Computational Visual Media
https://doi.org/10.1007/s41095-021-0227-7 Vol. 8, No. 1, March 2022, 63–77

Research Article

DualFace: Two-stage drawing guidance for freehand portrait sketching

Zhengyu Huang[1], Yichen Peng[1], Tomohiro Hibino[1], Chunqi Zhao[2], Haoran Xie[1] (✉), Tsukasa Fukusato[2], and Kazunori Miyata[1]

Abstract Special skills are required in portrait painting, such as imagining geometric structures and facial detail for final portrait designs. This makes it a difficult task for users, especially novices without prior artistic training, to draw freehand portraits with high-quality details. In this paper, we propose dualFace, a portrait drawing interface to assist users with different levels of drawing skills to complete recognizable and authentic face sketches. Inspired by traditional artist workflows for portrait drawing, dualFace gives two-stages of drawing assistance to provide global and local visual guidance. The former helps users draw contour lines for portraits (i.e., geometric structure), and the latter helps users draw details of facial parts, which conform to the user-drawn contour lines. In the global guidance stage, the user draws several contour lines, and dualFace then searches for several relevant images from an internal database and displays the suggested face contour lines on the background of the canvas. In the local guidance stage, we synthesize detailed portrait images with a deep generative model from user-drawn contour lines, and then use the synthesized results as detailed drawing guidance. We conducted a user study to verify the effectiveness of dualFace, which confirms that dualFace significantly helps users to produce a detailed portrait sketch.

Keywords portrait painting; user interface; freehand sketching; generative model; drawing guidance

1 Japan Advanced Institute of Science and Technology, Ishikawa, 9231211, Japan. E-mail: Z. Huang, huang.zhengyu@jaist.ac.jp; Y. Peng, puckikk@gmail.com; T. Hibino, t_hibino@jaist.ac.jp; H. Xie, xie@jaist.ac.jp (✉); K. Miyata, miyata@jaist.ac.jp.

2 The University of Tokyo, Tokyo, 1138654, Japan. E-mail: C. Zhao, shunnki.chou@ui.is.s.u-tokyo.ac.jp; T. Fukusato, tsukasafukusato@is.s.u-tokyo.ac.jp.

Manuscript received: 2021-01-14; accepted: 2021-03-15

1 Introduction

Portrait painting is an important art genre, representing a specific human from the real world or one's imagination. Some artists, together with their famous portrait drawings, have been widely adored for hundreds of years (e.g., *Mona Lisa* and *Girl with a Pearl Earring*). However, drawing portraits is cumbersome and requires special skills and capabilities (for example, spatial imagination and essential drawing skills), which are inaccessible to novices without prior artistic training. Therefore, the present paper aims to establish a user-friendly framework to support the process of drawing freehand portraits.

Several guidance-based systems have been proposed for supporting portrait drawing. For example, Portraitsketch [1] proposes a framework to display an artistically rendered sketch using tracing. However, the user must prepare a reference image in advance, which can be time consuming. Shadowdraw [2] and Sketchhelper [3] incorporate image retrieval methods with tracing tools to dynamically search for relevant images from a database instead of manual selection, and enable users to understand geometric structures of target designs (e.g., locations and proportions of facial parts). Although the approaches mentioned above can help users copy existing drawings, it is still difficult to explore novel portrait designs. In addition, these systems are unsuitable for drawing the details of portraits (particularly of facial parts) because they simply blend a set of relevant images. Thus, details of each image may be lost. Conversely, to explore new drawing designs in detailed, Ghosh et al. [4] and Zhu et al. [5] employ deep learning methods, especially with generative adversarial networks (GANs), to generate possible images with given color or edge constraints. However, the resulting image quality is still determined by the user's drawing skill, e.g., in locating facial parts, so it remains difficult for novices to design high-quality portrait drawings.

To address the problems above, we referred to conventional portrait drawing procedures [6], which explain that it is essential for novices to adopt two types of guidance: (i) global guidance, which helps users locate facial parts (geometric structures) with correct proportions and (ii) local guidance, which helps users design facial details (e.g., eye and nose). Nonetheless, previous research does not consider how to guide users to draw both global and local features of portraits, as far as we are aware. Thus, we first consider a method to automatically generate two types of visual guidance, global and local guidance, from user drawings (see Fig. 1). In the case of global guidance, as with Shadowdraw [2] and Sketchhelper [3], when the user draws contour lines on the canvas, the system dynamically searches amongst relevant images from a database and generates a blended image. In the case of local guidance, the system generates detailed facial portraits from the user-drawn contour lines using a GAN-based system, and displays one of them. Secondly, we implement a graphical user interface (GUI), called dualFace, that incorporates the above visual guidance and is able to switch between the two stages freely.

Our principal contributions in summary are:

- A two-stage guidance system that helps users design portrait drawings with data-driven global guidance and GAN-based local guidance.
- An optimization method to automatically generate detailed facial portraits with semantic constraints from user-drawn strokes. By using the generated portraits as drawing guidance, the user can explore the desired details without prior artistic training.
- A user study to demonstrate the benefits of our proposed system.

2 Related work

2.1 Sketch-based applications

Sketches are a high-level abstract visual representation without great visual detail. By analyzing the intent behind a user's freehand sketch, sketch-based interaction allows users intuitive access to various applications such as image retrieval [2, 7, 8], image editing [9–12], simulation control [13], block arrangement [14], and 3D modeling [15–17]. Among sketch-based systems, freehand portrait sketching is difficult for ordinary users due to the required drawing skills and capabilities, which are inaccessible to novices (e.g., with poor drawing skills). To address this issue, we aim to establish a user-friendly framework to support the process of freehand drawing of human faces.

2.2 Drawing assistance

Sketching system guidance has been thoroughly investigated [18–20], especially visual guidance that can be extracted from reference images. For example, geometric structures [21, 22] on the canvas can help a user in the process of freeform drawing of objects

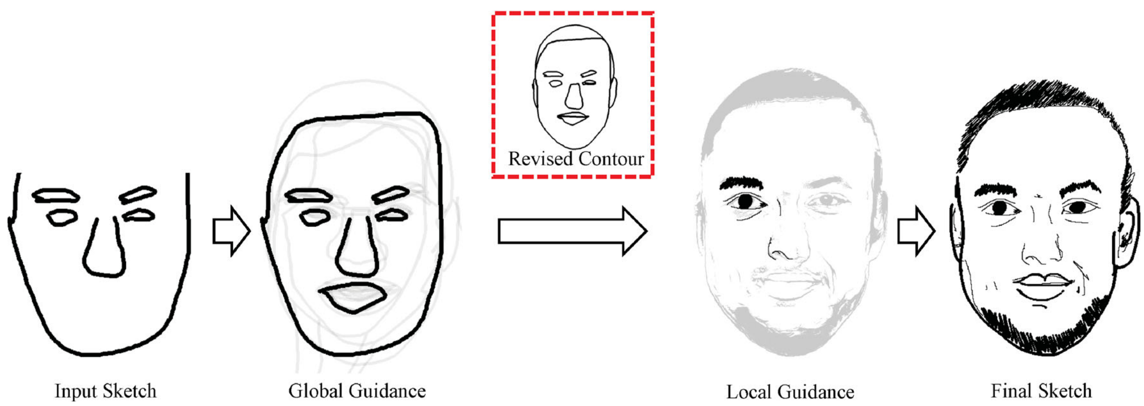


Fig. 1 Proposed portrait drawing interface, providing both global and local guidance from the input user sketch. The revised contour sketch in the back end is from the merged mask generated by our conversion algorithm from the input sketch, and is the reference for local guidance generation.

via tracing over the guidance [1, 23]. However, the user must select reference images, which can be time consuming. Lee et al. [2] and Choi et al. [3] dynamically search for relevant images in a large-scale database based on intermediate drawing results at drawing time and generate shadow guidance that suggests sketch completion to users. A similar drawing interface was designed for calligraphy practice [24]. With retrieval-based approaches, the visual guidance may be limited to the predefined database. To overcome this issue, image generation approaches can increase the variations from simple strokes, such as Drawfromdrawings [25] and MaskGAN [26]. Our framework combines both sketch-based retrieval and generation with optimization conversion from sketch-to-mask mapping.

