Convolutional Sketch Inversion

Yağmur Güçlütürk$^{(\boxtimes)}$, Umut Güçlü, Rob van Lier,
and Marcel A.J. van Gerven

Donders Institute for Brain, Cognition and Behaviour,
Radboud University, Nijmegen, The Netherlands
{y.gucluturk,u.guclu,r.vanlier,m.vangerven}@donders.ru.nl

Abstract. In this paper, we use deep neural networks for inverting face sketches to synthesize photorealistic face images. We first construct a semi-simulated dataset containing a very large number of computer-generated face sketches with different styles and corresponding face images by expanding existing unconstrained face data sets. We then train models achieving state-of-the-art results on both computer-generated sketches and hand-drawn sketches by leveraging recent advances in deep learning such as batch normalization, deep residual learning, perceptual losses and stochastic optimization in combination with our new dataset. We finally demonstrate potential applications of our models in fine arts and forensic arts. In contrast to existing patch-based approaches, our deep-neural-network-based approach can be used for synthesizing photorealistic face images by inverting face sketches in the wild.

Keywords: Deep neural network · Face synthesis · Face recognition · Fine arts · Forensic arts · Sketch inversion · Sketch recognition

1 Introduction

Portrait and self-portrait sketches have an important role in art. From an art historical perspective, self-portraits serve as historical records of what the artists looked like. From the perspective of an artist, self-portraits can be seen as a way to practice and improve one's skills without the need for a model to pose. Portraits of others further serve as memorabilia and a record of the person in the portrait. Artists most often are able to easily capture recognizable features of a person in their sketches. Therefore, hand-drawn sketches of people have further applications in law enforcement. Sketches of suspects drawn based on eye-witness accounts are used to identify suspects, either in person or from catalogues of mugshots (Fig. 1).

However, a challenging task that remains is photorealistic face image synthesis from face sketches in uncontrolled conditions. That is, at present, there exist no sketch inversion models that are able to perform in realistic conditions. These
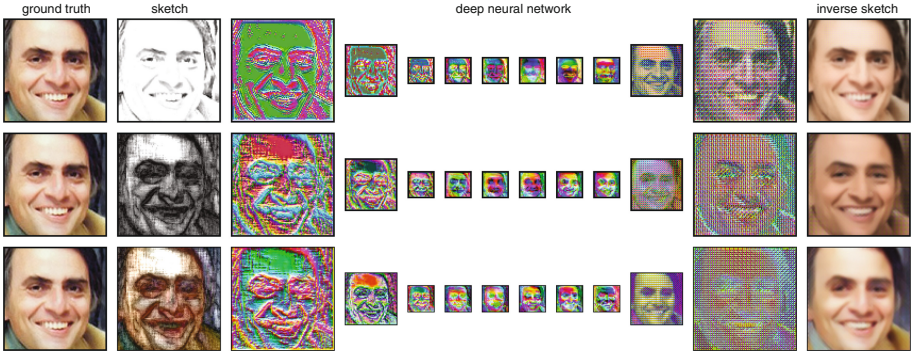

Fig. 1. Demonstration of our convolutional sketch inversion models. Our models invert face sketches to synthesize photorealistic face images. Each row shows the sketch inversion/image synthesis pipeline that transforms a different sketch of the same face to a different image of the same face via a different deep neural network. Each deep neural network layer is represented by the top three principal components of its feature maps.

conditions are characterized by changes in expression, pose, lighting condition and image quality, as well as the presence of varying amounts of background clutter and occlusions.

Here, we use DNNs to tackle the problem of inverting face sketches to synthesize photorealistic face images from different sketch styles in uncontrolled conditions. We developed three different models to handle three different types of sketch styles by training DNNs on datasets that we constructed by extending a well-known large-scale face dataset, obtained in uncontrolled conditions [21]. We test the models on another similar large-scale dataset [17], a hand-drawn sketch database [31] as well as on self-portrait sketches of famous Dutch artists. We show that our approach, which we refer to as *Convolutional Sketch Inversion* (CSI) can be used to achieve state-of-the-art results and discuss possible applications in fine arts, art history and forensics.

