

Special Section on ICAT-EGVE 2020

GAN-based image-to-friction generation for tactile simulation of fabric material

Shaoyu Cai ^{a,1}, Lu Zhao ^{b,1}, Yuki Ban ^c, Takuji Narumi ^d, Yue Liu ^{b,e,*}, Kening Zhu ^{a,f,*}^a School of Creative Media, City University of Hong Kong, Hong Kong Special Administrative Region of China^b Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing, PR China^c Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan^d Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan^e AICFVE of Beijing Film Academy, Beijing, PR China^f City University of Hong Kong Shenzhen Research Institute, Shenzhen, PR China

ARTICLE INFO

Article history:

Received 5 March 2021

Received in revised form 20 August 2021

Accepted 21 September 2021

Available online 25 September 2021

Keywords:

Supervised learning

Generative adversarial networks (GANs)

Haptic rendering

Electrovibration surface

Tactile simulation

Fabrics

ABSTRACT

The electrovibration tactile display could render the tactile feeling of different textured surfaces by generating the frictional force through voltage modulation. When a user is sliding his/her finger on the display surface, he/she can feel the frictional texture. However, it is not trivial to prepare and fine-tune the appropriate frictional signals for haptic design and texture simulation. In this paper, we present a deep-learning-based framework to generate the frictional signals from the textured images of fabric materials. The generated frictional signal can be used for the tactile rendering on the electrovibration tactile display. Leveraging GANs (Generative Adversarial Networks), our system could generate the displacement-based data of frictional coefficients for the tactile display to simulate the tactile feedback of different fabric materials. Our experimental results show that the proposed generative model could generate the frictional-coefficient signals visually and statistically close to the ground-truth signals. The following user studies on fabric-texture simulation show that users could not discriminate the generated and the ground-truth frictional signals being rendered on the electrovibration tactile display, suggesting the effectiveness of our deep-frictional-signal-generation model.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Various types of haptic feedback, such as electrovibration feedback [1], vibrotactile feedback [2,3], and thermal feedback [4,5], could be rendered physically on hardware devices. These devices allow users to understand the physical properties of virtual environments and can be used to improve the immersion and the realness in Virtual Reality (VR) [6–9]. Among different haptic-rendering devices, electrovibration tactile displays [1] could provide the frictional force feedback on the bare fingertips of users to simulate different surface properties, such as the geometric shapes [10,11], and the roughness of fabric textures [12]. This technology could be potentially applied for tactile simulation and rendering in VR [13], to allow users to feel and explore

the texture and the shapes in the virtual environments through haptic feedback [14–18].

However, the process of designing the appropriate frictional signals for tactile texture simulation could be challenging due to two main reasons. Firstly, the lateral forces are usually dependent upon the frictional properties, such as the frictional coefficient, the normal pressure force, and the displacements, on the real physical textured surface. However, most textured surfaces show irregular changes in the lateral force during the finger-sliding movement on the surface. This makes it challenging to build a general rule for the frictional simulation and prediction for a textured surface. Secondly, while it could be feasible to use the recorded sensor data to render tactile feedback on the electrovibration tactile display [19,20], it is often costly and time-consuming for data collection.

To address these gaps, we propose FrictGAN, a frictional-signal-generation framework based on Generative Adversarial Networks (GANs) [21] for tactile simulation of fabric material. As shown in Fig. 1, FrictGAN takes the visual images of the fabric materials as input, and generates the frictional signals accordingly. We define the frictional signals as the frictional coefficients on the finger-contact positions of the electrovibration

* Corresponding authors.

E-mail addresses: shaoyu.cai@my.cityu.edu.hk (S. Cai), lu.zhao@outlook.com (L. Zhao), ban@edu.k.u-tokyo.ac.jp (Y. Ban), narumi@cyber.t.u-tokyo.ac.jp (T. Narumi), liuyue@bit.edu.cn (Y. Liu), keninzhu@cityu.edu.hk (K. Zhu).

1 Authors contributed equally.

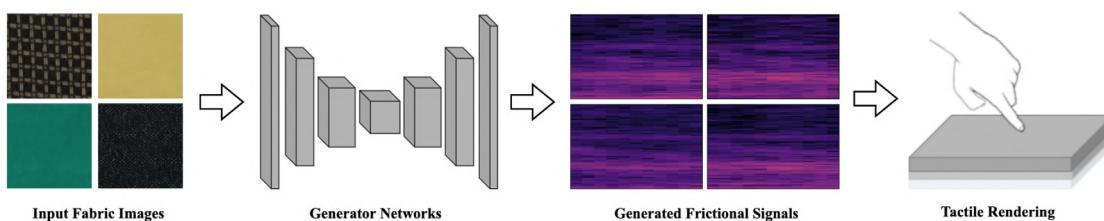


Fig. 1. The concept of our image-to-friction generation using the GAN-based method. Users could input the RGB fabric image to the Generator Network, and it will synthesize the frictional signals that could be applied on the electrovibration tactile display to simulate the haptic texture of the input fabrics.

tactile surface. Using HapTex database [22], we build a visual-to-frictional data set for training a conditional-GAN-based deep neural network [23], to generate the amplitude spectrogram for representing the frictional-coefficients data. The series of frictional coefficients further controls the electrovibration tactile display to control the lateral force corresponding to the user's finger displacement on the display surface. Extending our previous work [24], we evaluate the FrictGAN model by computing the root-mean-squared errors (RMSE) and t-Distributed Stochastic Neighbor Embedding (t-SNE) [25] of the ground-truth and the generated signals, and conduct a series of user-perception experiments to investigate the effectiveness of the FrictGAN-generated frictional signal on tactile simulation of fabrics. Our experimental results of frictional-signal generation show that FrictGAN could generate the amplitude spectrograms and the waveforms of the frictional-coefficients data visually and statistically close to the ground truth. Our user-perception studies show that the proposed model could generate the frictional signals suitable for tactile simulation both on trained and unseen/unknown types of fabric materials, with no significant difference in the user perception compared to the pre-recorded frictional signals from the real physical fabrics. Our source code, data set, and documentation are available at: <https://github.com/shaoyuca/Image-to-Friction-Generation>.