2.3 Portrait rendering

In the field of non-photorealistic rendering (NPR) of portraits [27], there are two typical kinds of existing approach. One is to extract contour lines from images [28–30]. While these can be useful for visual abstraction (e.g., preserving and enhancing local shapes), it is difficult to consider semantic constraints and capture specific styles. The other approach is to train a network that automatically generates artistic-like drawings from facial images [31–34]. In these problem settings, training a network requires pairs of facial images and portraits. However, it is challenging to construct pixel-based (dense) correspondences because facial components (e.g., eyes and nose) in portraits are manually located by artists. Yi et al. [35] combine a global network (for images as a whole) and a local network (for recognising each facial component) and transform high-quality portraits while preserving facial components. In this paper, we adopt a similar portrait rendering model to generate portrait drawings, and use them as local guidance.

3 User interface

In this section, we describe how users interact with the proposed two-stage user interface (see Fig. 2) to draw portraits with global and local guidance. Please refer to the accompanying video in the Electronic Supplementary Material (ESM) for details.

3.1 Drawing tool

As with commercial drawing tools, the system enables the user to draw black strokes on the canvas with a mouse-drag operation; the stroke width is manually determined using a slider. The system automatically records all the vertices of the strokes and the stroke

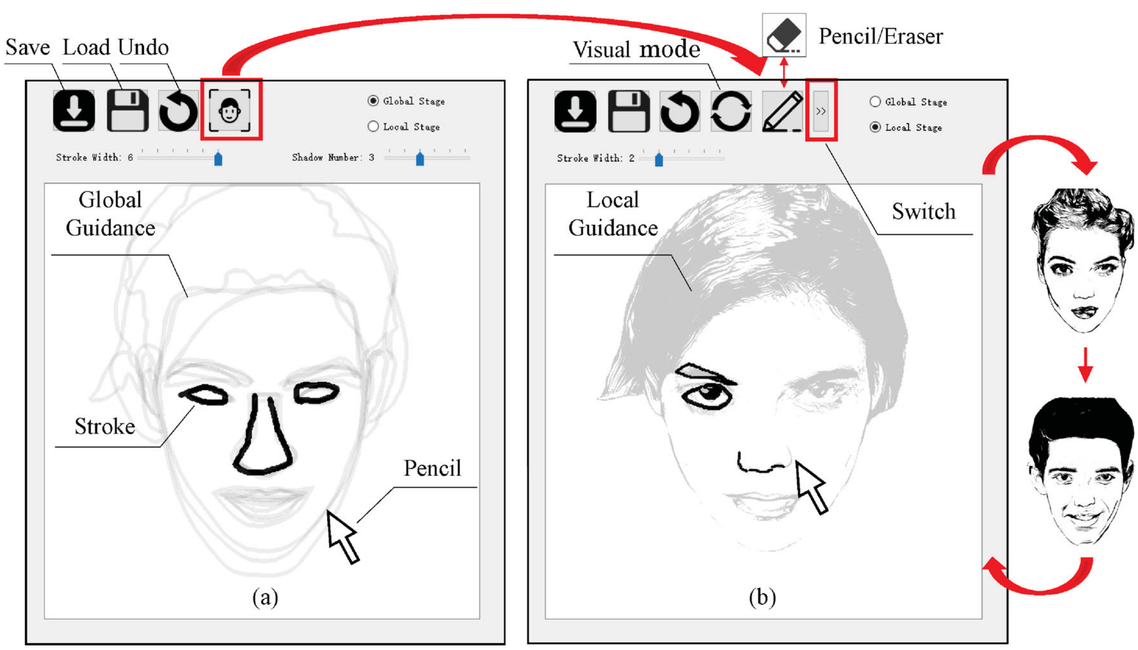


Fig. 2 Our two-stage guidance system. Given a user's intermediate drawing at runtime, the system generates: (a) global guidance by blending relevant images from the database and (b) local guidance (i.e., realistic facial portraits) using a model-based method.

order for the mask generation step. The eraser tool allows the user to click on a stroke, and the system deletes it. The undo tool can delete the last stroke from the stroke list. Our system can also load and save user-drawn strokes by clicking the appropriate buttons.

3.2 Visual guidance

Given user-drawn strokes, the system generates two types of visual guidance (global and local) for use in tracing. First, during global guidance, based on the user's intermediate drawing, the system dynamically searches for several relevant images from a database, and generates a "blended" image (global guidance) rather than a single image. With this global guidance, users can roughly understand locations and shapes of facial parts with correct proportions, as shown in Fig. 2(a). Second, during local guidance, the system generates several detailed facial portraits (guidance candidates) based on the user's intermediate drawings, and displays one of them instead of a blended image. The system has a switching function to change the generated images, so the user can search for the most reasonable local guidance. By using the local guidance, users can easily design local details such as eyes and nose; see Fig. 2(b). Note that the system allows the user to freely switch between global and local guidance modes by clicking the global/local radio button or the face icon button.

3.3 Rewind tool

In order to help users to draw the desired portraits, we provide a rewind tool in the proposed user interface. If users think that the local guidance does not meet their vision, the drawing process can be returned to the global stage by selecting the corresponding radio button, as shown in Fig. 2. Our drawing interface can automatically save the sketches while switching between global and local stages, so that users can revise their drawn contour sketches by reloading the recorded data.

4 Two-stage drawing guidance

Inspired by conventional portrait drawing processes, this work proposes dualFace, a two-stage framework for portrait drawing with both a global stage and a local stage for drawing guidance. The global stage provides interactive drawing guidance for each facial part. To help users achieve balanced facial contour drawing, we use a data-driven facial feature query by matching Gabor Local Line-based Features (GALIF) [36]. For the local stage, we adopt a GAN-based neural network to generate corresponding fine-grained sketches from a user's global stage rough contour sketch. Since we provide photo-realistic facial details in the local guidance, dualFace can help users concentrate on detailed drawing of facial features and improve their drawing skills. We believe the two-stage framework of dualFace may narrow the gap between novices and artists in portrait sketching due to the separation of global contour information from local facial details.

4.1 Global guidance

4.1.1 Overview

It is difficult to draw recognizable portraits with correct locations and portions of facial features, especially for novices. To solve this issue, dualFace first aims to help users to draw balanced facial contours. Figure 3 shows the workflow of global guidance, including data generation, contour matching, and interactive guidance. For the data generation step, face images are converted to contour images from a face database. For the contour matching step, local facial features are calculated and stored as feature vectors indexed into the database. For the interactive guidance step, the most similar candidates are retrieved in real time as shadow guidance. In contrast to the previous Shadowdraw work [2] with edge maps, we adopt labelled contour sketches for feature matching with the semantic sketch information. Thereby, each stroke of the user's drawing input can be matched with meaningful facial features for the next stage of local guidance.