2   Related Work

Prior work related to face sketches in computer vision has been mostly limited to synthesis of highly controlled (i.e. having neutral expression, frontal pose, with normal lighting and without any occlusions) sketches from photographs [7, 19, 26, 30, 34] (sketch synthesis) and photographs from sketches [7, 20, 30, 31, 33] (sketch inversion). Sketch inversion studies with controlled inputs utilized patch-based approaches and used Bayesian tensor inference [20], an embedded hidden Markov model [33], a multiscale Markov random field model [31], sparse representations [7] and transductive learning with a probabilistic graph model [30].

Few studies developed methods of sketch synthesis to handle more variation in one or more variables at a time, such as lighting [18], and lighting and pose [36]. In a recent study, Zhang et al. [35] showed that sketch synthesis by transferring

the style of a single sketch could be used also in uncontrolled conditions. In [35], first an initial sketch by a sparse representation-based greedy search strategy was estimated, then candidate patches were selected from a template style sketch and the estimated initial sketch. Finally, the candidate patches were refined by a multi-feature-based optimization model and the patches were assembled to produce the final synthesized sketch.

Recently, the use of deep convolutional neural networks (DNNs) in image transformation tasks, in which one type of image is transformed into another, has gained tremendous traction. In the context of sketch analysis, DNNs were used to tackle the problems of sketch synthesis and sketch simplification. For example, [34] has used a DNN to convert photographs to sketches. They developed a DNN with six convolutional layers and a discriminative regularization term for enhancing the discriminability of the generated sketch against other sketches. Furthermore, [24] has used a DNN to simplify rough sketches. They have shown that users prefer sketches simplified by the DNN more than they do those by other applications 97 % of the time.

Some other notable image transformation problems include colorization, style transfer and super-resolution. In colorization, the task is to transform a grayscale image to a color image that accurately captures the color information. In style transfer, the task is to transform one image to another image that captures the style of a third image. In super-resolution, the task is to transform a low-resolution image to a high-resolution image with maximum quality. DNNs have been used to tackle all of these problems with state-of-the art results [3,5,6,9, 13,15].

3   Semi-simulated Datasets

For training and testing our CSI model, we made use of the following datasets:

- *Large-scale CelebFaces Attributes (CelebA) dataset* [21]. The CelebA dataset contains 202,599 celebrity face images and 10,177 identities. The images were obtained from the internet and vary extensively in terms of pose, expression, lighting, image quality, background clutter and occlusion. Each image in the dataset has five landmark positions and 40 attributes. These images were used for training the networks.
- *Labeled Faces in the Wild (LFW) dataset* [17]. The LFW dataset contains 13,233 face images and 5749 identities. Similar to the CelebA dataset, images were obtained from the internet and vary extensively in terms of pose, expression, lighting, image quality, background clutter and occlusion. A subset of these images (11,990) were used for testing the networks.
- *CUHK Face Sketch (CUFS) database* [31]. The CUFS database contains photographs and their corresponding hand-drawn sketches of 606 individuals. The dataset was formed by combining face photographs from three other databases and producing hand-drawn sketches of these photographs. Concretely, it consists of 188 face photographs from the Chinese University of Hong Kong

(CUHK) student database [31] and their corresponding sketches, 123 face photographs from the AR Face Database [22] and their corresponding sketches, and 295 face photographs from the XM2VTS database [23] and their corresponding sketches. Only the 18 sketches that are showcased at the website of the CUFS database (six from each sub-database) were used in the current study. These images were used for testing the networks.

– *Sketches of famous Dutch artists.* We also used the following sketches: (i) Self-Portrait with Beret, Wide-Eyed by Rembrandt, 1630, etching, (ii) Two Self-portraits and Several Details by Vincent van Gogh, 1886, pencil on paper and (iii) Self-Portrait by M.C. Escher, 1929, lithograph on gray paper. These images were used for testing the networks.

3.1   Preprocessing

Similar to [4], each image was cropped and resized to 96 pixels × 96 pixels such that:

– The distance between the top of the image and the vertical center of the eyes was 38 pixels.
– The distance between the vertical center of the eyes and the vertical center of the mouth was 32 pixels.
– The distance between the vertical center of the mouth and the bottom of the image was 26 pixels.
– The horizontal center of the eyes and the mouth was at the horizontal center of the image.

