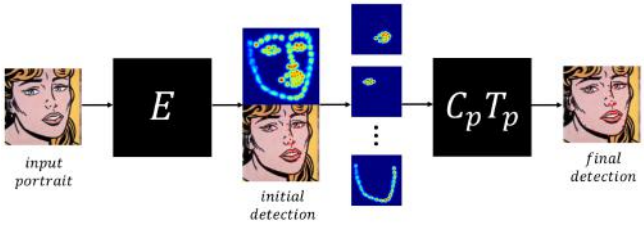



Fig. 5. Overview of our proposed Framework for Landmark Detection in Artistic Portraits. The Estimation Step ($E$) with STN sub-network obtains a coarse landmark shape according to the peak response points in the heatmaps obtained by the Heat-Maps Network. The initial detection is then followed by *part-based* Correction and Tuning ($C_pT_p$).
Face taken from Anxious Girl, 1964 ©Estate of Roy Lichtenstein courtesy Image-Duplicator [Fair Use] via (http://bit.ly/2VscEZq).

Finally, the landmark locations are fine-tuned based on weighted regularized mean shift. We enhance the frameworks performance on artistic portraits by using artistic augmentation in the training procedure. In addition, we add an STN component to the network and use a feature-based correction step, to reduce the dependency between the different facial features. Figure 5 shows an overview of our proposed Framework for Landmark Detection in Artistic Portraits.

## 5.1 Estimation Step

*Heat-Maps Network.* Following Zhang et al. [2018] framework, a regressor is used to regress an ideal heatmap for each landmark point in a data-driven manner. The ideal heatmap of the $i$-th landmark for image $\mathcal{I}$ is defined as a single-channel image $\mathcal{M}^i$ with the same resolution as $\mathcal{I}$, the value at position $\mathbf{z}$ is defined as $\mathcal{M}^i_{\mathbf{z}} = \mathcal{N}(\mathbf{z}; \mathbf{x}^*_i, \sigma^2\mathbf{I})$, where $\mathbf{x}^*_i$ is the ground truth location of the $i$-th landmark, and $\sigma$ is the variance of a blurring kernel creating the map.

Fig. 6. Artistic face augmentation is performed in two steps 1. Style transfer from random artworks, and 2. geometric deformation by perturbing landmark positions and warping the image. leftmost images from [Minear and Park 2004], used with permission.

The training dataset consists of pairs $D = \{(\mathcal{I}, \mathbf{l}^*)\}$, where $\mathbf{l}^* = [\mathbf{x}_1, ..., \mathbf{x}_n]^T \in \mathbb{R}^{n\times 2}$ is the ground truth positions of $n$ landmarks embedded in image $\mathcal{I}$. The objective of the regressor becomes estimating the network weights $\lambda$ that minimize the following L2 loss function:

$$\mathcal{L}(\lambda) = \sum_{(\mathcal{I},\mathbf{l}^*)\in D} \sum_i \left\| \mathcal{M}^i - \phi^i(\mathcal{I}, \lambda) \right\|^2 \quad (1)$$

where $\phi^i(\mathcal{I}, \lambda)$ is the output of the $i$-th channel of the heatmap network, on the input image $\mathcal{I}$.

*Spatial Transformer Network.* The Spatial Transformer Network (STN) predicts the transformation parameters from its input, applies it on a grid, and samples the input according to the warped grid using bilinear interpolation. We used the STN on the heat-maps before upsampling, predicting an affine transformation (6 parameters) and applying it to the heat-maps, hence our STN outputs the warped heatmaps. The architecture of the STNs localisation network, used to predict the transformation parameters, is detailed in our supplemental material.

## 5.2 Artistic Data Augmentation

To our knowledge, there is no large-scale dataset of 68-landmarks annotated artistic portraits. To overcome this absence, our key idea is to augment an annotated natural face dataset by transforming it to be more similar to artistic portraits. Following our observations from Section 4, we divide the artistic augmentation pipeline into two separate processes - first texture augmentation is applied on the image and then geometric augmentation of the results (see Figure 6).

Using artistic data augmentation, as opposed to basic augmentation, we can increase dramatically the size and variability of the original dataset and bring the augmented natural face domain closer to the artistic one (see Figure 3, yellow points).

*Texture Style Augmentation.* For the texture augmentation process, we follow the texture modelling approach proposed by Gatys et al. [2015] and use the Gram Matrix of the VGG16 feature maps. To create a texture-augmented copy of an input face image we use the image-optimization style transfer method proposed in their work.

Given a content image $\mathcal{I}_c$ and a style image $\mathcal{I}_s$, the algorithm tries to seek a stylised image $\mathcal{I}$ that minimises the following objective:

$$\mathcal{I}^* = \arg\min_{\mathcal{I}} \alpha\mathcal{L}_c(\mathcal{I}_c, \mathcal{I}) + \beta\mathcal{L}_s(\mathcal{I}_s, \mathcal{I})$$

where $\mathcal{L}_c$ compares the content representation of a given content image to that of the stylised image, and $\mathcal{L}_s$ compares the Gram-based style representation derived from a style image to that of the stylised image. $\alpha$ and $\beta$ are used to balance the content component and style component in the stylised result.

Given a face dataset $D = \{(\mathcal{I}, \mathbf{l}^*)\}$ and an art dataset $\{\mathcal{I}_{art}\}$ (containing general artistic images not necessarily portraits and not containing artists from our "artistic-faces" dataset), we randomly sample style images from $\{\mathcal{I}_{art}\}$ and create $K$ style-textured augment versions for each image $\mathcal{I}$ in a pre-processing stage. The resulting dataset is $D = \{(\mathcal{I}, \mathbf{l}^*, \{\mathcal{I}_k^t\}_{k=1}^K)\}$, where $\mathcal{I}_k^t$ is the $k$'th texture-stylized version of image $\mathcal{I}$.

We chose to use the Neural Style Transfer algorithm for texture augmentation, as it creates more credible and visually appealing results than model-based style transfer methods. The artistic texture augmentation brings the natural face domain closer to the embedding of artistic faces.

*Geometric Style Augmentation.* Geometric augmentation is applied by randomly distorting the input pair $(\mathcal{I}, \mathbf{l}^*)$ of face image $\mathcal{I}$ and ground truth landmarks $\mathbf{l}^*$. Applying geometric deformations on the image itself may result in over-distortion of the facial features. Instead, we apply a set of simple random perturbations to alter $\mathbf{l}^*$, and receive the new landmark positions $\mathbf{l}'$. Then, we warp the Image $\mathcal{I}$ according to the new landmark positions $\mathbf{l}'$ using Thin Plate Spline (TPS) interpolation [Bookstein 1989]. This creates the geometric-style augmented version $(\mathcal{I}^g, \mathbf{l}')$ of the image and its landmarks. Geometric style augmentation is fast enough to be performed online during training.