4.1.2 Data generation

It is challenging to collect an enormous number of artist-designed portraits for face retrieval. Instead of artistic portraits, we generate semantic label masks [37] by utilizing a bilateral segmentation network (BiSeNet) pre-trained on the CelebAMask-HQ dataset [26]. Each pixel in the masks has a facial label ID from facial images (e.g., eyes, nose, and mouth). We adopted the contour function of the OpenCV library for the line drawing functions. The contours of facial components are extracted from the semantic label masks with balanced facial features. Note that the

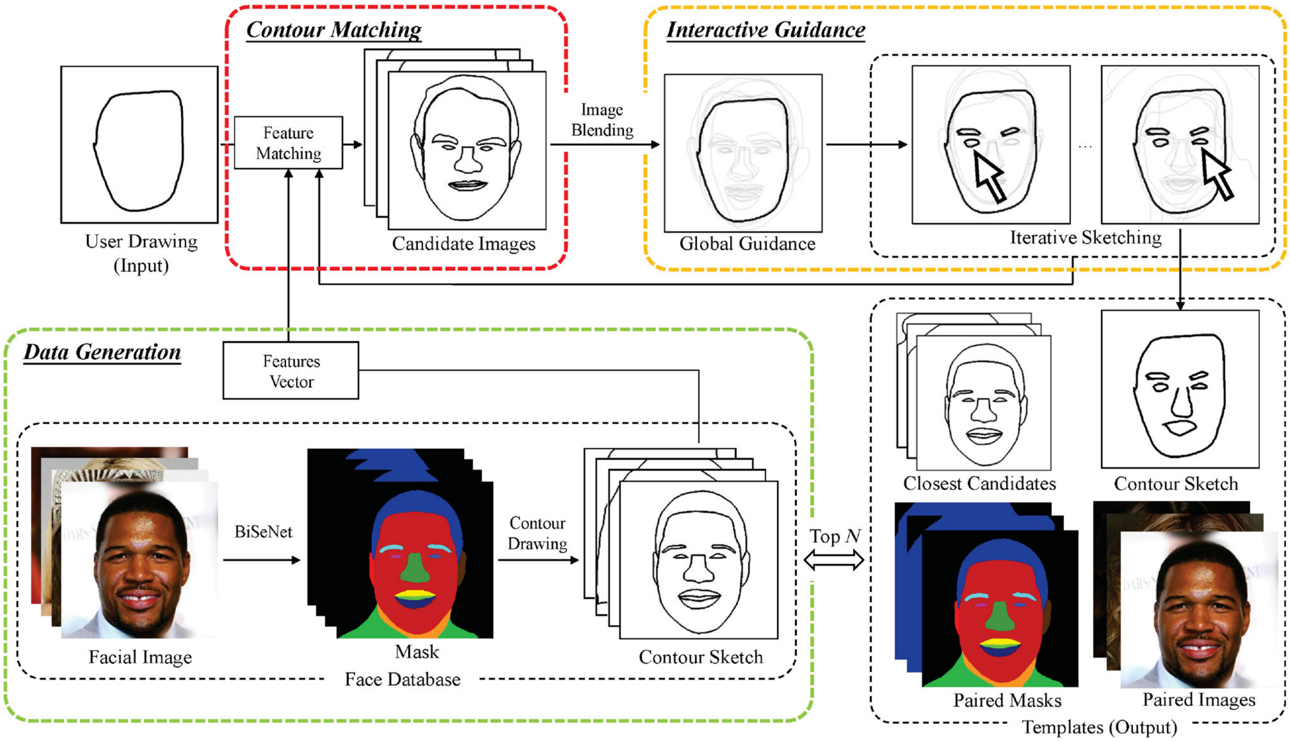


Fig. 3 Global guidance stage consists of three steps: data generation, contour matching, and interactive guidance. The contour sketches in our database are extracted from masks as source images, which are more meaningful for feature matching, and lead to better drawing guidance than previous work [2].

contour images are stored with the corresponding original face images, which are used for sketch retrieval in the global stage and for system input in the local stage.

4.1.3 Contour matching

To explore the closest contour sketches from the database for guidance according to a user's incomplete freehand sketch in real time, we use GALIF features for sketch retrieval and local shape matching [36]. For the online query method, the user sketch is encoded as a histogram. We calculate the similarity with the stored contour images in our database to obtain the closest contour images.

4.1.4 Interactive guidance

Following the shadow drawing interface [2], the top N relevant retrieval results in the face database are merged as a shadow image by image blending ($N = 3$ in our implementation). Benefiting from the interactive global guidance for portrait drawing, users understand the locations and shapes of each facial part. The global guidance is updated in real time for each drawing stroke. Under the help of global guidance for portrait sketching, the user can complete the contour sketch to express the rough shape and locations of facial parts meeting their drawing intent.

4.2 Local guidance

4.2.1 Overview

In order to guide users to draw details of facial components (e.g., irises and eyelashes), dualFace provides local guidance using relevant templates extracted from our database in the global stage. Local guidance for portrait sketching includes mask generation and portrait sketch generation (Fig. 4). For the mask generation step, user strokes in the global stage are recorded and converted to face masks based on the top N relevant templates ($N = 3$ in our implementation). For the portrait sketch generation step, all templates can generate fine-grained portrait sketches, and the user can select the most desirable one as the reference for further drawing. Note that the input contour sketch is not required to contain all facial parts, and the missing parts can be completed automatically with our stroke-mask mapping optimization.

GAN-based neural networks are used in local guidance for mask and portrait sketch generation.

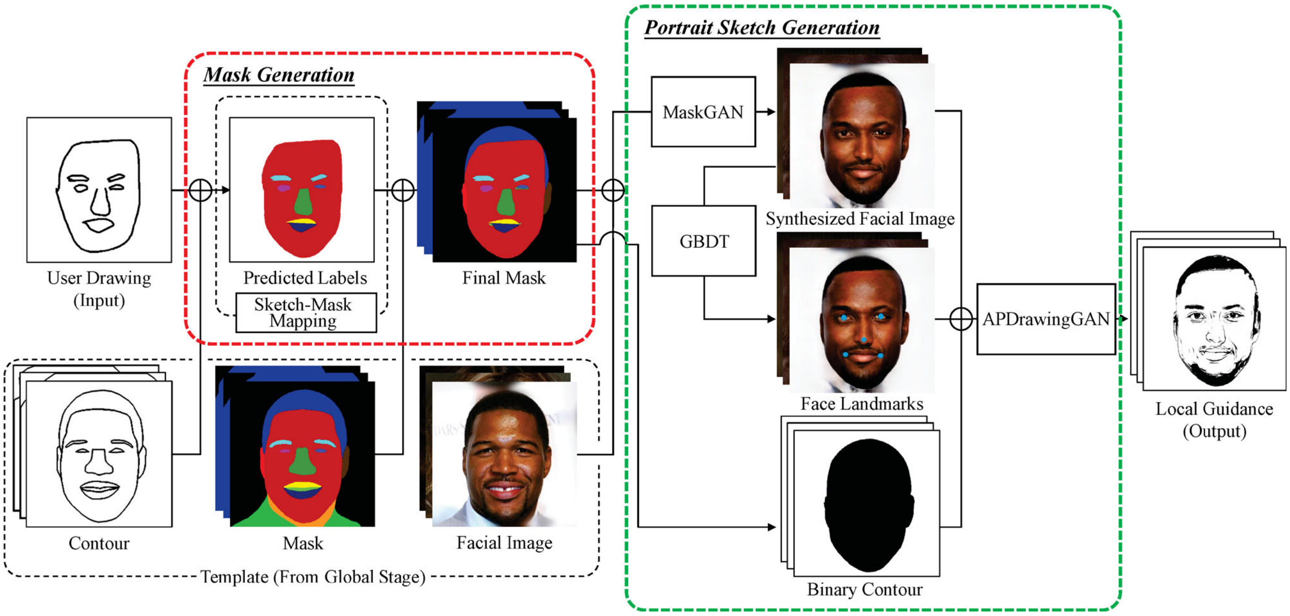


Fig. 4 Local guidance stage consists of two steps: mask generation and portrait sketch generation.

In our implementation, we adopt MaskGAN [26] to generate portrait images matching the facial contour sketch and APdrawingGAN [35] to transfer the portrait images into artistic portrait sketches. Note that the two generative models are trained independently. To connect these two models, facial landmarks are calculated with a Gradient Boosting Decision Tree (GBDT) [38], and the binary background mask is converted from the merged mask.