2. Related works

2.1. Texture simulation on electrovibration displays

The electrovibration tactile display can generate changeable electrostatic forces between the sliding fingertip and the display surface by modulating the driving voltages, to create a controllable attraction force and alter the surface friction [26]. Bueno et al. [12] found that contact friction plays an important role in the tactile perception of textile fabrics during bare fingertip sliding, suggesting the potential of fabric texture simulation using the electrovibration surface. However, precisely controlling the tactile signals on the electrovibration tactile display for high-fidelity texture simulation remains a challenge. Some previous works simulated the visual textures or the geometric shapes on the electrovibration tactile display through periodic wave signals [27,28]. However, such solutions may not be effective for simulating heterogeneous textured surfaces, such as real-world fabrics, because the periodic wave signals could not reflect the anisotropic change of tactile properties on the surfaces.

Recently, the data-driven or measurement-based methods for tactile modeling and rendering has emerged [29]. The data-driven algorithms with the sensor data pre-recorded from real-world physical materials could reproduce the change of tactile characteristics at different positions on the material surface. Ilkhani et al. showed that such data-driven methods could outperform the periodic wave signals on the texture rendering, especially for the anisotropic haptic textures (e.g., plastics) [20,30]. Osgouei et al. [31] presented an inverse dynamics model for generating

the voltage signals to mimic real-world textures on the electrovibration tactile display. As an extended work [32], they conducted user studies to compare the tactile feedback generated by the pre-recorded signals with the real-world physical material surfaces, and the results showed the model could achieve a high reality of rendering virtual textures on the electrovibration tactile display. Zhao et al. [33] simulated the tactile perception based on a public haptic database [34] by weighting the friction signal and the acceleration signal of the material surface. They also considered the real-time exploring speed of the user's finger in the tactile simulation. Jiao et al. [19] proposed a data-driven tactile-rendering algorithm based on the displacement-based frictional coefficients during the finger-sliding movement on the fabrics. With the same data-acquisition system, Jiao et al. constructed HapTex [22], a database of frictional coefficients recorded while a finger is sliding on 120 types of fabric surfaces.

While the aforementioned data-driven approaches could achieve considerable performance on tactile texture rendering and simulation, they often required pre-recorded frictional signals at hand. However, the signal-recording procedure could be high-cost and time-consuming. As HapTex [22] provides an aggregated database, including both visual and frictional tactile signals, of bare fingertips sliding on 120 types of fabrics, it could be potentially used for predicting the frictional signals from the visual image of the surface. Similar works have been conducted in the deep image-to-vibration generation [35–37]. In this paper, we adopt a GAN-based deep-learning structure on image-to-friction generation and show that the generated frictional signals could be rendered on an electrovibration tactile display for realistic fabric texture rendering.

2.2. Tactile signals synthesis from visual images

Some prior works have explored the image-based tactile rendering on electrovibration tactile displays. Wang and Sun [38] proposed a tactile rendering algorithm by extracting the shape features from the input visual images. Wu et al. [39] presented an image-based generative model for tactile signals by mapping electrostatic signals and image textures. Specifically, they used the Roberts filter to extract the texture information from an image, and modulated the friction between the fingertip and the electrostatic display. Similarly, Kim et al. [28] extracted the local gradient geometric features of visual images to create the tactile perception of 3D shapes on the electrovibration tactile device. These methods inferred the tactile signals based on the image textures, but the generated tactile data might not be suitable for real-world material simulation as it is not based on the real physical material samples.

Recently, Generative Adversarial Networks (GANs) show the considerable performance on high-dimensional data generation, such as image-to-image [23,40], text-to-image [41,42], and audio-to-image [43,44]. Some researchers also adopted the GAN-based methods for cross-modal visual-tactile data generation to predict the tactile information (e.g., vibrotactile/force/thermal signals)

based on the visual input (e.g., texture/material images or videos). As a preliminary attempt, Ujitoko and Ban [35] presented TactGAN, for the vibrotactile signal generation from the textured-surface images or the label attributes. Similarly, Li et al. [45] proposed a GAN-based method to learn the visual-tactile association by identifying the category labels of input images to guide the generation of the vibrotactile signals. Liu et al. [36] built the image-to-tactile cross-modal perception based on CycleGAN [40], and developed a handheld vibration device to render the generated tactile signals for visually impaired people.

The works mentioned above focused on vibrotactile signal generation with the availability of the comprehensive data sets [34,46]. To our best knowledge, there is no attempt for image-based frictional signal generation for real-world material simulation. Generally, GAN could generate a series of data, such as image super-resolution [47], sketch-based image generation [48] and audio-visual generation [43], that are closed to the “real” data samples corresponding to the input from the same or other domain. Noted that the CNN-based encoder has also been used in data forecast, to construct the functional mapping between the input and the output data based on the distributions in the latent space [49]. However, existing research showed that GAN-based generative networks outperformed the CNN-based solution in cross-modal data generation [50]. In this work, we propose a GAN-based framework for frictional signal generation from the visual images based on the augmented HapTex database [22].