To define the perturbation, we first determine the characteristic of each facial feature: eyes, nose, mouth, face, and use random deformations of the location, scale and aspect-ratio for each facial feature independently. The landmarks of each facial feature $f$ are transformed accordingly:

$$\mathbf{l}'_f = F(\mathbf{l}^*_f; s_{x_f}, s_{y_f}, t_{x_f}, t_{y_f})$$

where $\mathbf{l}^*_f$ is a subset of the ground truth landmarks, belonging to a certain facial feature $f$, $s_{x_f}, s_{y_f}$ and $t_{x_f}, t_{y_f}$ are the scaling and translation factors for the x and y coordinates of feature $f$ respectively. $F$ is a global scaling and translation transform, and $\mathbf{l}'_f$ is the resulting transformed landmarks of facial feature $f$. The scaling and translation factors are sampled from a uniform distribution with given lower and upper bounds for each transformation parameter (see our supplemental material for details).

### 5.3 Correction Step

At inference time, the image $\mathcal{I}$ is fed into the pre-trained NN to obtain the heat-maps $\mathcal{M} = \{\mathcal{M}^i\}$. The facial landmarks are initially located at the peak response positions in the heatmaps. These landmark predictions then go through a correction step by fitting them to a pre-trained point distribution model (PDM). The PDM models a shape $\mathbf{s}$ with global rigid transformations (scaling, in-plane rotation, and translation) as well as non-rigid variations (head poses and expressions). To calculate the PDM model, all the training shape are first aligned. The model is comprised of the mean shape $\bar{\mathbf{s}}$, and shape components $\Phi$. $\Phi$ is the collection of eigenvectors corresponding to the $m$ largest eigenvalues, which are obtained by applying PCA to the covariance matrix of the aligned training shapes. Following Zhang et al., we incorporate information obtained from the heatmaps into the PDM regularization process. We apply non-uniform regularization by considering the landmarks reliabilities. This leads to better outlier correction and produces a more robust shape correction.

Using one PDM imposes a global constraint on the shape $\mathbf{s}$ to correct all 68 facial landmarks. such regularization works well for natural faces, owing to the homogeneous nature of the human face. However, artistic portraits have significant variability in the facial features geometric location and proportion. Using one global constraint for all facial features will not allow much feature deviations from the canonical face shape. For this reason, we use a feature-based correction approach, i.e. we create a separate PDM model for each facial feature. At inference time, landmark subsets belonging to different facial features are fitted independently and create more accurate results (see Figure 7).

### 5.4 Tuning Step

After computing the initial shape $\mathbf{s}_0$ from the correction step, Zhang et al. propose a tuning step where the estimated shape is fine-tuned iteratively using weighted regularized mean shift, in which the confidence of the heatmap is integrated into a weighted version of the Regularized Landmark Mean-Shift (RLMS) [Saragih et al. 2011] framework. The tuning step alternates between computing the move step from response maps and regularizing it with the shape model's constraint (PDM).

For landmark detection in natural faces, the combination of the evidence from the response maps and the global parametric shape prior further improves the accuracy of facial landmark detection. For landmark detection in the artistic domain, using a global constraint for all facial features regularizes the output to the canonical face shape, and the detection performance deteriorates. For this reason, we use the tuning step to correct only the landmarks belonging to the outline of the face, i.e. landmarks belonging to the jaw ($EC_pT_{p:jaw}$) or the union of the jaw and eye brows ($EC_pT_{p:outline}$). The choice of a model can depend on the expected geometric style captured. While the first achieves better quantitative results, the second is better in capturing more extreme artistic styles (see 4th column in Figure 7).

## 6 EXPERIMENTS

### 6.1 Implementation

We implemented our NN using the Tensorflow framework[3]. The NN takes an input of a $256 \times 256$ face image and outputs a set of 68 heatmaps with the same resolution. During training, in addition to artistic augmentation we also use basic augmentation; we randomly

[3]http://tensorflow.org/


Fig. 7. Examples for Feature-based Landmark Correction. 1st and 3rd rows includes results of (**ECT + A**) using global correction and tuning, 2nd and 4th rows show results with part-based correction and Tuning ($\mathbf{EC_pT_p + A}$). By fitting each facial part independently, the landmarks can deviate from the canonical face shape, thus producing more accurate results.
From left to right: Portrait of the Mechanical, 1917 by Amedeo Modigliani courtesy WikiArt [Public Domain] via (http://bit.ly/2QbnSjL), Sitting Woman in a Green Blouse, 1913 by Egon Schiele courtesy WikiArt [Public Domain] via (http://bit.ly/2YDYKWb), The Annunciation, from The Life of the Virgin by Israhel van Meckenem courtesy The Met Collection [Public Domain] via (http://bit.ly/2VLmzhu), Portrait of Sawamura Tosho in the Role of Karukaya Doshin, 1834 by Utagawa Kunisada courtesy Van Gogh Museum [Public Domain] via (http://bit.ly/2VLmDxK).

flip the input image horizontally and crop a $248 \times 248$ arbitrary sub-image from it. Then, we rotate it with a random angle from $-30°$ to $30°$ before rescaling it back to $256 \times 256$. For the texture augmentation, we create $K = 9$ stylized copies for each input image, which is randomly selected out of the non-portrait training subset of WikiArt dataset. The parameters used for geometric style augmentation can be found in our supplemental material. The variance $\sigma$ of the 2D gaussians in the ideal heat-maps is set to 6. For training our NN, we use batch size of 6, and weight decay of $10^{-5}$, the learning rate is set to $10^{-4}$. We use ADAM optimization [Kingma and Ba 2015] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\hat{\epsilon} = 10^{-8}$. The network is trained for 115 epochs. Network weights are initialized using Xavier weight initialization [Y. and Bengio 2010]. We use the Menpo Project [2014] for performing image augmentation and landmark correction (PDM). For w-RLMS we use the code provided by [Zhang et al. 2018]. More implementation details can be found in our supplemental material. Our code will be publicly available.

## 6.2 Data Sets

*Natural Faces Data sets.* We use the 68-point annotations provided by Sagonas et al. [2013]. These annotations are provided for three existing in-the-wild datasets (LFPW [Belhumeur et al. 2013], HELEN [Le et al. 2012] and AFW [Zhu and Ramana 2012]), and the challenging dataset called IBUG. These annotations are split into the following subsets:

- the **training** set (3148 images) consisting of LFPW training images, HELEN training images, and AFW.
- the **challenging** subset (135) of IBUG.
- the **common** subset (554) of LFPW testing set, and HELEN testing set.
- the **full** set (689) contains the union of the above subsets.

The above annotations were actually provided as a training/validation set for the 300W face alignment competition, which used another set of images strictly for evaluation, called 300W test-set. The 300W test-set consists of 600 images split into two subsets, indoor and outdoor, which are said to have been drawn from a similar distribution as the IBUG dataset. We refer to the 300W test-set as the test subset. We use the training subset to train our algorithm, while the full, challenging, common and test subsets are used for evaluation.

*Artistic Data sets.* For texture style augmentation, we utilized the non-portrait subset of the PBN dataset (introduced in Section 3). Out of the non-portrait subset, we remove all artworks of the artists included in the "Artistic-Faces" dataset. This includes around 84,000 images of various styles and genres. We split these images into training art dataset of (80% of the images), which is used to texture-augment the training set, and evaluation art dataset (20% of the images), which is used to texture-augment the full, challenging, common and test sets. We use the "Artistic-Faces" dataset presented in Section 3 for evaluation. We also use the caricature subset of the WebCaricature dataset [Huo et al. 2018], containing 6042 caricatures, labeled with 17 facial landmarks for evaluating our framework.