4.2.2 Mask generation

For portrait image generation, conventional approaches adopt a facial mask with manually defined label information as shown in Fig. 4 (red dashed box is mask generation, with different colors indicating facial labels). However, it is a boring and time consuming task to manually label a portrait drawing. To alleviate the manual labor and adapt it to freehand sketching, we use automatic sketch-to-mask mapping with an optimization algorithm to generate facial mask from the contour sketch from the user drawing.

We first calculate the shape similarity F between user-drawn strokes S and regions of face template mask M. Any single stroke $\boldsymbol{s} \in S$ can be regarded as sequential vertices, where $\boldsymbol{s} = \{\boldsymbol{p}_i \mid i = 1, \cdots, N\}$. Then, we obtain the correspondence between two regions using the following equation:

$$\begin{aligned} F(S, M) &= \min_{\boldsymbol{s}} \sum_{\boldsymbol{s} \in S} \mathrm{Dis}(\boldsymbol{s}, m_k) \\ &= \min_{\boldsymbol{p}} \sum_{\boldsymbol{s} \in S} \Big(\frac{1}{N} \sum_{\boldsymbol{p} \in \boldsymbol{s}} L_2(\boldsymbol{p}, m_k)\Big) \\ &\text{s.t. label}(\boldsymbol{s}) = k \text{ and } m_k \in M \end{aligned} \tag{1}$$

where $\mathrm{Dis}(\boldsymbol{s}, m_k)$ denotes the distance between a single stroke $\boldsymbol{s}$ and m_k (region of M whose label ID is k). $\mathrm{Dis}(\boldsymbol{s}, m_k)$ consists of $\mathrm{dis}(\boldsymbol{p}, M)$, which denotes the average of L_2 distance from all vertices $\boldsymbol{p} \in \boldsymbol{s}$ to m_k. $\mathrm{label}(\boldsymbol{s})$ is the discriminant function to calculate the label ID of $\boldsymbol{s}$ decided by majority vote for vertex $\boldsymbol{p} \in \boldsymbol{s}$, calculated by the following equations:

$$\begin{cases} \mathrm{label}(\boldsymbol{s}) = \mathrm{argmax}_{\boldsymbol{p}}\, C_{\boldsymbol{p} \in \boldsymbol{s}}(V(\boldsymbol{p}, M)) \\ V(\boldsymbol{p}, M) = k^* = \mathrm{argmin}_{\boldsymbol{p}}\ \mathrm{dis}(\boldsymbol{p}, m_k) \end{cases} \tag{2}$$

where $C_{\boldsymbol{p} \in \boldsymbol{s}}(\cdot)$ is the aggregation function for stroke $\boldsymbol{s}$ to count the number of vertices with the same label ID. The discriminant function $V(\boldsymbol{p}, M)$ can determine the label ID k^* for a single vertex $\boldsymbol{p}$ in M by finding the minimum distance of $\boldsymbol{p}$ in each region of M.

The sketch-to-mask mapping algorithm is described in Algorithm 1. The user's strokes are classified using the labels in the matching mask, and strokes with the same labels are merged as a new stroke. Then, a contour (concave hull) of each new stroke is calculated as a new mask to replace the old one in the matching mask.

Algorithm 1 Sketch-to-mask mapping

```
Input: Strokes S, matched mask M
Output: User-defined mask M*
M* = zeros(M.shape);
Number of elements of mask M, m ← len(M);
for k=1:m do
    Merge strokes with same label ms;
    Mask region in same label: mask = M[k];
    if mask is none then
        continue;
    end
    ms= MergeStrokes(s ∈ S s.t. label(s) == k) ;
    if ms is none then
        M*[k] =mask;
    end
    M*[k]=ConcaveHull(ms);
end
return M*
```

In terms of the correspondence between the user sketch and face template, we transfer semantic labels of facial components in the facial template to each region of the user-drawn stroke (e.g., hair, mouth, eyes). Then, we replace the corresponding template regions with ones of user-drawn regions if existing and merge the user's stroke feature into the mask. Note that the contour sketch can be auto-completed even if the user input sketch is partial. Finally, we replace user-drawn regions (partial sketch) and the corresponding template regions, and generate a complete label mask.

4.2.3 Portrait sketch generation

Generating facial images with details from rough sketches is an under-determined problem. An end-to-end GAN-based model requires extensive artistic drawing with similar styles for training, which is expensive and time consuming. To solve this issue, we divide this problem into sketch-to-portrait image generation and artistic rendering for simplification, as shown in Fig. 4. We first generate a realistic facial image using the MaskGAN network based on the complete label mask, corresponding face image, and the face template from the global stage. Then, we convert the face image to a portrait sketch using the APDrawingGAN network for artistic rendering. We obtain the locations of facial components based on GBDT and binary contours of the background from the final mask to connect the two generative networks of mask and portrait sketch generation. Note that global features of the generated local references are restricted by the user's contour sketch.

4.3 Implementation

In our implementation, dualFace was programmed in Python as a real-time drawing application on the Windows 10 platform. A workstation with a 3.7 GHz Intel Core i9 10900KF CPU, two 5.10 GHz NVIDIA RTX2080ti GPUs, and 64 GB RAM was used as the test environment. In addition, 518 images of size 512×512 were taken from the CelebAMask-HQ dataset and converted to contour sketches. GALIF features were extracted for sketch retrieval in the stage of global guidance. For the implementation of local guidance, we used MaskGAN for mask generation consisting of the Dense Mapping Network (DMN) for image generation and U-Net like MaskVAE for mask editing, pre-trained on CelebAMask-HQ with more than 200,000 images. We used APDrawingGAN for portrait sketch generation with a hierarchical GAN structure using U-Net with skip connections for each facial feature (i.e., left eye, right eye, nose, and mouth). In this work, we utilized the pre-trained models with 300 epochs of training on the APDrawing dataset (140 face images and corresponding portrait drawings by an artist).

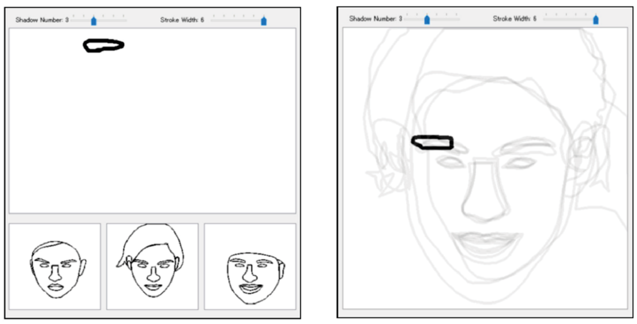
Our prototype system requires, on average, 0.36 s for image retrieval in global guidance after mouse release, and 2.78 s for each portrait image generation in local guidance. Note that image generation was conducted only once, meaning dualFace can provide effective feedback for portrait drawing. Because dualFace generates facial images for local guidance, there are no reference images or labels available as ground truth for quantitative evaluation. Therefore, we conducted a user study to verify our approach in a qualitative way.

5 User study

To evaluate the usefulness of the proposed user interface (UI) for dualFace, we compared dualFace with two conventional drawing interfaces: a suggestive drawing UI (Fig. 1(a)) and shadow drawing UI (Fig. 1(b)). The suggestive drawing UI provides the three most related contours in sub-windows below the main canvas from the face database. The shadow drawing UI provides the blended shadow image from dualFace's global stage, like Shadowdraw [2].