3.2   Sketching

Each image in the CelebA and LFW datasets was automatically transformed to a line sketch, a grayscale sketch and a color sketch. Sketches in the CUFS database and those by the famous Dutch artists were further transformed to line sketches by using the same procedure.

Color and grayscale sketch types are produced by the same stylization algorithm [8]. To obtain the sketch images, the input image is first filtered by an edge-aware filter. This filtered image is then blended with the magnitude of the gradient of the filtered image. Then, each pixel is scaled by a normalization factor resulting in the final sketch-like image.

Line sketches which resemble pencil sketches were generated based on [2]. Line sketch conversion works by first converting the color image to grayscale. This is followed by inverting the grayscale image to obtain a negative image. Next, a Gaussian blur is applied. Finally, using color dodge, the resulting image is blended with the grayscale version of the original image.

It should be noted that synthesizing face images from color or grayscale sketches is a more difficult problem than doing so from line sketches since many details of the faces are preserved by line sketches while they are lost for other sketch types.

4   Models

We developed one DNN for each of the three sketch styles based on the style transfer architecture in [15]. Each of the three DNNs was based on the same architecture except for the first layer where the number of input channels were either one or three depending on the number of color channels of the sketches. The architecture comprised three convolutional layers, five residual blocks [12], two deconvolutional layers and another convolutional layer. Each of the five residual blocks comprised two convolutional layers. All of the layers except for the last layer were followed by batch normalization [14] and rectified linear units. The last layer was followed by batch normalization and hyperbolic tangent units. All models were implemented in the Chainer framework [27]. Table 1 shows the details of the architecture.

Table 1. Deep neural network architectures. BN; batch normalization with decay = 0.9, $\epsilon = 1e - 5$, ReLU; rectified linear unit, con.; convolution, dec.; deconvolution, res.; residual block, tanh; hyperbolic tangent unit. Outputs of the hyperbolic tangent units are scaled to [0, 255]. x/y indicates the parameters of the first and second layers of a residual block. +x indicates that the input and output of a block are summed and no activation function is used.

Layer	Type	in_channels	out_channels	ksize	Stride	Pad	Normalization	Activation
1	con.	1 or 3	32	9	1	4	BN	ReLU
2	con.	32	64	3	2	1	BN	ReLU
3	con.	64	128	3	2	1	BN	ReLU
4	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
5	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
6	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
7	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
8	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
9	dec.	128	64	3	2	1	BN	ReLU
10	dec.	64	32	3	2	1	BN	ReLU
11	con.	32	3	9	1	4	BN	tanh

4.1   Estimation

For model optimization we used Adam [16] with parameters $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ and mini-batch size = 4. We trained the models by iteratively minimizing the loss function for 200,000 iterations. The loss function comprised three components. The first component is the standard Euclidean loss for the targets and the predictions (pixel loss; ℓ_p). The second component is the Euclidean loss for the feature-transformed targets and the feature-transformed predictions (feature loss) [15]:

$$\ell_f = \frac{1}{n} \sum_{i,j,k} \left(\phi\left(t\right)_{i,j,k} - \phi\left(y\right)_{i,j,k} \right)^2 \tag{1}$$

where n is the total number of features, $\phi(t)_{i,j,k}$ is a feature of the targets and $\phi(y)_{i,j,k}$ is a feature of the predictions. Similar to [15], we used the outputs of the fourth layer of a 16-layer DNN (relu_2_2 outputs of the VGG-16 pretrained model) [25] to feature transform the targets and the predictions. The third component is the total variation loss for the predictions:

$$\ell_{tv} = \sum_{i,j} \left(\left(y_{i+1,j} - y_{i,j}\right)^2 + \left(y_{i,j+1} - y_{i,j}\right)^2 \right)^{0.5} \tag{2}$$

where $y_{i,j}$ is a pixel of the predictions. A weighted combination of these components resulted in the following loss function:

$$\ell = \lambda_p \ell_p + \lambda_f \ell_f + \lambda_{tv} \ell_{tv} \tag{3}$$

where we set $\lambda_p = \lambda_f = 1$ and $\lambda_{tv} = 0.00001$.