## 6.3 Evaluation Metrics

For fair comparison, the evaluation metrics are chosen as the common protocols in the literature ([Ren et al. 2016], [Zhu et al. 2016], [Lv et al. 2017]). The primary metric is the Normalized Mean Error (NME), which could be calculated as $\frac{1}{n}\sum_{i=1}^{n}\frac{\|\mathbf{x}_i - \mathbf{x}_i^*\|_2}{d}$, where $d$ denotes the normalized distance and $n$ is the number of facial landmarks involved in the evaluation.

To maintain consistency with previous works we report our results using two different normalizing distances, inter-ocular distance i.e. the distance between the outer eye corners, and inter-pupil distance i.e. the distance between the eye centers. In the artistic faces domain where the different facial features can be localized in various ways, the distance between eyes is less suitable. Therefore, we also report our results using the diagonal of the bounding box as the normalizing distance. We also use Cumulative Error Distribution (CED) curve to compare our results to previous works. In the supplemental material Area Under the Curve (AUC) @ 0.08 error and failure rate are also reported.

## 6.4 Comparison to State-of-the-Art

*Results on natural faces (NF).* In Table 1 we compare our approach with recently proposed state of-the-art algorithms ([Dong et al. 2018], [Kowalski et al. 2017], [Zhang et al. 2018]) on natural face

Table 1. Comparisons of inter-ocular Normalized Mean Error (%) on the 300W test sets of natural faces (NF). Last two rows are our method.

| Method | Common | Challenging | Full |
|---|---|---|---|
| MDM [2016] | 4.83 | 10.14 | 5.88 |
| SAN [2018] | 3.41 | 7.55 | 4.24 |
| DAN [2017] | 3.19 | 5.24 | 3.59 |
| DAN-MENPO [2017] | **3.09** | **4.88** | **3.44** |
| ECT [2018] | 4.66 | 7.96 | 5.31 |
| $\mathbf{EC_pT_p}$ + **A** (ocular) | 3.29 | 6.34 | 3.89 |
| $\mathbf{EC_pT_p}$ + **A** (pupil) | 4.56 | 9.16 | 5.46 |

images. We use the detector cropped bounding-boxes of the 300W test sets as provided by [Sagonas et al. 2013]. For this comparison, the error is normalized by the inter-ocular distance to maintain consistency with the 300W competition. Since [Zhang et al. 2018] provided results normalized by the inter-pupil distance, we provide this measure also for our method. Compared to the ECT method, the performance on natural faces deteriorates by 15% on the challenging set and by 3% on the full set, however we improve the performance on the common set by 2%. Our method ($\mathbf{EC_pT_p}$ + **A**) achieves results that are comparable with recent state-of-the-art methods. This implies that adding artistic augmentation to the training procedure, and a part-based correction for inference, enables using the same framework for detecting facial landmarks on both natural and artistic faces, and can be used as a detection component in cross-domain applications (see Section 7 for potential applications).

*Results on natural faces with artistic augmentation (NF+A).* When adding artistic augmentation to the test set to mimic real art portraits, our method performs better than state-of-the-art methods. We use the 300W test sets, with artistic augmentation, and the error is normalized by the inter-ocular distance. For the augmented test sets we use ground truth bounding box to crop the input images, calculated using the minimum and maximum coordinates of ground truth face landmarks. We compare our approach with best performing state of-the-art algorithms in Table 2 ([Dong et al. 2018], [Kowalski et al. 2017], [Zhang et al. 2018]). To obtain results of these methods we use the official implementation released by the authors. For our methods ($ECT + A$, $EC_pT_p + A$), we used our own implementation. Our method outperforms the other methods because landmark detection methods for natural faces are trained on natural face data, and therefore deals poorly with the wide range of variations in texture and geometry of the artistically augmented test sets. In addition, many landmark detection algorithms use a shape prior or regularizer, which constrains the landmarks shape to the canonical human face and thus deteriorate their performance in artistic-style inputs.

*Results on Artistic-Faces.* We show the performance of different facial landmark detection algorithms on real artistic portraits in Table 3 using the NME measure. We use the Artistic-Faces dataset (Section 3) using ground truth bounding box to crop the input images, calculated using the minimum and maximum coordinates of ground truth landmarks. We use 3 different normalization methods for comparison; inter-ocular distance, inter-pupil distance and the

Table 2. Comparisons of inter-ocular Normalized Mean Error (%) on the 300W test sets with artistic augmentation (NF+A). Last two rows are our method.

| Method | Common-A | Challenging-A | Full-A | Test-A |
|---|---|---|---|---|
| DAN | 16.05 | 26.26 | 18.05 | 18.66 |
| DAN-MENPO | 15.74 | 24.89 | 17.54 | 18.32 |
| SAN | 9.74 | 22.65 | 12.27 | 14.50 |
| MDM | 7.07 | 17.00 | 9.01 | 10.67 |
| ECT | 6.26 | 17.51 | 8.46 | 10.52 |
| ECT+A | 3.98 | 11.05 | 5.36 | 5.82 |
| $\mathbf{EC_pT_{p:jaw}}$ + **A** | **3.67** | 10.88 | **5.08** | **5.52** |
| $\mathbf{EC_pT_{p:out}}$ + **A** | 3.69 | **10.80** | 5.09 | 5.55 |

Table 3. Comparisons of Normalized Mean Error (%) on the Artistic-Faces dataset (AF). Last two rows are our method.

| Method | inter-pupil NME | inter-ocular NME | BB diagonl NME |
|---|---|---|---|
| MDM | 8.04 | 5.63 | 2.47 |
| SAN | 8.03 | 5.60 | 2.49 |
| DAN | 7.43 | 5.21 | 2.27 |
| DAN-MENPO | 7.25 | 5.08 | 2.25 |
| ECT | 7.44 | 5.21 | 2.32 |
| ECT+A | 6.89 | 4.81 | 2.14 |
| $\mathbf{EC_pT_{p:jaw}}$ + **A** | 6.57 | 4.59 | 2.04 |
| $\mathbf{EC_pT_{p:out}}$ + **A** | **6.47** | **4.52** | **2.01** |

diagonal of the ground truth bounding-box. Similar to the NF+A test sets, our method outperforms state-of-the-art methods that were trained on natural face data. Note that artistic augmentation can be added to the training procedure of any algorithm to enhance performance on artistic portraits. In addition, any global shape regularizer or prior, should be modified accordingly for further improvement. We also provide AUC and failure rate comparisons in the supplemental material.

To examine another artistic domain we evaluate our model on the caricature subset of the WebCaricature dataset [Huo et al. 2018]. For this purpose, a subset of 16 out of 68 landmarks is used. One landmark (top of hairline) cannot be obtained out of the 68 landmarks model [Sagonas et al. 2013], and is removed from the ground-truth annotation for comparison. Our method $EC_pT_p + A$ achieves inter-ocular NME of 10.34%, compared to 13.23% using the *ECT* method [Zhang et al. 2018].