5.1 Evaluation procedure

We invited 14 participants in our study (graduate students, nine males and five females). All participants were asked to draw portraits with a pen tablet (WACOM with 22.4 cm × 14.0 cm drawing area) and an LCD monitor (126.2 cm × 83.7 cm). All participants were asked to draw freely and undirectedly with as many details as possible, using all three drawing interfaces: suggestive UI, shadow UI, and ours, in random order. They first drew freely on the tablet until they felt comfortable using the devices before the user study. We instructed all participants in use of dualFace with a manual. We asked the participants to draw each facial mask in a properly-closed curve. All participants were required to draw carefully and choose the most anticipated references for local guidance from multiple generated candidates after they completed the global stage. Finally, we administered the questionnaire to all participants after they finished the user study.



(a) (b)

Fig. 5 Drawing interfaces used in our user study: (a) suggestive drawing UI and (b) shadow drawing UI.

The questions in our questionnaire were designed to confirm the effectiveness of global and local guidance, and to provide an overall evaluation of using dualFace, as shown in Table 1. All questions adopted a five-point Likert scale (1 for strongly disagree, 5 for strongly agree).

5.2 Drawing evaluation

After all participants completed the comparison study, the other 25 participants joined the online questionnaire for drawing quality evaluation. All participants were asked to score up 12 portrait sketches (four for each drawing UI). We confirmed two questions about the qualities of the spatial relationship and facial details for all portrait sketches. We adopted five-point Likert scales for all questions (1 for very poor, 5 for very good). A good spatial relationship of portrait sketch means the well-balanced facial parts, and good facial details mean that each facial part has finely detailed drawing, such as eyes and mouth. We explained the meanings of the two qualities to all participants before the online questionnaire.

6 Results

We now discuss the implementation results of dualFace, evaluation results, user feedback, and our observations from our user study.

6.1 Visual guidance

Figure 6 shows some examples sketched with dualFace. Users can achieve the desired local guidance according to their freehand contour sketches from the global guidance. If a user's sketch is incomplete, it can be

Table 1 Questionnaire results in our user study

#	Question	Score	Mean	SD
Global	Clear and easy to understand?		4.5	0.6
	Feedback is meaningful and helpful?		4.1	1.0
	Easy to follow and use?		4.1	0.8
Local	Clear and easy to understand?		4.6	0.8
	Feedback is meaningful and helpful?		4.0	1.1
	Easy to follow and use?		4.0	1.0
Overall	Help me learn how to draw faces?		4.2	0.9
	Useful for helping learn how to draw faces?		4.1	0.6
	Useful for helping improve face drawing skill?		3.9	1.2

Fig. 6 Some examples of our implementation results. *User sketch*: results drawn under the global guidance. *Revised contour*: matched facial masks in local guidance. *Local guidance*: generated portrait sketch image for user reference. Portrait images in local guidance were selected by users as closest alternatives to user drawing expectations. *Final result*: final outcome of user drawing.

completed automatically and revised with our sketch-to-mask matching optimization. The last column of Fig. 6 shows an example of a partial sketch. Although the user only has sketched the left eye and eyebrow contour on the global stage, the proposed system can still work well.

In comparison with the drawing interface of ShadowDraw [2], dualFace has no limitation on facial details in drawing guidance. If we blend the relevant templates (face images with details), it is difficult to distinguish the facial references with the loss of facial details, as shown in Fig. 7. Therefore, ShadowDraw can only support drawing guidance of simple subjects without photo-realistic details.

In local guidance, mask generation plays an important role in meeting the user's intent in freehand drawing. To verify this issue, we compared the system results with and without mask generation, as shown in Fig. 8. In the case without mask generation, the feature lines in the user's contour sketch did not conform to the generated portrait drawing, as shown in Fig. 8(left). Meanwhile, more plausible results were achieved by the proposed framework with mask generation.

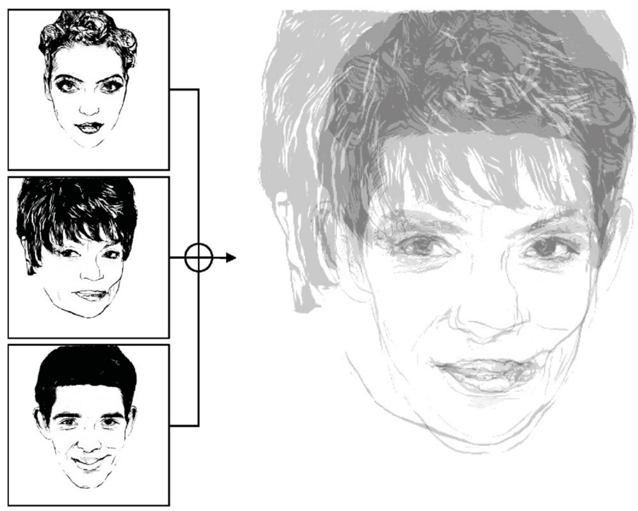


Fig. 7 A limitation of ShadowDraw: the blended image (right) has difficulty preserving details of facial images (left).

6.2 User evaluation

The results of the questionnaire are given in Table 1. Participants were asked to score dualFace by answering nine questions in total (three for global guidance, three for local guidance, and three for overall evaluation). The mean scores of all questions are above 3.9, verifying that the proposed drawing interface dualFace is easy to understand and follow at a high level. For overall user experience, all participants thought our system could help them to draw portraits well and improve their drawing skills. Because dualFace provides guidance on a whole portrait sketch to the participants, users may want

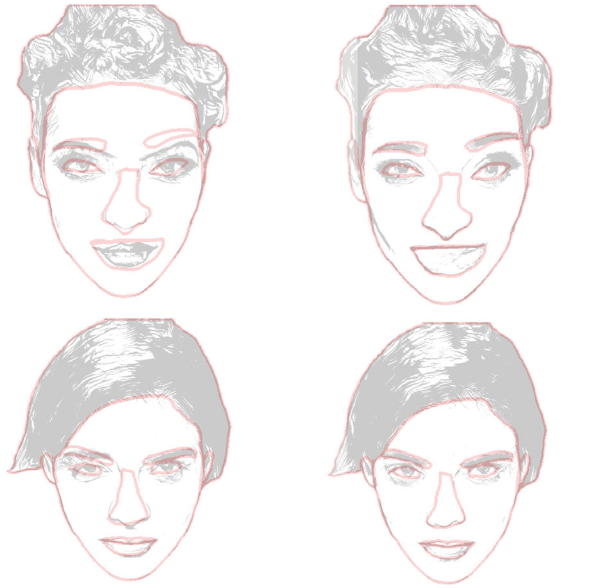


Fig. 8 Results with and without the mask generation process. Mismatches are obvious between the user's contour sketch (red lines) and the generated local guidance without mask generation (left).

to practice basic drawing skills such as arrangements of straight lines or curves. We plan to improve the current drawing interface to help users practice basic drawing skills in the near future.

Figure 9 shows the results of an evaluation study of portrait sketches from our online questionnaire. The proposed drawing interface achieved comparatively high scores in drawing evaluations of both spatial relationship and facial details, with average scores of 4.5 and 4.32, respectively. Therefore, dualFace can guide users to achieve better portrait drawings with correct facial spatial relationships and detailed facial features, whereas the other drawing interfaces may fail to provide them.

Figure 10 shows portrait sketches from our comparison of the suggestive drawing UI, shadow drawing UI, and our dualFace UI. We found that dualFace can not only help users with weak or middling drawing skills to achieve much better portrait sketches, but also help highly-skilled users to complete high quality portrait sketches different from their customary styles of painting. Participants were asked to score their drawing skill using a five-point Likert scale.

Participants' comments about system usage include "I think dualFace is useful, especially for helping freehand drawing." We also received comments about our guidance system, such as "Local guidance with mask generation fits my stroke more than the one without it" and "Local guidance was surely based on my own, but it looked like a monster." All this feedback indicated that mask generation can increase variation in sketches but sometimes generates unnatural facial images. This issue can be solved with other neural rendering approaches or a larger face database. We hope to improve the current prototype to help users draw from different viewpoints and with higher matching rates to user's drawn strokes.