3. Method and implementation

We aim to build the mapping between vision (e.g., visual images) and touch (e.g., frictional signals on the electrovibration tactile surfaces) for tactile simulation of the fabric surface. This could be considered as the problem of cross-modal visual-to-tactile generation. Inspired by the previous works on cross-modal data generation [36,41,43] and data-driven texture rendering on electrostatic tactile displays [19,32], we propose a deep-learning-based method of frictional-coefficients data generation for the tactile simulation of fabrics.

Fig. 2 shows the framework of our frictional-coefficients generation and rendering system. The system first converts the RGB visual image of fabric material into a gray-scale, which will be taken as the input for the GAN-based generative model, and generates the waveform-based frictional-coefficients data to determine the driving voltages on the electrovibration tactile display for tactile rendering of the fabric materials.

3.1. Frictional-coefficient signals generation model

The displacement-based frictional coefficients data could be processed as the spatial time-series signals, and be converted into 2D representations (e.g., amplitude spectrograms). Inspired by the previous work on cross-modal visual-tactile data generation [35,36], we adopt the structure of conditional GAN (cGAN) [51] as the backbone of our signal-generation model, and utilize the Griffin-Lim algorithm [52] for converting the generated amplitude spectrograms to the waveform of the frictional-coefficients data.

3.1.1. Model architecture

The green part of **Fig. 2** demonstrates the architecture of our proposed frictional-coefficients generation model. The model consists of a pair of generator G and discriminator D that are commonly adopted in the cGAN-based framework for image-to-image generation [23]. The generator G contains two parts, namely an encoder network and a decoder network. The encoder network takes the visual image as input, and outputs a 128-dimensional

latent vector. The decoder synthesizes the amplitude spectrogram from the latent space represented as the concatenation of the latent vector and a randomized 50-dimensional noise vector. The generated and the ground-truth spectrograms and the cropped visual image (i.e. input) are then passed to the discriminator D to distinguish the ground-truth and the generated amplitude spectrograms based on the conditional input. The final generated spectrogram is converted into the waveform by the Griffin-Lim algorithm [52].

We adopt the Convolutional-Neural-Networks (CNN) structure for all the network components in our model. For the generator G , we implement the U-net backbone [53] to build a series of skip connections between the layer i and the layer $n - i$ of the generator G where n represents the total number of layers in G . In addition, we implement the down-sampling and the up-sampling layers, the batch normalization, and the ReLU units in G ; and three dropout units are added in the first three layers of the decoder. We also add a ReLU function in the last layer of the generator for the spectrogram generation in the final stage. For the discriminator D , we adopt the PatchGAN structure [23] which takes a channel-wise concatenated vector from the cropped input image and the generated/ground-truth amplitude spectrograms as the final input, and outputs a patch-based vector. Each layer in the discriminator consists of the down-sampling layer, the batch normalization, and the Leaky ReLU units. After finishing the training procedure, we remove the discriminator D and only utilize the generator G to generate the amplitude spectrogram, and convert it to the wave-format frictional signals for tactile rendering on the electrovibration tactile display. For more details on the network architecture, please refer to our source code.

3.1.2. Objective functions

We adopt the cGAN structure to learn the mapping from the image data x and the random noise vector z , to the amplitude spectrogram y . The objective mapping of the model is as below:

$$G : x, z \rightarrow y. \quad (1)$$

Following the original cGAN [51] structure, we define the objective function of our proposed model as:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{(x)}, y \sim p_{(y)}} [\log D(y|x)] + \mathbb{E}_{x \sim p_{(x)}, z \sim p_{(z)}} [\log(1 - D(G(z, x)|x))]. \quad (2)$$

where $p_{(x)}$, $p_{(y)}$ and $p_{(z)}$ represent the distribution of the image domain, the spectrogram domain and the random noise, respectively. However, using only this original cGAN objective function (Eq. (2)) may cause gradient vanishing [54]. To solve the problem, we implement the Wasserstein GAN (WGAN) [55] for more stable generator G and discriminator D training. That is, we replace the original GAN loss as the Wasserstein GAN loss L_{WGAN} as below:

$$L_{\text{WGAN}} = -\mathbb{E}_{x \sim p_{(x)}, y \sim p_{(y)}} [D(y|x)] + \mathbb{E}_{x \sim p_{(x)}, \tilde{y} \sim p_{(\tilde{y})}} [D(\tilde{y}|x)] \quad (3)$$

Inspired by the previous work on spectrogram-based tactile signal generation [36], we also include the Manhattan distance, which is also known as the L_1 distance, as our pixel-wise loss. Thus, the final objective function becomes:

$$\arg \min_G \max_D L_{\text{WGAN}} + \lambda L_{L_1} \quad (4)$$

In this equation, L_{WGAN} is the WGAN-loss, and L_{L_1} is the pixel-wise L_1 loss between the real and the generated spectrograms. λ is the hyper-parameter, which is set as 100.