## 6.5 Ablation Study

We verify the significance of each component in our proposed method for landmark detection in art. Figure 8 shows the comparison regarding CED curves of the ECT [2018], ECT+A and $EC_pT_p + A$ methods. As can be seen, there is a performance improvement by adding each one of the components to the landmark detection framework. Artistic augmentation ("+**A**") increases the robustness of the neural network to a wide variety of textures and geometric input styles, and the part-based landmark correction ("$\mathbf{C_pT_p}$") further

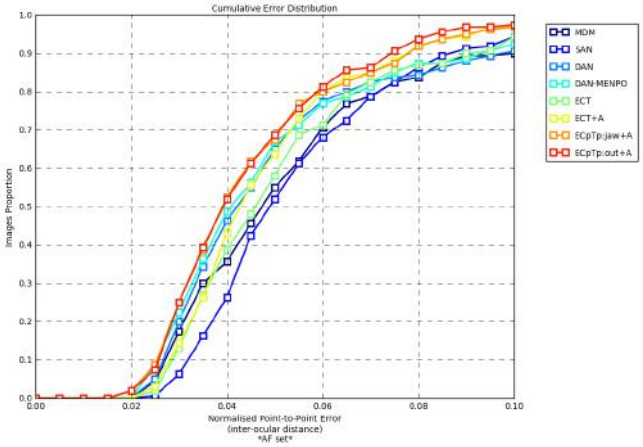



Fig. 8. Comparison of cumulative errors distribution (CED) curves on the Artistic-Faces dataset (normalized by inter-ocular distance)

increases the performance by reducing the dependency between the different facial features, which is necessary due to position and shape variations of facial features in artworks.

We examine the effect of adding each Artistic Augmentation component separately: Geometric and Texture-style augmentation to the Network training. To isolate the effect of each component, we conducted 4 tests which differ only in the type of augmentation. In all tests we use the exact same parameters to train the Heat-Map Network, and only change the type of augmentation performed on the training set. We refer to each test by the name of its augmentation; **Basic**, **Texture**, **Geometry** and **Geometry+Texture (G+T)** (see Figure 9). We report the test results in Table 4. For each augmentation type we report the results of the network predictions (E), as well as the results of the full *ECT* method and our $EC_pT_p$ algorithm. For this comparison we use our own implementation of *ECT*. As seen in Table 4, using only Texture augmentation improves the performance compared to using basic augmentation. Interestingly, it shows that using only geometric augmentation deteriorates the performance compared to using basic augmentation. We believe that this is due to the fact that the network "struggles" to deal with unknown textures. When omitting texture augmentation, it is beneficial for the network to learn the canonical face, so it can "expect" the facial features to appear in certain areas of the input image, and by that overcome uncertainties that arises from new input textures. The best results are achieved using both texture and geometric augmentation in the training procedure.

Facial landmark detection is a mature field and many algorithms achieve impressive results, it is therefore difficult to attain large improvement over state-of-the-art methods in terms of average performance. However, examining the results qualitatively reveals a significant improvement in capturing the geometric style of portraits. Figure 10 shows a number of results of our proposed method ($\mathbf{EC_pT_p + A}$) on the Artistic-Faces dataset. In Figure 7 a detailed example was shown where similar average distances actually have significant qualitative differences in matching the facial features' shape.

Table 4. Comparisons of Normalized Mean Error (%) on the Artistic-Faces dataset (AF) using different Augmentation strategies.

| Method | E | ECT | $EC_pT_p$ |
|---|---|---|---|
| Basic | 5.93 | 5.33 | 5.37 |
| Geometry | 6.75 | 5.60 | 5.93 |
| Texture | 5.10 | 4.84 | 4.63 |
| G+T | **5.06** | **4.81** | **4.52** |

*Note:* We refer to each test by the name of its augmentation; **Basic**, **Texture**, **Geometry** and **Geometry+Texture (G+T)**.


Fig. 9. Training images with different Augmentation strategies. these datasets were created to examine the effect of adding each Artistic Augmentation element, i.e. Geometric and Texture style augmentation to the Network training procedure. leftmost image from [Minear and Park 2004], used with permission.

Figure 11 shows a comparison of artists landmark distributions using different methods for detection. Adding artistic augmentation to the training procedure and part-based correction allows for further deviation from the canonical face, resulting in better modelling of the artist's geometric style. Compared to the ECT algorithm, each of the added components to the framework brings the predicted artist distribution closer to the ground truth artist distribution. In Table 5 we verify quantitatively the advantage of using our method vs. ECT . We measure the average KL divergence distance between the distributions of feature locations (as in Figure 11) comparing the ground truth to two landmark detection methods: using ECT, and our proposed method ($\mathbf{EC_pT_p + A}$). Our method shows improvement for almost all artists. By achieving better modelling of the artist geometric style, our method can be used for style analysis and synthesis applications as presented in the next section.

## 7 APPLICATIONS

Our method for facial landmark detection in art allows both the analysis and synthesis of artistic portraits. In this section, we present some applications for our method.

To analyze a set of novel artistic portraits, we use the same preprocessing depicted in Section 3 to obtain a set of 256×256 images for any desired artist(s) $\{\mathcal{I}_{artist}\}$. Next, we use our automatic landmark detection method to detect facial landmarks on each portrait and represent them as the landmark vector $\mathbf{l}_{artist}$. This creates a an annotated dataset $D_{artist} = \{(\mathcal{I}_{artist}, \mathbf{l}_{artist})\}$.

### 7.1 Artist's Geometric Style

By analyzing the geometric information in $\{\mathbf{l}_{artist}\}$, we can model the shape and proportions of a specific artist or a specific set of

Table 5. KL divergence on the Artistic-Faces dataset (AF) compared to Ground-Truth distributions

| Method | Modigliani | Comics | Schiele | Leger | Matisse | Hindu | Meckenem | Bratby | Chagall | Kisling | Dix | Picasso | Raphael | Lichtenstein | Kunisada | Gogh |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ECT | 0.46 | 1.61 | 0.68 | 0.54 | 0.95 | 0.45 | 1.34 | 2.24 | 1.44 | **0.31** | 3.82 | 0.99 | 1.22 | 0.36 | 2.34 | 0.85 |
| $EC_pT_p$ + A | **0.28** | **0.7** | **0.52** | **0.29** | **0.36** | **0.32** | **0.59** | **1.05** | **0.5** | 0.36 | **0.61** | **0.65** | **0.68** | **0.26** | **1.04** | **0.37** |


Fig. 10. Example results on the Artistic-Faces dataset. 1st row includes the result of the ECT algorithm [2018], 2nd row includes results of the ECT algorithm with Artistic Augmentation (ECT+A), 3rd row includes results of our proposed method ($EC_pT_p + A$).
From left to right: Portrait of Paul F. Schmidt, 1921 by Otto Dix courtesy WikiArt [Public Domain US] via (http://bit.ly/2LRIWxx), Girl with A Black Cat, 1910 by Henri Matisse courtesy WikiArt [Public Domain US] via (http://bit.ly/2W7B1jr), Spaniard, 1906 by Pablo Picasso courtesy WikiArt [Public Domain US] via (http://bit.ly/2VPxQha), Self-Portrait, 1889 by Vincent van Gogh courtesy NGA Images [Public Domain] via (http://bit.ly/2WSdpN0), Saint Stephen by Israhel van Meckenem courtesy NGA Images [Public Domain] via (http://bit.ly/2Hrt7b6), Special Exhibition of Buddhist Icons at the Tenmangu Shrine, 1849-1851 by Utagawa Kunisada courtesy Van Gogh Museum [Public Domain] via (http://bit.ly/2JLi53f).