6.3 User satisfaction

To verify user satisfaction in terms of whether the proposed drawing interface helped users match their objectives, we conducted a user evaluation for the three aforementioned interfaces: suggestive drawing UI, shadow drawing UI, and ours (Fig. 5). We recruited 10 graduate students in this evaluation and a questionnaire was conducted afterwards.

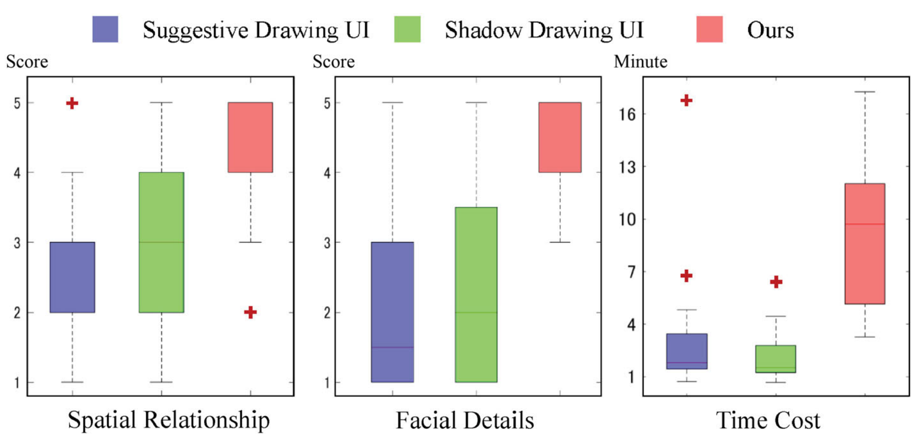


Fig. 9 Left, center: evaluation results of spatial relationship and facial details in portrait drawings. Right: time to complete portrait sketching for each drawing interface.

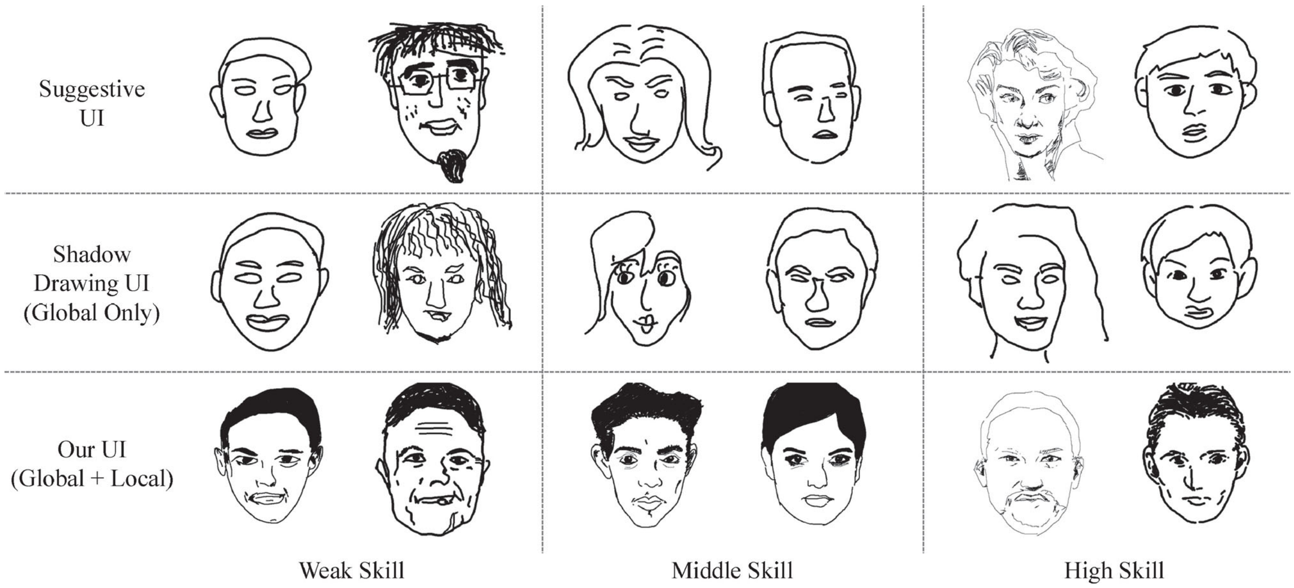


Fig. 10 Drawing results from six participants. Each column corresponds to the same participant's drawing.

We confirmed two ideas in the questionnaire. The average score for the question "Do you think in dualFace your rough sketch matches the detailed guidance as per your expectation?" scored 4.33 (1, not matched at all; 5, well matched). The average score for the question "How would you rate your satisfaction of drawing with dualFace comparing with the other two interfaces?" is 4.44. Therefore, dualFace is verified to enable users to draw portraits that match their visions. We also interviewed the participants for further feedback on user experience. Comments on the final drawings include: "My drawing was better than I thought it would be", "There were plenty of details in my drawing which makes it look better". For usability of dualFace, the users thought that "It is interesting that it generated details accordingly" and "It can automatically generate details, but also beautify the face (drawing)", consistent with our findings in Table 1.

7 Discussion

7.1 Time

We measured the time taken for of portrait drawing for each drawing interface, as shown in Fig. 9. The minimum time among all sketches using dualFace was 4 min 15 s, and the maximum was 17 min 15 s. Although user's drawing skills may differ, the drawing results which took longer led to better drawing results and more facial details. The average time taken is around 10 min. We believe that our local guidance can not only provide enough detailed features for users to follow but also stimulate user's creativity if they intended to spend more time using dualFace.

7.2 System interactivity

Compared to related sketch to facial image generation approaches [39–41], the main contribution of dualFace is providing interactive feedback to users, helping them to improve their drawing skills. In earlier work, users cannot get any help from the system until the drawing is completed, so it is usually essential for user's to have drawing skills. Although DeepFaceDrawing can generate a high-quality facial image from a rough sketch with shadow guidance [42], it is difficult to improve user's drawing skills because they use edge maps extracted from images as guidance without separating local and global facial information. In contrast, dualFace can provide interactive sketch support with two-stage guidance for both global features and local facial details. Our system can provide balanced facial information in real time, so that users, especially novices, can concentrate on learning how to sketch a balanced facial contour.

7.3 Generation diversity

To meet user's drawing expectations, it is necessary to ensure generation diversity in the facial image database. In this work, facial diversity could be influenced by database size and mask generation. However, the best size of the image database for retrieval in the global stage is a hyper-parameter because it is difficult to find a suitable criterion to automatically evaluate whether generative guidance of dualFace matches user's vision, for size optimization. In our implementation, thus, we selected around 500 typical facial images manually covering different facial types and shapes of facial parts. Our selection strategy ensured completed facial parts with clear contours in a frontal view, avoid overlapping parts with hair or glasses. For mask generation, we can improve the diversity of the generated results for local guidance with multiple references. Figure 11 shows the facial references to users from the global stage from which users can select the most satisfying generated image as local guidance for facial detail drawing. All references maintained the shape restrictions (red lines) from sketch input.

8 Conclusions

In this work, we proposed a portrait drawing interface with two-stage global and local guidance. First, we generate a shadow image to provide locations of facial parts when drawing strokes as global guidance. After specifying locations of facial parts as a contour sketch, we then generate detailed facial images from user contour sketches with face mask and portrait drawing generation networks in local guidance. The proposed user interface, dualFace, was verified to be useful and satisfactory in portrait drawing for users with different levels of drawing skill. We believe that our work contributes to accelerate freehand drawing interfaces.