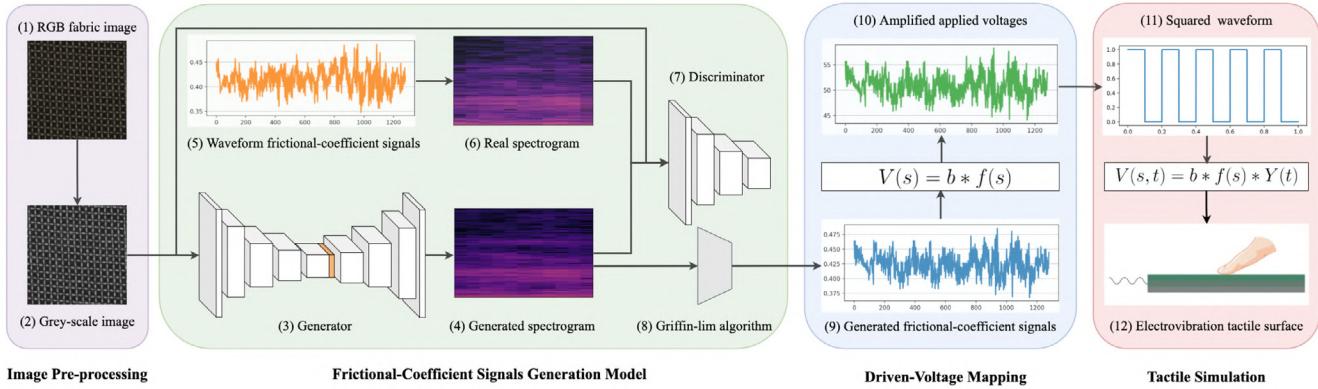


Fig. 2. The illustration of our system implementation. Firstly, the RGB fabric image (1) is converted as the gray-scale image as the texture image (2) during the pre-processing procedure and fed into the Generator (3) with a randomized 50-d noise vector (the orange part) in the latent space for our spectrogram generation (4). Meanwhile, the pre-recorded frictional-coefficients data (5) is computed as the spectrogram format (6) through STFT as the ground-truth data and passed into the discriminator (7) with the concatenation of the cropped gray-scale image. After the training stage, we remove the discriminator (7) and utilize the Griffin-Lim algorithm (8) to transform the generated spectrogram into the wave-format frictional-coefficient signal (9). It will be mapped to the applied voltage (10). Then we modulate the driving voltage and the frequency of the squared carrier wave (11) for the electrovibration tactile displays rendering (12). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Texture modeling and rendering

Following the procedure adopted in HapTex [22], we map the driving voltages to the frictional coefficients, and modulate the carrier wave for texture simulation on the electrovibration tactile display. We amplify the frictional-coefficients signals as the applied voltages at the specific position on the electrovibration display surface according to the following formula:

$$V(s) = b * f(s) \quad (5)$$

where $V(s)$ denotes the applied voltage at the position s on the electrovibration tactile display, b is the amplification factor, and $f(s)$ represents the value of the displacement-based frictional coefficients signal. As the haptic perception of electrovibration depends on the spatial and the temporal change of the waveform signal [56], we add a high-frequency carrier-wave function with the frequency f_c denoted as $Y(t)$ to modulate the driving voltage on the electrovibration tactile display. Specifically, the driving voltage at the position s and the timestamp t could be calculated as:

$$V(s, t) = b * f(s) * Y(t) \quad (6)$$

The amplification factor b and the carrier wave frequency f_c could be determined by users subjectively matching the virtual and the physical textures [19]. We will discuss this process in Section 6.

4. Data preparation

The HapTex database [22] presents both visual images and position-based frictional coefficients collected during the bare-finger sliding on the corresponding fabric surface. The previous work on haptic texture rendering [19] shows that the frictional-coefficient data pre-recorded from the fabric surfaces could be used for rendering realistic virtual texture on the electrovibration tactile display. The HapTex database [22] was constructed under a similar process using the same data-acquisition system as the aforementioned study [19]. Therefore, it is reasonable to assume that it would be effective to adopt the HapTex data for texture rendering on an electrovibration tactile display.

This database could be used to create haptic textures on the electrovibration tactile display device through psychophysical experiments [19]. HapTex contains in total of 120 types of fabric material within 10 categories (i.e. Velvet, Cotton, Leather, Fiber, Chiffon, Wool, Nylon, Polyester, Linen, and Silk). To train our visual-to-frictional generation model, we select 10 types of fabric materials (randomly choose one from each category) to construct our data set for model training and testing. Note that we exclude some contaminated samples which contain stains in the visual image, as such features may affect the process of model training and signal generation. The selected samples (from F1 to F10) are shown in the first column of Fig. 3.

4.1. Visual data & frictional tactile data

We take the visual image of the fabric material as the input for our frictional-signal generation model. The original HapTex database [22] contains the RGB images of different types of fabric materials in the resolution of 2362×2362 pixels mapping to the size of 200×200 mm 2 of the physical fabric samples. We crop these original RGB images into the size of 1024×1024 pixels, mapping to 100×100 mm 2 in the physical material, as the input size of our signal-generation model. Since the tactile signal may not strongly depend on the colors of the fabrics [35], we then convert the RGB images to the gray-scale images to reduce the bias in model training. The second column of Fig. 3 demonstrates the examples of the gray-scale visual images for the selected fabric samples.

We utilize the position-based frictional coefficients as our tactile data, for the generative model training and the electro-vibration hardware control. The frictional tactile data consists of a series of measured frictional-coefficients data points related to the positional displacements on the 200×200 mm 2 physical fabric samples. Mapping to the visual image, each original frictional tactile signal of the selected fabric material includes 2000 data points with a finger-sliding distance of 155 mm approximately. To acquire the 2D representation of the frictional data, we compute the amplitude spectrogram from the wave-format frictional coefficients using Short-time Fourier transform (STFT), with a 512-hamming window and a 128-hop size.