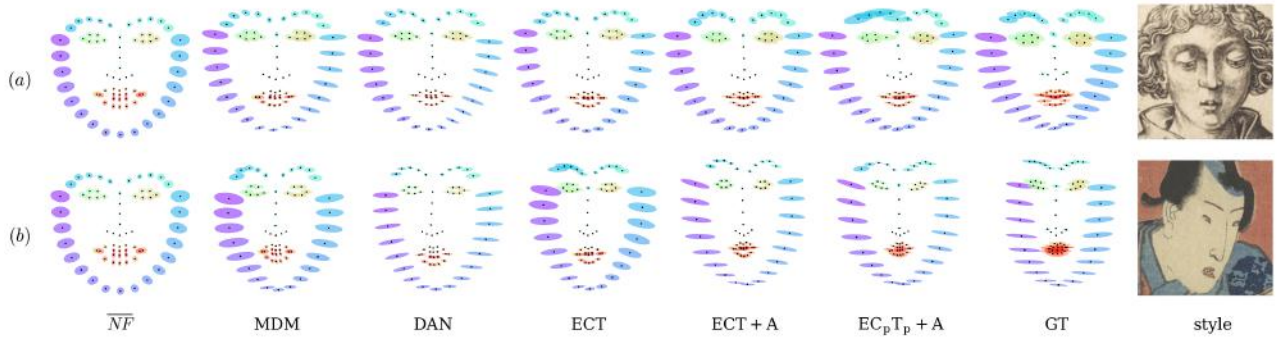



Fig. 11. Comparison of artist distributions using different methods for landmark detection. (a) contains artwork distributions of Israhel van Meckenem. (b) contains artwork distributions of Utagawa Kunisada. Adding Artistic Augmentation to the training procedure (ECT+A) results in a network more robust to variations in texture and geometry, hence producing more accurate results. Adding Part-based Correction ($EC_pT_p + A$) allows for further deviation from the canonical face and allows to capture the specific artist's geometric style (compare to GT, the ground truth).
Face in the 2nd row taken from Tasogare, 1830-1839 by Utagawa Kunisada courtesy Van Gogh Museum [Public Domain] via (http://bit.ly/2YxsHqH).

images. This information can be used for portrait style classification, finding similarities between different artists, and serve as a tool for learning artistic portrait style. This information can also be used as the input for generative algorithms (see Section 7.4).

To define a mathematical model of the geometric style of an artist (based either on a portrait or a set of portraits), we compare the portraits' landmarks $\mathbf{l}_{artist}$, to the landmarks of a natural "average-face" $\mathbf{l}_{avg}$. The average-face landmarks' positions are computed

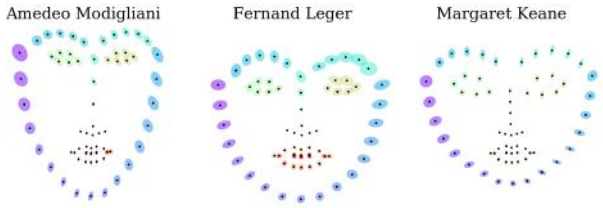



Fig. 12. Artist distributions using our proposed method for landmark detection in art. Our framework is able to capture a wide variety of artistic styles. Observing artist distributions, we can identify the signature style of each artist; an elongated face and nose for Modigliani, a short face with wide chin and large distance between the eyes for Leger, a small face with large eyes for Keane. For more distribution examples see our supplemental material.

from the 300-W training set. By computing the offsets of the portraits from the "average-face" we are able to capture the artists' facial feature variations. We uniformly normalize $\{\mathbf{l}_{artist}\}$ and $\mathbf{l}_{avg}$ dividing by the image size without changing face proportions or rotation, and align their center points (a point on the tip of the nose). For each pairs of matching landmark points we define an offset vector by measuring the difference of their positions: $v_i = (x_i - z_i)$, where $x_i \in \mathbf{l}_{artist}, z_i \in \mathbf{l}_{avg}$. Using the collection of offset vectors $\{\mathbf{v}_{artist}\}$, we calculate the mean vector for a set of portraits (for example, of a specific artist) $\mu_{artist}$ and the covariance matrix $\Sigma_{artist}$ of the reshaped artist offset vectors $\{\mathbf{v}'_{artist}\} \in \mathbb{R}^{1\times 136}$. Lastly, similar to [Berger et al. 2013] we fit a Multivariate Gaussian Model to the artist data $\mathcal{P}_{artist} \sim \mathcal{N}(\mu_{artist}, \Sigma_{artist})$.

Using $\{\mathbf{l}_{artist}\}$ we can visualize the distribution of facial landmarks of a specific artist. As illustrated in Figure 12, such visualizations can highlight the differences in the geometric style of different artists. Observing these distributions, we can identify the style of artists: an elongated face and nose for Modigliani, a small face with large eyes for Keane, etc. For more geometric distribution examples see the supplemental material.

## 7.2 Average-Portraits

Using our detection framework to obtain an annotated dataset $D_{artist} = \{(\mathcal{I}_{artist}, \mathbf{l}_{artist})\}$, we can calculate the artist mean facial shape $\mu_{artist}$ (similar to Section 7.1), and create average portraits representing different artists. To create an average portrait, we simply warp the collection of annotated portraits $D_{artist}$ to the artists mean shape $\mu_{artist}$, and calculate the mean RGB values of the warped portraits collection.

Figure 13 shows average portraits of artist collections included in the Artistic Faces Dataset (10 portraits each). For comparison, we also show the average portraits obtained using ground-truth landmarks, landmarks obtain with ECT, and by calculating the mean RGB values of the original portraits, without any alignment. Our framework enables creating an average portrait which is meaningful, informative and representative of both texture and geometric style of the artist. Without any alignment, we achieve average portraits which are blurry and ambiguous. Comparing to the average portraits obtained using the ECT framework, our portraits are closer to the ground-truth average portraits, while the ECT portraits tend to be blurry due to inaccurate landmark detection. This stresses the importance of capturing the fine details, enabled by our method.

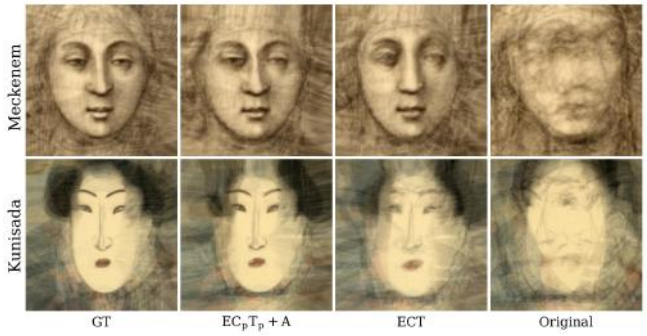



Fig. 13. Average Portraits. In the 1st column, ground-truth landmarks are used to create average portraits by aligning the facial features, in the 2nd column we use landmarks obtained from our frameworks, in the 3rd column we show average portraits obtained using the ECT framework, the 4th column contains average portraits obtained by calculating the mean RGB values of the original portraits, without any alignment. Our framework enables creating an average portrait which is representative of both texture and geometric style of the artist.