Because the proposed system converts user's sketches to masks by matching strokes with the example mask, the contour sketch must contain exact shape information. DualFace can only support drawing portraits in a realistic style due to use of real photos in the face database. It is difficult to achieve high-level semantic sketches such as emotional faces and exaggerated cartoon style drawing, because it is currently challenging to determine the shapes of facial parts. If the strokes for facial parts are not closed curves, this may lead to indeterminate contours of facial parts. Figure 12 shows an input sketch with a smiling face which generates a strange mask with two separate pieces of nose. We plan to improve the representation of facial sketches and increase the robustness of dualFace, and weigh user's intent and portrait quality.

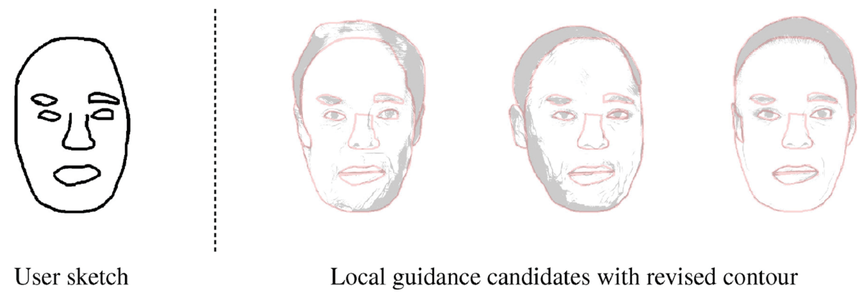


Fig. 11 Multiple reference candidates (right) generated from the user sketch (left) for local drawing guidance.

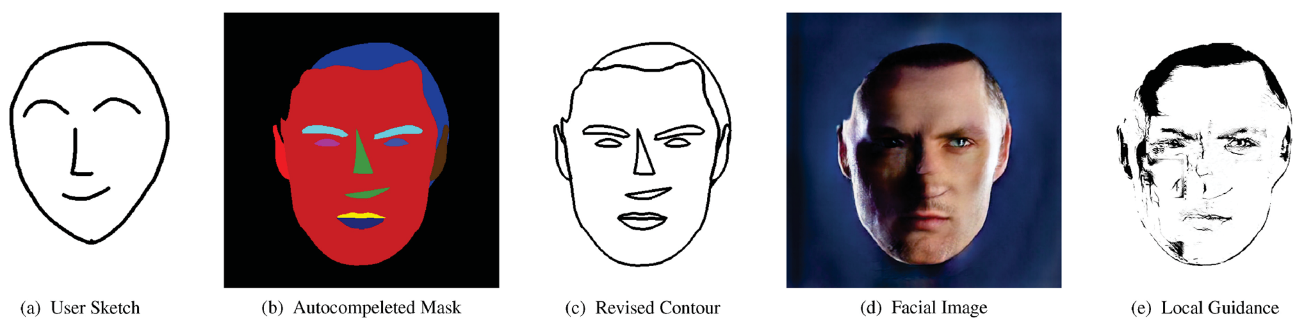


Fig. 12 Limitations of our work. An abstract sketch may not be converted to a reasonable mask (b): the mouth in the user's contour sketch (a) is wrongly regarded as a part of the nose. This caused degeneration of the generated image (d) and local guidance (e).

Acknowledgements

We thank the reviewers for their valuable comments. We also thank Toby Chong for his discussion of ideas. This work was supported by Grant from Tateishi Science and Technology Foundation, JSPS KAKENHI grant JP20K19845 and JP19K20316, Japan.

Electronic Supplementary Material Supplementary material is available in the online version of this article at `https://doi.org/10.1007/s41095-021-0227-7`.

References

[1] Xie, J.; Hertzmann, A.; Li, W.; Winnemöller, H. PortraitSketch: Face sketching assistance for novices. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, 407–417, 2014.

[2] Lee, Y. J.; Zitnick, C. L.; Cohen, M. F. ShadowDraw: Real-time user guidance for freehand drawing. *ACM Transactions on Graphics* Vol. 30, No. 4, Article No. 27, 2011.

[3] Choi, J.; Cho, H.; Song, J.; Yoon, S. M. SketchHelper: Real-time stroke guidance for freehand sketch retrieval. *IEEE Transactions on Multimedia* Vol. 21, No. 8, 2083–2092, 2019.

[4] Ghosh, A.; Zhang, R.; Dokania, P.; Wang, O.; Efros, A.; Torr, P.; Shechtman, E. Interactive sketch & fill: Multiclass sketch-to-image translation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 1171–1180, 2019.

[5] Zhu, J. Y.; Krähenbühl, P.; Shechtman, E.; Efros, A. A. Generative visual manipulation on the natural image manifold. In: *Computer Vision – ECCV 2016. Lecture Notes in Computer Science, Vol. 9909.* Leibe, B.; Matas, J.; Sebe, N.; Welling, M. Eds. Springer Cham, 597–613, 2016.

[6] Bradley, B. Drawing People: How to Portray the Clothed Figure. North Light/Writers Digest, 2003. Available at https://books.google.co.jp/books?id=lhdjDwAAQBAJ.

[7] Liu, L.; Shen, F. M.; Shen, Y. M.; Liu, X. L.; Shao, L. Deep sketch hashing: Fast free-hand sketch-based image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2298–2307, 2017.

[8] Yu, Q.; Liu, F.; Song, Y. Z.; Xiang, T.; Hospedales, T. M.; Loy, C. C. Sketch me that shoe. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 799–807, 2016.

[9] Dekel, T.; Gan, C.; Krishnan, D.; Liu, C.; Freeman, W. T. Sparse, smart contours to represent and edit images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3511–3520, 2018.

[10] Portenier, T.; Hu, Q. Y.; Szabó, A.; Bigdeli, S. A.; Favaro, P.; Zwicker, M. FaceShop: Deep sketch-based face image editing. *ACM Transactions on Graphics* Vol. 37, No. 4, Article No. 99, 2018.

[11] Tseng, H. Y.; Fisher, M.; Lu, J. W.; Li, Y. J.; Kim, V.; Yang, M. H. Modeling artistic workflows for image generation and editing. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science, Vol. 12363.* Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 158–174, 2020.

[12] Yang, S.; Wang, Z. Y.; Liu, J. Y.; Guo, Z. M. Deep plastic surgery: Robust and controllable image editing with human-drawn sketches. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science, Vol. 12360.* Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 601–617, 2020.

[13] Hu, Z. Y.; Xie, H. R.; Fukusato, T.; Sato, T.; Igarashi, T. Sketch2VF: Sketch-based flow design with conditional generative adversarial network. *Computer Animation and Virtual Worlds* Vol. 30, Nos. 3–4, e1889, 2019.

[14] Peng, Y. C.; Mishima, Y.; Igarashi, Y.; Miyauchi, R.; Okawa, M.; Xie, H. R.; Miyata, K. Sketch2Domino: Interactive chain reaction design and guidance. In: Proceedings of the Nicograph International, 32–38, 2020.

[15] Fukusato, T.; Noh, S.-T.; Igarashi, T.; Ito, D. Interactive meshing of user-defined point sets. *Journal of Computer Graphics Techniques* Vol. 9, No. 3, 39–58, 2020.

[16] Igarashi, T.; Hughes, J. F. A suggestive interface for 3D drawing. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, 173–181, 2001.

[17] He, Y.; Xie, H. R.; Zhang, C.; Yang, X.; Miyata, K. Sketch-based normal map generation with geometric sampling. In: Proceedings of the SPIE 11766, International Workshop on Advanced Imaging Technology, 117661B, 2021.

[18] Flagg, M.; Rehg, J. M. Projector-guided painting. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, 235–244, 2006.

[19] Igarashi, T.; Matsuoka, S.; Kawachiya, S.; Tanaka, H. Interactive beautification: A technique for rapid geometric design. In: Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology, 105–114, 1997.

[20] Laviole, J.; Hachet, M. PapARt: Interactive 3D graphics and multi-touch augmented paper for artistic creation. In: Proceedings of the IEEE Symposium on 3D User Interfaces, 3–6, 2012.