4.2. Visual-to-frictional data set

To construct the visual-frictional data set for model training and testing, we adopt the data augmentation as follow:

For visual data, we first apply a sliding-window with the size of 1024×1024 pixels to move horizontally across and crop the original fabric image (2362×2362 pixels) in HapTex with the offset of two pixels between two consecutive windows. To this end, we totally acquire 669 RGB images for each kind of fabric material after the aforementioned augmentation. On the other hand, to map the pixel-wise visual images and the positional frictional-coefficients signals, we intercept the first 100 mm frictional signals including 1280 data points, which is corresponding to the size of 1024×1024 fabric image. Then we apply the sliding window with the size of 1280 data points to horizontally move from the beginning to the end of a frictional-coefficients signal with the offset of 1/12 mm between two consecutive windows based on the previous sliding-window settings of visual images, to build the visual-to-frictional cross-modal data pairs as our augmented data set. To this end, we augment both the visual images (1024×1024) and the amplitude spectrograms (257×11), for each category of fabric materials, resulting in 10 types of materials $\times 669 \times 2 = 13380$ visual and tactile data in total. For the visual images and the spectrograms data of each fabric material, we randomly split the data-pairs with the ratio of $8 : 1 : 1$ (training : validation : testing) as our data set.

5. Technical experiment: Frictional signals generation

5.1. Evaluation metrics

To evaluate the performance of our generative model, we compare the generated results with the ground-truth data. We first compare the amplitude spectrograms between the original data and the generated data. Specifically, we qualitatively compare the generated spectrogram from our proposed model and the real spectrogram on the visual appearance. Inspired by the previous work on GAN-based time-series data generation [57] and haptic texture generation [35,58], we then utilize t-Distributed Stochastic Neighbor Embedding (t-SNE) [25] to visualize the high-dimensional latent vectors in a 2D space as another qualitative evaluation metric. As our quantitative evaluation metric, we also examine the root-mean-squared errors (RMSE) and the mean absolute errors (MAE) on the generated 2D spectrogram data and 1D displacement-based waveform frictional-coefficients data.

5.2. Experimental settings and training details

We train our image-to-friction generation model using the data set described in Section 4.2. In our experiments, we implement our network with the Tensorflow framework 2.1.0 and train it on an Nvidia Geforce GTX 2080Ti GPU. The frictional signal processing and the Griffin-Lim algorithm [52] are implemented using the librosa [59] library, and the t-SNE results and the plots are generated with Scikit-learn [60] library in Python. We follow the WGAN-based training settings used by Arjovsky et al. [55], and use the RMSprop optimizer with a batch size of 8. The learning rates for the generator and the discriminator are $5e-5$ in our experiments. All model weights are initialized with the Xavier normal initializer [61]. According to our preliminary experiments, we set the number of training iterations as 200 epochs. With the aforementioned hardware and software configuration, it took about one day for training the GAN-based model, and each image-to-signal generation took approximately 0.02 s through our trained model. The final loss values of the Generator and the Discriminator were 0.8354 and 0.9456, respectively.

5.3. Experimental results and analysis

Fig. 3 illustrates the visual comparison of the spectrograms (generated vs real) and the waveform-based (the horizontal X-axis represents the displacement) frictional signals (generated vs real). Specifically, the third and the fourth columns show the generated and the real spectrograms, respectively. The comparison of these two columns demonstrates the visual similarity between the generated and the real spectrograms qualitatively. We also compare the generated and the ground-truth frictional coefficients in the waveforms. As shown in the last column of **Fig. 3**, the orange curves represent the generated signals, and the blue curves illustrate the real signals. The comparison on the waveform of frictional signals shows a similar trend and level of fluctuation between the generated and the ground-truth signals.

Furthermore, we compute the t-SNE visualization of the high-dimensional latent vectors (128-D) output by the encoder to compare the distribution of the training and the testing frictional data in a 2D space. This approach of dimension reduction has been applied to evaluate the generation of high-dimensional data (i.e., images [62] or time-series data [63]). In addition, some previous works [35,58] related to haptic texture generation also adopted t-SNE visualization for their evaluation of generative models. **Fig. 4** shows that the texture encoder has learned to cluster the same type of fabric material together, as the t-SNE results of the latent vectors for both the testing data (the “star” marks) and the training data (the “dot” marks) largely overlap with each other in the 2D space. The results also reveal that the samples of different fabric materials could be grouped into visually distinguishable clusters in a 2D space. These observations indicate that the generative model could learn the encoded features of the tactile signals for the same material category [64], and generate the outputs close to the real samples [57].

For the quantitative evaluation, **Tables 1** and **2** show the average RMSE and MAE of the generated spectrograms and frictional-coefficient signals for each fabric material type, respectively. The overall average RMSE value of spectrograms for all the generated samples across the 10 selected materials is 0.257 (SD = 0.064), and 0.028 (SD = 0.004) for the generated displacement-based frictional-coefficients signals, achieving the ratio of RMSE and the mean value of 5.22%–10.21%, averagely 6.94% for frictional-coefficient-signals. Similarly, the overall average MAE value of spectrograms is 0.069 (SD = 0.014), and 0.023 (SD = 0.003) for the waveform frictional-coefficients signals, which reaches 4.14%–8.42% (averagely 5.13%) on the ratio of MAE and average frictional-coefficient signals.

5.4. Generalization to new/unseen data

To study the generalizability of our model on generating the frictional signals for some new/unseen data, we randomly select three types of new data (denotes as Fiber-N, Wool-N and Silk-N) which are not included in the training data set, and pre-process their visual data and frictional data with the same aforementioned data-preparation procedure in Section 4. **Fig. 5** shows the visual textured images in gray-scale and the comparison of the waveform-based frictional-coefficients data (generated vs ground-truth). The trends and the values of the generated signals (the orange curves) are visually close to the ground-truth signals (the blue curves). The RMSE values of generated spectrograms and displacement-based frictional-coefficient-signals are 0.317 and 0.033 (Fiber-N), 0.402 and 0.034 (Wool-N) and 0.454 and 0.055 (Silk-N), separately, averagely 0.391 and 0.041, for these three types of new/unseen data. Additionally, the t-SNE visualization (**Fig. 4**) also illustrates that the generated frictional signals of selected three types of new data (Fiber-N, Wool-N and Silk-N) in the 2D space are closed to the clusters of the materials