## 7.3 Style Signatures

Using our observations from section 4, and several sizes and ratios described in literature we define a vector of 99 dimensions as the *geometric* style signature of a portrait. This signature vector includes the following set of values. Feature aspect-ratios are computed as $f_H : f_W$, for any facial feature $f$, where $f_H$ and $f_W$ represent the height and width of the feature's bounding box. For all pairs of features $f^1, f^2$, the relative proportion are computed as $f^1_H : f^2_H$, $f^1_W : f^2_W$ and $f^1_H : f^2_W$. The relative location of all inner facial features (mouth, eyes and eyebrows) are calculated as the normalized difference between the feature center point $\bar{f}$ and the face center point $O$ in both axes: $FW^{-1}\left|\bar{f}_x - O_x\right|, FH^{-1}\left|\bar{f}_y - O_y\right|$, where $FW = \max(\mathbf{l}_x) - \min(\mathbf{l}_x)$ and $FH = \max(\mathbf{l}_y) - \min(\mathbf{l}_y)$.

Other values included in the geometric style signature are facial proportions that have been reported to have correlation with beauty and attractiveness (see full list in the supplemental material). We use two approaches proposed in literature: Neoclassical Canons [Farkas et al. 1985], and Golden Ratios [Schmid et al. 2008].

For a holistic representation of artistic style, we combine the geometry signature with texture style embedding proposed by [Gatys et al. 2015]. To balance the effect of each component, we use principal component analysis (PCA) to reduce the dimensionality of each representation vector. The reduced texture and geometry style vectors are then concatenated into a single style signature. Figure 14 shows a visualization of the embedding of style signatures of the Artistic-Faces dataset using T-SNE. In the left subfigure we use the texture style representation proposed by Gatys et al. In the right subfigure we use our combined texture and geometry signature.

Adding geometric information to the texture style embedding reveals differences and similarities that were not captured by texture alone, and allows better examination of artists style similarities.

### 7.4 Geometry-Aware Style Transfer

Most style transfer methods deal only with texture style transfer, without considering the artistic geometric style. In this section, we present a method for Geometry + Texture style transfer and show results of various artistic models. Our geometry-aware style transfer is performed using a portrait image bank for texture style transfer, and an artist-specific geometric-style model for geometric style transfer.

To stylize a natural face image $\mathcal{I}$ of arbitrary size, we use our landmark detection framework to extract facial landmarks $\mathbf{l}$, matching to the face crop of input image $\mathcal{I}$.

For geometric stylization, we will use an artists' portrait image bank $\{\mathcal{I}_{artist}\}$, and build the geometric style model $\mathcal{P}_{artist}$ described in Section 7.1. $\mathcal{P}_{artist}$ is then sampled to produce an offset vector $\mathbf{v}'_{artist}$. Each landmark $\mathbf{l}$ is aligned to $\mathbf{v}'_{artist}$ by its center point, and then perturbed by simply adding the matching offset to its point: $l'_i = (x_i + v'_i)$, where $x_i \in \mathbf{l}, v'_i \in \mathbf{v}'_{artist}$. The whole input image $\mathcal{I}$ is stylized by warping to the stylized landmarks $\mathbf{l}'$ using TPS interpolation, resulting in $\mathcal{I}^g$ - a geometrically stylized image.

For texture style transfer we use the algorithm proposed by [Gatys et al. 2015] described in section 5, where $\mathcal{I}^g$ serves as the content image, and a random sample from $\{\mathcal{I}_{artist}\}$ serves as the style image, resulting in a portrait that is stylized in both geometry and texture $\mathcal{I}^{g+t}$.

Figure 16 and 1 show example results of our proposed application for geometry-aware style transfer. We show stylization results of the same input image, using seven different artistic style models. The first row contains the results of the geometric stylization stage ($\mathcal{I}^g$). For comparison, the second row contains the results of using the algorithm of [Gatys et al. 2015] on the input image, without performing geometric stylization ($\mathcal{I}^t$). The third row contains the stylization results of our geometry + texture application for style transfer ($\mathcal{I}^{g+t}$). By combining both textural and geometrical elements for modelling artistic style, we achieve image stylization which is more visually appealing, maintains higher artistic credibility, and present higher variation between images stylized using different artistic models. Our method also enables combining geometry and texture style components of different artists, creating stylized images with inventive new styles (see Figure 18).

Our method's ability to capture fine details in artistic portraits allows to model the artist's geometric style more accurately. This is demonstrated qualitatively in Figure 13, where artists portraits are much better aligned to produce an average protrait (compare our framework $EC_pT_p + A$ to $ECT$). We can also produce stylization results that are more similar geometrically to original artworks. In Figure 15 we compare geometric sytlization of a portrait based on models of the artist Utagawa Kunisada, using our landmark detection $EC_pT_p + A$ vs. $ECT$ landmark detection of [Zhang et al. 2018].

The images created using the ECT-model, partially captures Kunisada's geometric style (elongated face). However, the inner facial features are drawn closer to the canonical face, and key style elements are lost compared to our model (e.g. small slanted eyes, small nose-to-mouth distance, small mouth etc.). See our supplemental material for additional stylization examples.

## 8 CONCLUSIONS

In this work, we have presented the first Facial Landmark Detection framework for general-style Artistic Portraits. To achieve that, we explored the differences between the natural and artistic faces domains. Using these observations, we proposed a method for Artistic Augmentation, which brings the natural face domain closer to the artistic one. Such augmentation provides the means to learn a detection algorithm that is more robust to variations in texture and geometry. The algorithm uses a feature-based correction and tuning steps, to reduce the dependency between the different facial features, allowing the landmarks to deviate from the canonical face shape. Our method outperforms state-of-the-art methods of landmark detection on the Artistic Faces dataset.

We demonstrated several applications for our method. We presented a method for analyzing the geometric style of artists, and defined a style signature for portrait artworks containing both a geometric and a texture-based parts. We also presented a synthesis application for geometry-aware portrait style transfer.

*Limitations:* Our method cannot yet handle strong shape variations, where the facial features themselves have a distinctly different shape than natural facial features, such as in Manga or cartoons (see Figure 17). In the future, we plan to adapt our framework to styles with strong shape variations. We also plan to investigate the possibility of using geometric style analysis and style signatures for other applications such as classification and matching of artists and portraits.

## ACKNOWLEDGMENTS

The authors would like to thank the reviewers and Michael Rubinstein for many suggestions. This work was partially supported by the Israel Science Foundation grant number 2216/15 and a grant from Amazon for web services.

## REFERENCES

J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. 2014. Menpo: A Comprehensive Platform for Parametric Image Alignment and Visual Deformable Models. In *Proceedings of the ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 679–682.

P.-N. Belhumeur, D.-W. Jacobs, D.-J. Kriegman, and N. Kumar. 2013. Localizing parts of faces using a consensus of exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 12 (2013), 2930–2940.

Itamar Berger, Ariel Shamir, Moshe Mahler, Elizabeth Carter, and Jessica Hodgins. 2013. Style and Abstraction in Portrait Sketching. *ACM Transactions on Graphics 32(4) (SIGGRAPH Conference Proceedings)* 32, 4, Article 55 (2013), 12 pages.