The use of the feature loss to train models for image transformation tasks was recently proposed by [15]. In the context of super-resolution, [15] found that replacing pixel loss with feature loss gives visually pleasing results at the expanse of image quality because of the artefacts introduced by the feature loss.

In the context of sketch inversion, our preliminary experiments showed that combining feature loss and pixel loss increases image quality while maintaining visual pleasantness. Furthermore, we observed that a small amount of total variation loss further removes the artefacts that are introduced by the feature loss. Therefore, we used the combination of the three losses in the final experiments. The quantitative results of the preliminary experiments in which the models were trained by using only the feature loss are provided in the Appendix (Tables 4 and 5).

4.2   Validation

First, we qualitatively tested the models by visual inspection of the synthesized face images (Fig. 2). Synthesized face images matched the ground truth photographs closely and persons in the images were easily recognizable in most cases. Among the three styles of sketch models, the line sketch model (Fig. 2, first column) captured the highest level of detail in terms of the face structure, whereas the synthesized inverse sketches of the color sketch model (Fig. 2, third column) had less structural detail but was able to better reproduce the color information in the ground truth images compared to the inverted sketches of the line sketch model. Sketches synthesized by the grayscale model (Fig. 2, second column) were less detailed than those synthesized by the line sketch model. Furthermore, the color content was less accurate in sketches synthesized by the grayscale model than those synthesized by both the color sketch and the line sketch models. We found that the line model performed impressively in terms of
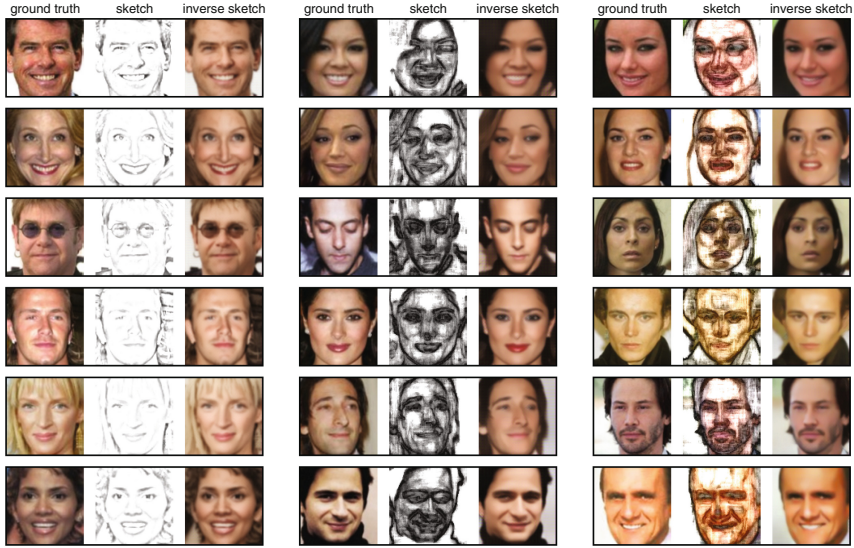
Fig. 2. Examples of the synthesized inverse sketches from the LFW dataset. Each distinct column shows examples from different sketch styles models, i.e. line sketch model (column 1), grayscale sketch model (column 2) and colour sketch model (column 3). First image in each column is the ground truth, the second image is the generated sketch and the third one is the synthesized inverse sketch.

matching the hair and skin color of the individuals even when the line sketches did not contain any color information. This may indicate that along with taking advantage of the luminance differences in the sketches to infer coloring, the model was able to learn color properties often associated with high-level face features of different ethnicities.

Then, we quantitatively tested the models by comparison of the peak signal to noise ratio (PSNR), structural similarity (SSIM) and standard Pearson product-moment correlation coefficient R of the synthesized face images [32] (Table 2). PSNR measures the physical quality of an image. It is defined as the ratio between the peak power of the image and the power of the noise in the image (Euclidean distance between the image and the reference image):

$$\text{PSNR} = \frac{1}{3} \sum_k 10 \log_{10} \frac{\max \text{DR}^2}{\frac{1}{m} \sum_{i,j} \left(t_{i,j,k} - y_{i,j,k} \right)^2} \tag{4}$$