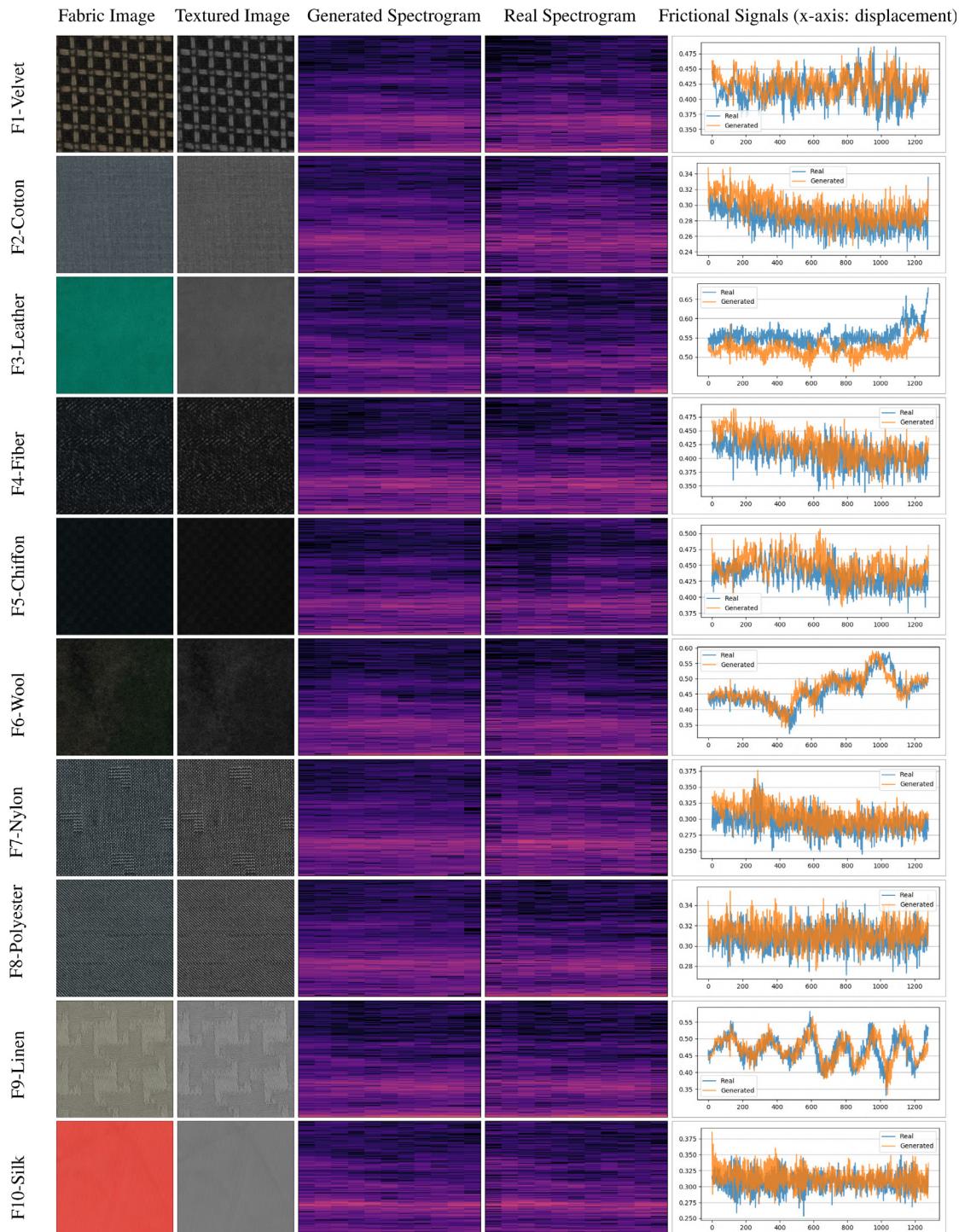


Fig. 3. The RGB fabric images (first row) and the gray-scale textured images (second row), and the comparison of generated/real spectrograms (third & fourth rows) and displacement-based waveforms (the horizontal X-axis represents the displacement) of the frictional signals (The orange line represents the generated and the blue line shows the real/ground-truth signals in the last row). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that are included in the training set, suggesting the model could generate reasonable tactile signals of the new data based on the learned knowledge. For example, the generator could generate the data of the unseen silk fabric (red “cross” marks) that is close to the cluster of silk (red “circle” marks) in the 2D space of t-SNE visualization, as they may share similar tactile features on texture surface.

6. User study: Tactile simulation on electrovibration tactile display

After experimenting with the technical performance of our model, we investigate the user perception on fabric simulation using the generated frictional-coefficients data. Specifically, we study how users would distinguish the generated frictional signals and the pre-recorded/ground-truth signals being rendered on the electrovibration tactile display.