F.-L. Bookstein. 1989. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 6 (1989), 567–585.

A. Bulat and G. Tzimiropoulos. 2016. Convolutional aggregation of local evidence for large pose face alignment. In *Proceedings of the British Machine Vision Conference (BMVC)*.

A. Bulat and G. Tzimiropoulos. 2017. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 1021–1030.


Fig. 14. T-SNE visualization of the Artistic-Faces dataset using two different style embedding: only texture-based (left) and adding our geometric style representation (right). Notice that by adding geometric information, artworks of artists with consistent geometric style such as Amedeo Modigliani (red points) are clustered closer together. His artworks (c) and (d) are clustered together on the right even though they contain different color schemes. On the other hand, artworks of artists with inconsistent geometric style get more scattered in the embedding space on the right. Otto Dix (yellow points) has a "caricature-like" geometric style, exaggerating features based on the subjects. This geometric style is inconsistent and varies between artworks. More observations on these charts can be found in the supplemental material.

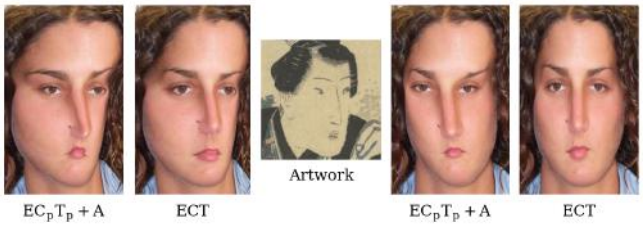



Fig. 15. Geometric Stylization. For this example, we built 2 geometric models of the artist Utagawa Kunisada, using the subset of his artworks out of the Artistic-Faces Dataset. In the images stylized by the ECT-based model, the inner facial features are drawn closer to the canonical face, and by that some of the key style elements are lost (e.g. small slanted eyes, small nose-to-mouth distance, etc.). Our model successfully captures the geometric style of the artist, and produces geometrically stylized images which are closer to the original artworks.

Kaidi Cao, Jing Liao, and Lu Yuan. 2018. CariGANs: Unpaired Photo-to-caricature Translation. In *SIGGRAPH Asia 2018 Technical Papers (SIGGRAPH Asia '18)*. Article 244, 14 pages.

X. Cao, Y. Wei, F. Wen, and J. Sun. 2014. Face Alignment by Explicit Shape Regression. *Proceedings of the International Journal of Computer Vision (IJCV)* 107, 2 (2014), 177–190.

W.-T. Chu and W.-W. Li. 2017. Manga FaceNet: Face Detection in Manga based on Deep Neural Network. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*. 412–415.

T.-F. Cootes and C.-J. Taylor. 1992. Active Shape Models - 'Smart Snakes'. In *Proceedings of the British Machine Vision Conference (BMVC)*. 266–275.

P. Dollar, P. Welinder, and P. Perona. 2010. Cascaded Pose Regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1078–1085.

X. Dong, Y. Yan, W. Ouyang, and Y. Yang. 2018. Style Aggregated Network for Facial Landmark Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 379–388.

L.-G. Farkas, T.-A. Hreczko, J.-C. Kolar, and I.-R. Munro. 1985. Vertical and Horizontal Proportions of the Face in Young Adult North American Caucasians: Revision of Neo-classical Canons. *Plastic and Reconstructive Surgery (PRS)* 75, 3 (1985), 328âĂŞ337.

L.-A. Gatys, A.-S. Ecker, and M. Bethge. 2015. A Neural Algorithm of Artistic Style. (2015). arXiv:arXiv:1508.06576

G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens. 2017. Exploring the structure of a Real-time, Arbitrary Neural Artistic Stylization Network. In *Proceedings of the British Machine Vision Conference (BMVC)*.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative Adversarial Nets. In *Proceedings of the Neural Information Processing Systems (NIPS)*. 2672–2680.

Jing Huo, Wenbin Li, Yinghuan Shi, Yang Gao, and Hujun Yin. 2018. WebCaricature: a benchmark for caricature recognition. In *British Machine Vision Conference*.

P.-T. Jackson, A. Atapour-Abarghouei, S. Bonner, T. Breckon, and B. Obara. 2018. Style Augmentation: Data Augmentation via Style Randomization. (2018). arXiv:arXiv:1809.05375

M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. 2015. Spatial Transformer Networks. In *Advances in Neural Information Processing Systems (NIPS)*. 2017–2025.

S. Jha, N. Agarwal, and S. Agarwal. 2018. Bringing Cartoons to Life: Towards Improved Cartoon Face Detection and Recognition Systems. (2018). arXiv:arXiv:1804.01753

Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song. 2017. Neural Style Transfer: A Review. (2017). arXiv:arXiv:1705.04058

J. Johnson, A. Alahi, and L. Fei-Fei. 2016. Perceptual Losses for Realtime Style Transfer and Super-Resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 694–711.

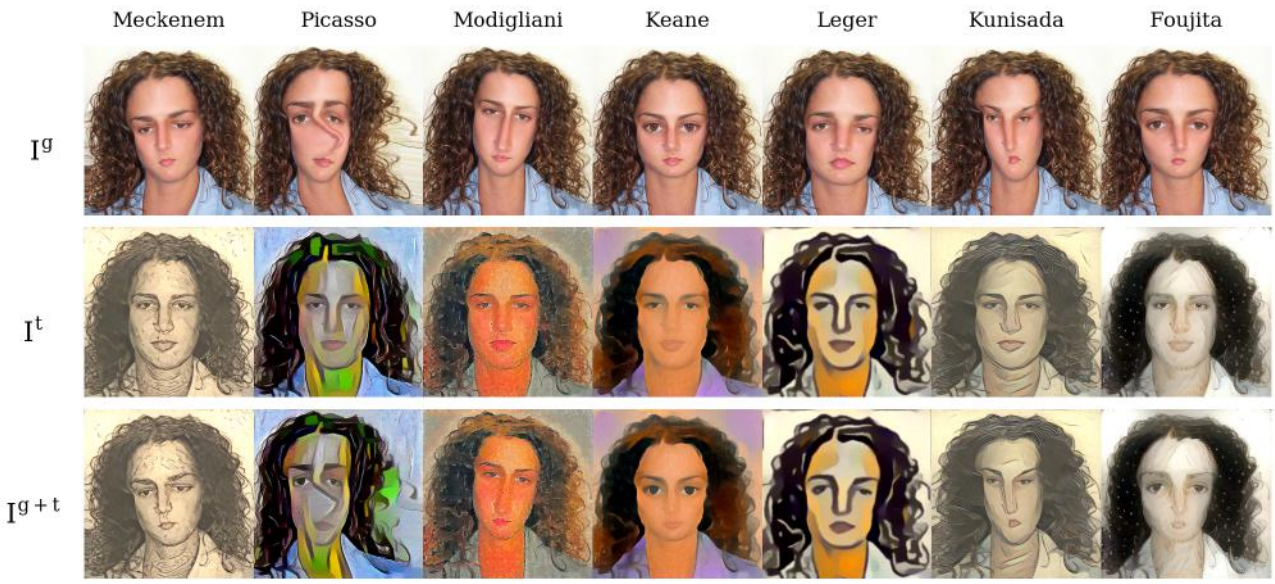



Fig. 16. Geometry + Texture Style Transfer. 1st row contains the results of the geometric stylization stage ($I^g$). 2nd row contains the results of using the algorithm of [Gatys et al. 2015] on the input image, without performing geometric stylization ($I^t$). 3rd row contains the stylization results of our g+t style transfer application ($I^{g+t}$). By combining both textural and geometrical elements for modelling artistic style, we achieve image stylization which is more visually appealing, maintains higher artistic credibility, and present higher variation between images stylized using different artistic models.