where DR is the dynamic range, and m is the total number of pixels in each of the three color channels. SSIM measures the perceptual quality of an image. It is defined as the multiplicative combination of the similarities between the image and the reference image in terms of contrast, luminance and structure:

$$\text{SSIM} = \frac{1}{3} \sum_k \frac{1}{m} \sum_{i,j} \frac{\left(2\mu\left(t_{i,j,k}\right)\mu\left(y_{i,j,k}\right) + C_1\right)\left(2\sigma\left(t_{i,j,k}, y_{i,j,k}\right) C_2\right)}{\left(\mu\left(t_{i,j,k}\right)^2 \mu\left(y_{i,j,k}\right)^2 + C_1\right)\left(2\sigma\left(t_{i,j,k}\right)^2 \sigma\left(y_{i,j,k}\right)^2 C_2\right)}$$

(5)

where $\mu\left(t_{i,j,k}\right)$, $\mu\left(y_{i,j,k}\right)$, $\sigma\left(t_{i,j,k}\right)$, $\sigma\left(y_{i,j,k}\right)$ and $\sigma\left(t_{i,j,k}, y_{i,j,k}\right)$ are means, standard deviations and cross-covariances of windows centered around i and j. Furthermore, $C_1 = (0.01 \max \text{DR})^2$ and $C_2 = (0.03 \max \text{DR})^2$. Quality of a dataset is defined as the mean quality over the images in the dataset.

Table 2. Comparison of physical (PSNR), perceptual (SSIM) and cor relational (R) quality measures for the inverse sketches synthesized by the line, grayscale and color sketch-style models. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	SSIM	R
Line	**20.1158 ± 0.0231**	**0.8583 ± 0.0003**	**0.9298 ± 0.0005**
Grayscale	17.6567 ± 0.0263	0.6529 ± 0.0008	0.7458 ± 0.0020
Color	19.2029 ± 0.0293	0.7154 ± 0.0008	0.8087 ± 0.0017

The inversion of the line sketches resulted in the highest quality face images for all three measures (20.12 for PSNR, 0.86 for SSIM and 0.93 for R). In contrast the inversion of the grayscale sketches resulted in the lowest quality face images for all measures (17.65 for PSNR, 0.65 for SSIM and 0.75 for R). This shows that both the physical and the perceptual quality of the inverted sketch images produced by the line sketch network was superior than those by the other sketch styles.

Finally, we tested how well the line sketch inversion model can be transferred to the task of synthesizing face images from sketches that are hand-drawn and not generated using the same methods that were used to train the model. We considered only the line sketch model since the contents of the hand-drawn sketch database that we used [31] were most similar to the line sketches.

We found that the line sketch inversion model can solve this inductive transfer task almost as good as it can solve the task that it was trained on (Fig. 3). Once again, the model synthesized photorealistic face images. While color was not always synthesized accurately, other elements such as form, shape, line, space and texture were often synthesized well. Furthermore hair texture and style, which posed a problem in most previous studies, was very well handled by our CSI model. We observed that the dark-edged pencil strokes in the hand-drawn sketches that were not accompanied by shading resulted in less realistic inversions (compare e.g. nose areas of sketches in the first and second rows with those in the third row in Fig. 3). This can be explained by the lack of such features in the training data of the line sketch model, and can be easily overcome by including training examples more closely resembling the drawing style of the sketch artists.

For all the samples from the CUFS database, the PSNR, the SSIM index and the R of the synthesized face images were 13.42, 0.52, and 0.67, respectively

Fig. 3. Examples of the synthesized inverse sketches from the CUFS database. First image in each column is the ground truth, the second image is the sketch hand-drawn by an artist and the third one is the inverse sketch that was synthesized by the line sketch model.

(Table 3). Among the three sub-databases of the CUFS database, the quality of the synthesized images from the CUHK dataset was the highest in terms of the PSNR (15.07) and R (0.83). While the PSNR and R values for the AR dataset was lower than those of the CUHK dataset, SSIM did not differ between the two datasets. The lowest quality inverted sketches were produced from the sample sketches of the XM2GTS database (with 13.42 for PSNR, 0.42 for SSIM and 0.41 for R).