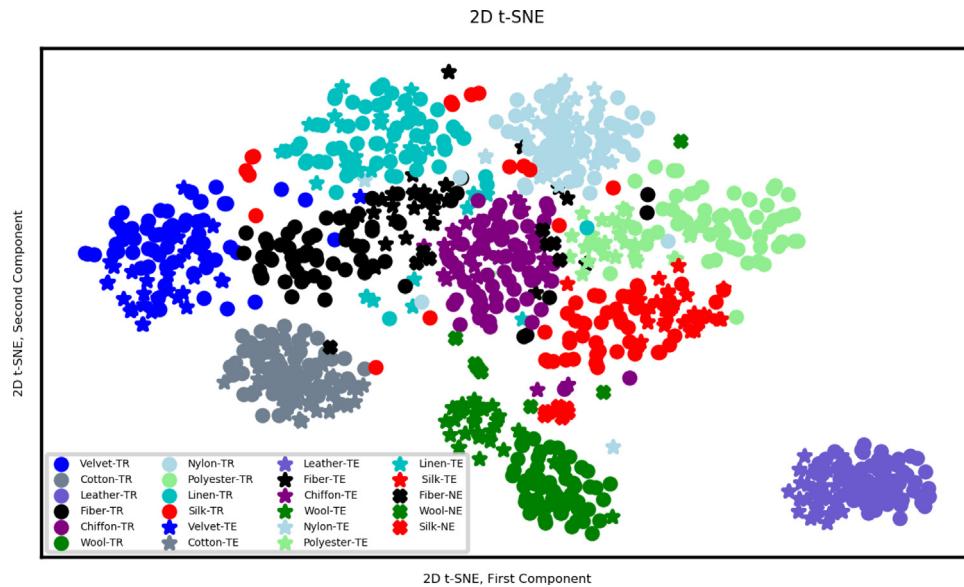


Fig. 4. The 2D t-SNE visualization of the tactile data. The “dot” mark denotes the training data (with the suffix “-TR”) and the “star” mark denotes the testing data (with the suffix “-TE”). The “cross” mark denotes the generated data from the new/unseen materials not in the training set (with the suffix “-NE”). The same color indicates the same category of fabric materials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The RMSE values of the generated spectrograms and the reconstructed frictional-coefficients signals for different materials.

Fabric materials	RMSE (SD)		Average signals	RMSE/Average signals
	Spectrograms	Signals		
F1-Velvet	0.0772(0.0075)	0.0232(0.0021)	0.4136	5.59%
F2-Cotton	0.2437(0.0123)	0.0205(0.0008)	0.2817	7.05%
F3-Leather	0.9104(0.0441)	0.0515(0.0025)	0.5581	10.21%
F4-Fiber	0.2629(0.0944)	0.0278(0.0024)	0.4052	6.66%
F5-Chiffon	0.2181(0.1108)	0.0301(0.0104)	0.4308	6.61%
F6-Wool	0.2430(0.1076)	0.0309(0.0065)	0.4684	6.73%
F7-Nylon	0.1215(0.0305)	0.0175(0.0022)	0.2922	5.86%
F8-Polyester	0.1015(0.0291)	0.0160(0.0011)	0.3069	5.22%
F9-Linen	0.2692(0.1240)	0.0433(0.0094)	0.4583	9.23%
F10-Silk	0.1252(0.0757)	0.0194(0.0024)	0.3068	6.27%
Average	0.2573(0.0636)	0.0280(0.0040)	0.3922	6.94%

Table 2
The MAE values of the generated spectrograms and the reconstructed frictional-coefficients signals for different materials.

Fabric materials	MAE (SD)		Average signals	MAE/Average signals
	Spectrograms	Signals		
F1-Velvet	0.0334(0.0021)	0.0180(0.0016)	0.4136	4.37%
F2-Cotton	0.0602(0.0046)	0.0166(0.0007)	0.2817	5.91%
F3-Leather	0.1274(0.0055)	0.0470(0.0019)	0.5581	8.42%
F4-Fiber	0.0778(0.0167)	0.0221(0.0020)	0.4052	5.48%
F5-Chiffon	0.0962(0.0318)	0.0246(0.0091)	0.4308	5.72%
F6-Wool	0.0724(0.0208)	0.0235(0.0045)	0.4684	5.14%
F7-Nylon	0.0434(0.0062)	0.0137(0.0018)	0.2922	4.70%
F8-Polyester	0.0511(0.0130)	0.0127(0.0009)	0.3069	4.14%
F9-Linen	0.0779(0.0267)	0.0343(0.0077)	0.4583	7.48%
F10-Silk	0.0462(0.0157)	0.0154(0.0020)	0.3068	5.03%
Average	0.0686(0.0143)	0.0228(0.0032)	0.3922	5.13%

6.1. Participants

Twelve participants (seven males, five females) are recruited from a local university, averagely aging 27.17 years old ($SD = 2.62$). Two of them have previous experience with electrovibration tactile displays. One participant is left-handed, and the others are right-handed. A small gratuity worth around 5 US Dollars is provided for the completion of the study.

6.2. Apparatus

As we use the HapTex database in our implementation, we adopt the same setup and procedure in Jiao et al.’s work [19] for our electrovibration-based fabric texture rendering through the pre-recorded/generated frictional-coefficient signals to ensure the rendering effectiveness. Specifically, we implement a customized electrovibration tactile display (Fig. 6(a)) following the same actuation principle of the TeslaTouch device [1]. To