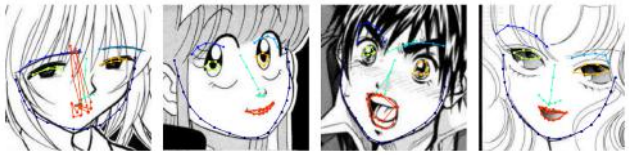


Fig. 17. Limitations: our method cannot yet handle very large shape variations, where the facial features themselves have a distinctly different shapes than natural facial features (images by: Yagami Ken, Shinzawa Motoei, Deguchi Ryusei, Aida Mayumi. from [Ogawa et al. 2018]), used with permission.

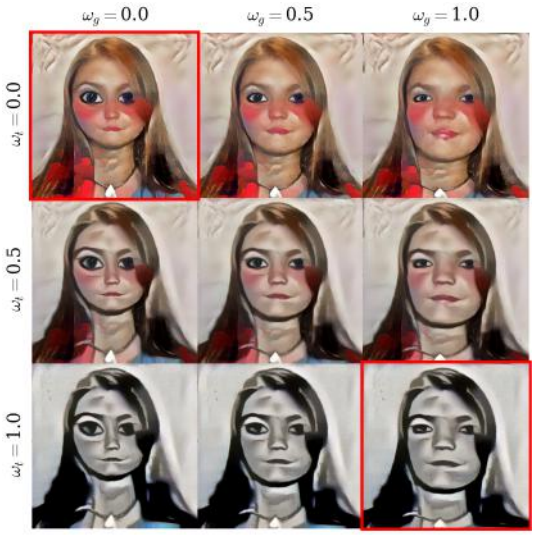



Fig. 18. Style Interpolation Combinations: each image corresponds to a different set of style weights ($\omega_g$, $\omega_t$), combining the geometric and texture style of the two artists - Keane and Leger (marked with red rectangles).

Parneet Kaur, Hang Zhang, and Kristin J. Dana. 2017. Photo-Realistic Facial Texture Transfer. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2017), 2097–2105.

V. Kazemi and J. Sullivan. 2014. One Millisecond Face Alignment with an Ensemble of Regression Trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1867–1874.

Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2015).

M. Kowalski, J. Naruniec, and T. Trzcinski. 2017. Deep Alignment Network: A Convolutional Neural Network for Robust Face Alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2034–2043.

V. Le, J. Brandt, Z. Lin, L. Bourdev, and T.-S. Huang. 2012. Interactive facial feature localization. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 679–692.

C. Li and M. Wand. 2016. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2479–2486.

S. Li, X. Xu, L. Nie, and T.-S. Chua. 2017. Laplacian-steered Neural Style Transfer. In *Proceedings of ACM on Multimedia Conference (ACM)*. 1716–1724.

J. Lv, X. Shao, J. Xing, C. Cheng, and X. Zhou. 2017. A Deep Regression Architecture with Two-Stage Re-initialization for High Performance Facial Landmark Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3691–3700.

M. Minear and D. Park. 2004. A Lifespan Database of Adult Facial Stimuli. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36 (12 2004), 630–3.

N.-V. Nguyen, C. Rigaud, and J.-C. Burie. 2017. Comic characters Detection using Deep Learning. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 3. 42–46.

Toru Ogawa, Atsushi Otsubo, Rei Narita, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2018. Object Detection for Comics using Manga109 Annotations. *CoRR* abs/1803.08670 (2018).

S. Ren, X. Cao, Y. Wei, and J. Sun. 2016. Face Alignment via Regressing Local Binary Features. *IEEE Transactions on Image Processing* 25, 3 (March 2016), 1233–1245.

C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 2013. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 397–403.

J.-M. Saragih, S. Lucey, and J.-F. Cohn. 2011. Deformable Model Fitting by Regularized Landmark Mean-Shift. *Proceedings of the International Journal of Computer Vision (IJCV)* 91, 2 (2011), 200–215.

K. Schmid, D. Marx, and A. Samal. 2008. Computation of face attractiveness index based on neoclassic canons, symmetry and golden ratio. *Pattern Recognition* 41, 8 (2008), 2710–2717.

Ahmed A. S. Seleim, Mohamed A. Elgharib, and Linda Doyle. 2016. Painting style transfer for head portraits using convolutional neural networks. *ACM Trans. Graph.* 35 (2016), 129:1–129:18.

Y. Shi, D. Deb, and A.-K. Jain. 2018. WarpGAN: Automatic Caricature Generation. (2018). arXiv:arXiv:1811.10100

M. Stricker, O. Augereau, K. Kise, and M. Iwata. 2018. Facial Landmark Detection for Manga Images. (2018). arXiv:arXiv:1811.03214

W. Sun and K. Kise. 2010. Similar Partial Copy Detection of Line Drawings Using a Cascade Classifier and Feature Matching. In *Proceedings of the International Workshop on Computational Forensics (IWCF)*. 121–132.

Y. Sun, X. Wang, and X. Tang. 2013. Deep Convolutional Network Cascade for Facial Point Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3476–3483.

G. Trigeorgis, P. Snape, M. Nicolaou, E. Antonakos, and S. Zafeiriou. 2016. Mnemonic Descent Method: A Recurrent Process Applied for End-to-End Face Alignment. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4177–4187.

D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky. 2016. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In *Proceedings of the International Conference on Machine Learning (ICML)*. 1349–1357.

X. Xiong and F. De la Torre. 2013. Supervised Descent Method and its Applications to Face Alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 532–539.

X. Glorot Y. and Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, Vol. 9. 249–256.

H. Yanagisawa, D. Ishii, and H. Watanabe. 2014. Face Detection for Comic Images with Deformable Part Model. In *Proceedings of the Image Electronics and Visual Computing Workshop (IIEEJ)*.

J. Yang, Q. Liu, and K. Zhang. 2017. Stacked Hourglass Network for Robust Facial Landmark Localisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025–2033.

X. Yu, F. Zhou, and M. Chandraker. 2016. Deep Deformation Network for Object Landmark Localization. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 52–70.

H. Zhang, Q. Li, Z. Sun, and Y. Liu. 2018. Combining Data-driven and Model-driven Methods for Robust Facial Landmark Detection. *IEEE Transactions on Information Forensics and Security* 13, 10 (2018), 2409–2422.

K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. 2016. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.

J.-Y. Zhu, T. Park, P. Isola, and A.-A. Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2242–2251.

S. Zhu, C. Li, C. Change Loy, and X. Tang. 2015. Face Alignment by Coarse-to-Fine Shape Searching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4998–5006.

S. Zhu, C. Li, C.-C. Loy, and X. Tang. 2016. Unconstrained Face Alignment via Cascaded Compositional Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3409–3417.

X. Zhu and D. Ramana. 2012. Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2879–2886.