Additional results on both computer-generated sketches and hand-drawn sketches are provided at https://github.com/yagguc/CSI due to space limitations.

Table 3. Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized from the sketches in the CUFS database and its sub-databases. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	SSIM	R
CUHK (6)	**15.0675 ± 0.3958**	0.5658 ± 0.0099	**0.8264 ± 0.0269**
AR (6)	13.8687 ± 0.7009	**0.5684 ± 0.0277**	0.7667 ± 0.0314
XM2GTS (6)	11.3293 ± 1.2156	0.4231 ± 0.0272	0.4138± 0.1130
All (18)	13.4218 ± 0.6123	0.5191 ± 0.0207	0.6690 ± 0.0591

5   Applications

5.1   Fine Arts

In many cases self-portrait studies allow us a glimpse of what famous artists looked like through the artists' own perspective. Since there are no photographic records of many artists (in particular of those who lived before the 19th century during which the photography was invented and became widespread) self-portrait sketches and paintings are the only visual records that we have of many artists. Converting the sketches of the artists into photographs using a DNN that was trained on tens of thousands of face sketch-photograph pairs results in very interesting end-products.

Here we used our DNN-based approach to synthesize photographs of famous Dutch artists Rembrandt, Vincent van Gogh and M.C. Escher from their self-portrait sketches[1] (Fig. 4). To the best of our knowledge, the synthesized photorealistic images of these artists are the first of their kind.

Fig. 4. Self-portrait sketches and synthesized inverse sketches along with a reference painting or photograph of famous Dutch artists: Rembrandt (top), Vincent van Gogh (middle) and M.C. Escher (bottom). Sketches: (i) Self-Portrait with Beret, Wide-Eyed by Rembrandt, 1630, etching. (ii) Two Self-portraits and Several Details by Vincent van Gogh, 1886, pencil on paper. (iii) Self-Portrait by M.C. Escher, 1929, lithograph on gray paper. Reference paintings: (i) Self-Portrait by Rembrandt, 1630, oil painting on copper. (ii) Self-Portrait with Straw Hat by Vincent van Gogh, 1887, oil painting on canvas.

---

[1] For simplicity, although different methods were used to produce these artworks, we refer to them as sketches.

Our qualitative assesment revealed that, the inverted sketch of Rembrandt synthesized from his 1630 sketch indeed resembles himself in his paintings (particulary his self-portrait painting from 1630), and Escher's to his photographs. We found that the inverted sketch of van Gogh synthesized from his 1886 sketch was the most realistic synthesized photograph among those of the three artists, albeit not closely matching his self-portrait paintings of a distinct post-impressionist style.

Although we do not have a quantitative way to measure the accuracy of the results in this case, results demonstrate that the artistic style of the input sketches influence the quality of the produced photorealistic images. Generating new training sketch data to match more closely to the sketch style of a specific artist of interest (e.g. by using the method proposed by [35]), and training the network with these sketches would overcome this limitation.

Sketching is one of the most important training methods that artist use to develop their skills. Converting sketches into photorealistic images would allow the artists in training to see and evaluate the accuracy of their sketches clearly and easily which can in turn become an efficient training tool. Furthermore, sketching is often much faster than producing a painting. When for example the sketch is based on imagination rather than a photograph, deep sketch inversion can provide a photorealistic guideline (or even an end-product, if digital art is being produced) and can speed up the production process of artists. Figure 3, which shows the inverted sketches by contemporary artists that produced the sketches in the CUFS database, further demonstrates this type of application. The current method can be developed into a smartphone/tablet or computer application for common use.

5.2 Forensic Arts

In cases where no other representation of a suspect exists, sketches drawn by forensic artists based on eye-witness accounts are frequently used by the law enforcement. However, direct use of sketches for automatically identifying suspects from databases containing photographs does not work well because these two face representations are too different to allow a direct comparison [29]. Inverting a sketch to a photograph makes this task much easier by reducing the difference between these two alternative representations, enabling a direct automatized comparison [31].