[21] Dixon, D.; Prasad, M.; Hammond, T. iCanDraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces. In: Proceedings of the 28th International Conference on Human Factors In Computing Systems, 897–906, 2010.

[22] Iarussi, E.; Bousseau, A.; Tsandilas, T. The drawing assistant: Automated drawing guidance and feedback from photographs. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, 183–192, 2013.

[23] Su, Q.; Li, W. H. A.; Wang, J.; Fu, H. Ez-sketching: Three-level optimization for error-tolerant image tracing. *ACM Transactions on Graphics* Vol. 33, No. 4, Article No. 54, 2014.

[24] He, Z. Z.; Xie, H. R.; Miyata, K. Interactive projection system for calligraphy practice. In: Proceedings of the Nicograph International, 55–61, 2020.

[25] Matsui, Y.; Shiratori, T.; Aizawa, K. DrawFromDrawings: 2D drawing assistance via stroke interpolation with a sketch database. *IEEE Transactions on Visualization and Computer Graphics* Vol. 23, No. 7, 1852–1862, 2017.

[26] Lee, C. H.; Liu, Z. W.; Wu, L. Y.; Luo, P. MaskGAN: Towards diverse and interactive facial image manipulation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5548–5557, 2020.

[27] Rosin, P. L.; Mould, D.; Berger, I.; Collomosse, J. P.; Lai, Y.-K.; Li, C.; Li, H.; Shamir, A.; Wand, M.; Wang, T.; Winnemöller, H. Benchmarking non-photorealistic rendering of portraits. In: Proceedings of the Symposium on Non-Photorealistic Animation and Rendering, Article No. 11, 2017.

[28] Chen, Z. Y.; Zhou, J. Y.; Gao, X. Y.; Li, L. S.; Liu, J. F. A novel method for pencil drawing generation in non-photo-realistic rendering. In: *Advances in Multimedia Information Processing - PCM 2008. Lecture Notes in Computer Science, Vol. 5353*. Huang, Y. M. R. et al. Eds. Springer Berlin Heidelberg, 931–934, 2008.

[29] Xie, D. G.; Zhao, Y.; Xu, D. An efficient approach for generating pencil filter and its implementation on GPU. In: Proceedings of the 10th IEEE International Conference on Computer-Aided Design and Computer Graphics, 185–190, 2007.

[30] Zhang, J. W.; Wang, R. Z.; Dan, X. Automatic genaration of sketch-like pencil drawing from image. In: Proceedings of the IEEE International Conference on Multimedia & Expo Workshops, 261–266, 2017.

[31] Li, C.; Wand, M. Combining Markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2479–2486, 2016.

[32] Liao, J.; Yao, Y.; Yuan, L.; Hua, G.; Kang, S. B. Visual attribute transfer through deep image analogy. *ACM Transactions on Graphics* Vol. 36, No. 4, Article No. 120, 2017.

[33] Silva, F.; de Castro, P. A. L.; Júnior, H. R.; Marujo, E. C. ManGAN: Assisting colorization of manga characters concept art using conditional GAN. In: Proceedings of the IEEE International Conference on Image Processing, 3257–3261, 2019.

[34] Zhu, J. Y.; Park, T.; Isola, P.; Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2242–2251, 2017.

[35] Yi, R.; Liu, Y. J.; Lai, Y. K.; Rosin, P. L. APDrawingGAN: Generating artistic portrait drawings from face photos with hierarchical GANs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10735–10744, 2019.

[36] Eitz, M.; Richter, R.; Boubekeur, T.; Hildebrand, K.; Alexa, M. Sketch-based shape retrieval. *ACM Transactions on Graphics* Vol. 31, No. 4, Article No. 31, 2012.

[37] Yu, C. Q.; Wang, J. B.; Peng, C.; Gao, C. X.; Yu, G.; Sang, N. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. In: *Computer Vision – ECCV 2018. Lecture Notes in Computer Science, Vol. 11217*. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Eds. Springer Cham, 334–349, 2018.

[38] Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1867–1874, 2014.

[39] Li, Y. H.; Chen, X. J.; Wu, F.; Zha, Z. J. LinesToFacePhoto: Face photo generation from lines with conditional self-attention generative adversarial networks. In: Proceedings of the 27th ACM International Conference on Multimedia, 2323–2331, 2019.

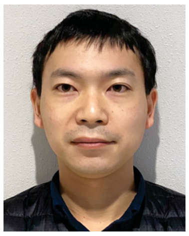
[40] Li, Y. H.; Chen, X. J.; Yang, B. X.; Chen, Z. H.; Cheng, Z. H.; Zha, Z. J. DeepFacePencil: Creating face images from freehand sketches. In: Proceedings of the 28th ACM International Conference on Multimedia, 991–999, 2020.

[41] Richardson, E.; Alaluf, Y.; Patashnik, O.; Nitzan, Y.; Cohen-Or, D. Encoding in style: A StyleGAN encoder for image-to-image translation. *arXiv preprint* arXiv:2008.00951v1, 2020.

[42] Chen, S.-Y.; Su, W.; Gao, L.; Xia, S.; Fu, H. DeepFaceDrawing: Deep generation of face images from sketches. *ACM Transactions on Graphics* Vol. 39, No. 4, Article No. 72, 2020.

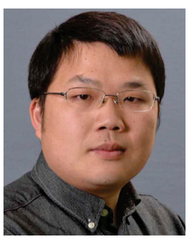
Zhengyu Huang is currently a Ph.D. candidate at Japan Advanced Institute of Science and Technology. He is engaged in research on computer graphics and human–computer interaction. He received his B.E. degree at Jinan University and M.E. degree at Northwest A&F University, China, in 2012 and 2019, respectively.

Yichen Peng is currently a master student at Japan Advanced Institute of Science and Technology. He is graduated from the Guangdong University of Technology in 2017, majoring in animation design. His research interests are in human computer interaction and computer graphics.

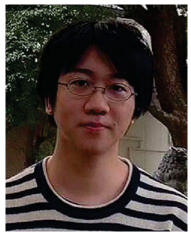


Tomohiro Hibino is currently a Ph.D. candidate at Japan Advanced Institute of Science and Technology. He is graduated from the University of Tokyo in 2013, majoring in agricultural informatics. His research interests are in virtual reality and computer graphics.

Chunqi Zhao is currently a Ph.D. student in the Creative Informatics Department in the University of Tokyo. He graduated from Fudan University with a B.E. degree in 2017 and from the University of Tokyo with an M.E. degree in 2020. His research interests include human–computer interaction and representation learning. He mainly works on improving experience for programming and creativity.



Haoran Xie has been a senior lecturer at Japan Advanced Institute of Science and Technology (JAIST) since 2020. He received his Ph.D. degree from JAIST in 2015. After that, he worked in the Computer Science Department in the University of Tokyo (2015–2018), and was an assistant professor in JAIST (2018–2020). His main research interests are in interactive computer graphics and user interfaces. He is a member of ACM.



Tsukasa Fukusato has been an assistant professor in the Computer Science Department in the University of Tokyo since 2017. He received his Ph.D. degree from the Department of Pure and Applied Physics at Waseda University in 2017. His main research areas are computer graphics and human computer interaction such as cartoon design. He is a member of ACM.

Kazunori Miyata has been a professor at the Japan Advanced Institute of Science and Technology (JAIST) since 2002. Prior to joining JAIST, he was an associate professor in the Department of Imaging Art at the Tokyo Institute of Polytechnics (1998–2002), and a researcher at IBM Japan (1984–1998). He received his B.S. degree from Tohoku University in 1984, and his M.S. and Ph.D. degrees from Tokyo Institute of Technology in 1986 and 1997 respectively. His research interests include computer graphics, media art, and multimedia applications. He is a member of ACM and IEEE.