To evaluate the potential use of our system for forensic applications, we performed an identification analysis (Fig. 5). In this analysis, we evaluated the accuracy of identifying a target face image in a very large set of candidate face images (LFW dataset containing over 11,000 images) from an (inverse) face sketch. The identification accuracies for the synthesized faces were always significantly higher than those for the corresponding sketched faces ($p \ll 0.05$, binomial test). While the identification accuracies for the color and grayscale sketches were very low (2.38 % and 1.42 %, respectively), those for the synthesized color and grayscale inverse sketches were relatively high (82.29 % and 73.81 %, respectively). On the other hand, identification accuracy of line sketches was already high, at 81.14 %

before inversion. Synthesizing inverse sketches from line sketches raised the identification accuracy to an almost perfect level (99.79 %).
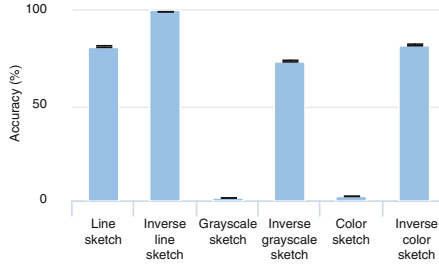

Fig. 5. Identification accuracies for line, grayscale and color sketches, and for inverse sketches synthesized by the corresponding models. Error bars show the bootstrap estimates of the standard errors.

6   Conclusions

In this study we developed sketch datasets, complementing well known unconstrained benchmarking datasets [17,21], developed DNN models that can synthesize face images from sketches with state-of-the-art performance and proposed applications of our CSI model in fine arts, art history and forensics. We foresee further computer vision applications of the developed methods for non-face images and various other sketch-like representations, as well as cognitive neuroscience applications for the study of cognitive phenomena such as perceptual filling in [1,28] and the neural representation of complex stimuli [10,11].

Appendix

Table 4. Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized by the line, grayscale and color sketch-style models trained using feature loss alone. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	SSIM	R
Line	14.8956 ± 0.0207	0.5931 ± 0.0006	0.6023 ± 0.0017
Grayscale	17.1654 ± 0.0277	0.6301 ± 0.0008	0.7175 ± 0.0022
Color	**18.9884 ± 0.0296**	**0.7072 ± 0.0008**	**0.7976 ± 0.0019**

Table 5. Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized from the sketches in the CUFS database and its sub-databases with the line sketch model trained using feature loss alone. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	SSIM	R
CUHK (6)	**14.6213 ± 0.4061**	0.5358 ± 0.0216	**0.8295 ± 0.0200**
AR (6)	14.1721 ± 0.4127	**0.5608 ± 0.0232**	0.7811 ± 0.0217
XM2GTS (6)	11.7158 ± 1.3050	0.4096 ± 0.0258	0.3817 ± 0.1341
All (18)	13.5030 ± 0.5639	0.5021 ± 0.0205	0.6641 ± 0.0658

References

1. Anstis, S., Vergeer, M., Lier, R.V.: Looking at two paintings at once: luminance edges can gate colors. i-Perception **3**(8), 515–518 (2012). http://dx.doi.org/10.1068/i0537sas
2. Beyeler, M.: OpenCV with Python Blueprints. Packt Publishing, Birmingham (2015)
3. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: International Conference on Computer Vision. Institute of Electrical and Electronics Engineers (IEEE), December 2015. http://dx.doi.org/10.1109/ICCV.2015.55
4. Cowen, A.S., Chun, M.M., Kuhl, B.A.: Neural portraits of perception: reconstructing face images from evoked brain activity. NeuroImage **94**, 12–22 (2014). http://dx.doi.org/10.1016/j.neuroimage.2014.03.018
5. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision (2014)
6. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016). http://dx.doi.org/10.1109/TPAMI.2015.2439281
7. Gao, X., Wang, N., Tao, D., Li, X.: Face sketch-photo synthesis and retrieval using sparse representation. IEEE Trans. Circ. Syst. Video Technol. **22**(8), 1213–1226 (2012). http://dx.doi.org/10.1109/TCSVT.2012.2198090
8. Gastal, E.S.L., Oliveira, M.M.: Domain transform for edge-aware image and video processing. ACM Trans. Graph. **30**(4), 1 (2011). http://dx.doi.org/10.1145/2010324.1964964
9. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. CoRR abs/1508.06576 (2015)
10. Güçlü, U., van Gerven, M.A.J.: Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J. Neurosci. **35**(27), 10005–10014 (2015). http://dx.doi.org/10.1523/JNEUROSCI.5023-14.2015
11. Güçlü, U., van Gerven, M.A.J.: Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. NeuroImage (2015). http://dx.doi.org/10.1016/j.neuroimage.2015.12.036
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR abs/1512.03385 (2015)
13. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Trans. Graph. **35**(4), 110 (2016)

14. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. CoRR abs/1603.08155 (2016)
16. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. CoRR abs/1412.6980 (2014)
17. Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H., Hua, G.: Labeled faces in the wild: a survey. In: Kawulok, M., Celebi, M.E., Smolka, B. (eds.) Advances in Face Detection and Facial Image Analysis, pp. 189–248. Springer, Heidelberg (2016). http://dx.doi.org/10.1007/978-3-319-25958-1_8
18. Li, Y.h., Savvides, M., Bhagavatula, V.: Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In: International Conference on Acoustics, Speech, and Signal Processing. Institute of Electrical and Electronics Engineers (IEEE) (2006). http://dx.doi.org/10.1109/ICASSP.2006.1660353
19. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A nonlinear approach for face sketch synthesis and recognition. In: Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers (IEEE) (2005). http://dx.doi.org/10.1109/CVPR.2005.39
20. Liu, W., Tang, X., Liu, J.: Bayesian tensor inference for sketch-based facial photo hallucination. In: International Joint Conference on Artificial Intelligence (2007)
21. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: International Conference on Computer Vision (2015)
22. Martinez, A.M., Benavente, R.: The AR-face database. CVC Technical report 24 (1998)
23. Messer, K., Matas, J., Kittler, J., Jonsson, K.: XM2VTSDB: The extended M2VTS database. In: Audio and Video-based Biometric Person Authentication (1999)
24. Simo-Serra, E., Iizuka, S., Sasaki, K., Ishikawa, H.: Learning to simplify: fully convolutional networks for rough sketch cleanup. ACM Trans. Graph. **35**(4), 121 (2016)
25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
26. Tang, X., Wang, X.: Face sketch synthesis and recognition. In: International Conference on Computer Vision. Institute of Electrical and Electronics Engineers (IEEE) (2003). http://dx.doi.org/10.1109/ICCV.2003.1238414
27. Tokui, S., Oono, K., Hido, S., Clayton, J.: Chainer: a next-generation open source framework for deep learning. In: Workshop on Machine Learning Systems at Neural Information Processing Systems (2015)
28. Vergeer, M., Anstis, S., van Lier, R.: Flexible color perception depending on the shape and positioning of achromatic contours. Front. Psychol. **6** (2015). http://dx.doi.org/10.3389/fpsyg.2015.00620
29. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: A comprehensive survey to face hallucination. Int. J. Comput. Vis. **106**(1), 9–30 (2013). http://dx.doi.org/10.1007/s11263-013-0645-9
30. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: Transductive face sketch-photo synthesis. IEEE Trans. Neural Netw. Learn. Syst. **24**(9), 1364–1376 (2013). http://dx.doi.org/10.1109/TNNLS.2013.2258174
31. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 1955–1967 (2009). http://dx.doi.org/10.1109/TPAMI.2008.222

32. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). http://dx.doi.org/10.1109/TIP.2003.819861
33. Xiao, B., Gao, X., Tao, D., Li, X.: A new approach for face recognition by sketches in photos. Sig. Process. **89**(8), 1576–1588 (2009). http://dx.doi.org/10.1016/j.sigpro.2009.02.008
34. Zhang, L., Lin, L., Wu, X., Ding, S., Zhang, L.: End-to-end photo-sketch generation via fully convolutional representation learning. In: International Conference on Multimedia Retrieval. Association for Computing Machinery (ACM) (2015). http://dx.doi.org/10.1145/2671188.2749321
35. Zhang, S., Gao, X., Wang, N., Li, J.: Robust face sketch style synthesis. IEEE Trans. Image Process. **25**(1), 220–232 (2016). http://dx.doi.org/10.1109/TIP.2015.2501755
36. Zhang, W., Wang, X., Tang, X.: Lighting and pose robust face sketch synthesis. In: European Conference on Computer Vision (2010)