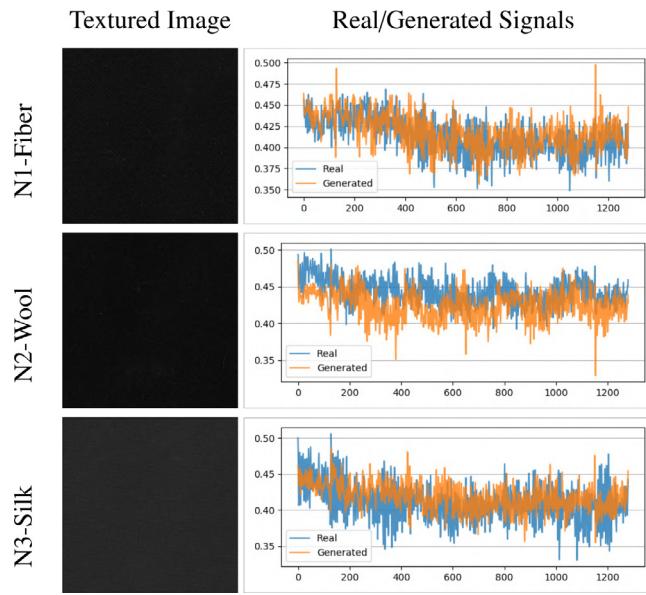


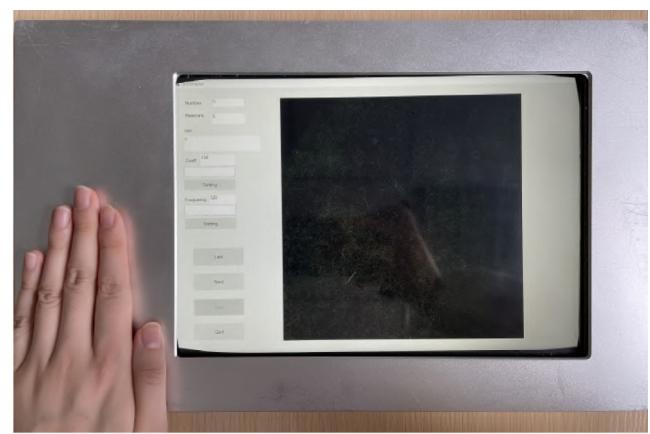
Fig. 5. The generated results of the unseen/new data (Fiber, Wool and Silk). The first column represents the textured image and the second column shows the generated (orange) and the real/ground-truth (blue) waveform-based frictional coefficients data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

facilitate the data recording, we develop the interactive interface shown in Fig. 6(b). The tactile display device consists of an infrared (IR) touch frame, a 3M Microtouch screen, a Microsoft Surface Pro3 tablet, and the controlling circuit. The haptic signal stimulus is activated and regulated in the controlling circuitry unit, which receives the commands from the tablet. The interactive and display area of the electrovibration device is 10.8 inches diagonally, which is determined by the size of the tablet. The resolution of the tablet screen is 1920×1280 pixels, and we scale the fabric image into 1280×1280 pixels to match the size of 100×100 mm² texture surface. Hence, the frictional signals could be controlled with the resolution of one pixel based on the user's finger position on the electrovibration surface. When a user moves his/her finger on the 3M Microtouch screen, the finger's position is tracked by the IR touch frame, and the electrovibration haptic feedback could be presented at the corresponding position on the screen. An Aluminum casing is installed around the device and connects with the ground pin of the controlling circuit to ensure the grounding status of the user's body for safety.

6.3. Stimuli

The electrovibration tactile display can generate various haptic textures by modulating the frequency, the voltage, and the waveform of the input signal. In our study, the frequency range of the driving signal is from 10 to 1000 Hz, and the voltage range is 0 to 255 Vpp. The current is limited to 0.5 mA for safety. In terms of the carrier wave, existing research showed that humans are more sensitive to the tactile stimuli generated by the square wave than the sinusoidal wave for the frequencies lower than 60 Hz, and no significant difference between these two waveforms at frequencies greater than 60 Hz [56]. Thus, we select the square wave as the carrier wave for the input signal of the device. The frequency value and the amplification factor for each material simulation are determined by the psychophysical experiment which will be described in the latter part.

To reduce the subjects' workload and time during the experiment, we randomly select 5 types of fabrics from the 10 selected



(a) The prototype of our electrovibration tactile device.



(b) The user-perception and -response interfaces for the tactile simulation experiment.

Fig. 6. (a) The experimental apparatus and (b) the interactive interface.

fabric materials (F1-Velvet, F5-Chiffon, F6-Wool, F9-Linen and F10-Silk in our current experiment). In addition, we include the 3 unseen/unknown fabric materials, namely N1-Fiber, N2-Wool, and N3-Silk, to further investigate the user perception of the frictional signal generated from the new data. To this end, we simulate 8 types of fabric materials by rendering their real and generated frictional signals on the electrovibration tactile display. The physical fabric materials are shown in Fig. 7, and used as the tactile reference in this study.

6.4. Study design

We conduct a within-subject study to evaluate the performance of tactile rendering on fabric materials using the generated frictional signals. During the study, the participants could move their finger horizontally across the electrovibration tactile surface to perceive the virtual texture. To create a realistic texture simulation, we design a two-phase user-perception experiment: parameter determination and virtual texture comparison. The first phase is designed for determining the amplification factor and the frequency of the input signal for the high-fidelity tactile rendering, and the second is for comparing the texture-rendering quality of the generated and the pre-recorded/ground-truth signals. The whole experiment for each participant takes approximately 50 min in total, with the two phases as below:

Phase 1: Parameter Determination (15–20 min). In this phase, the ground-truth frictional signal is rendered on the electrovibration tactile display for the baseline of tactile simulation determination. For each fabric material type, the participant is

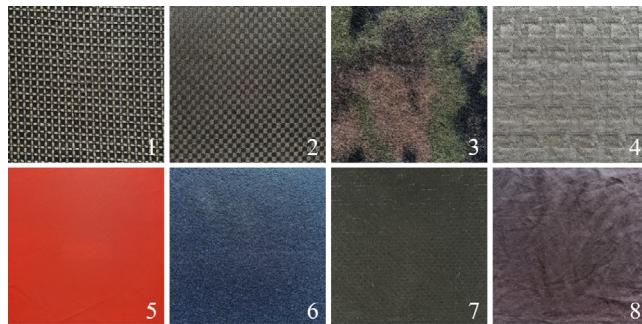


Fig. 7. The physical fabric materials to be rendered on the electrovibration tactile display. 1: F1-Velvet, 2: F5-Chiffon, 3: F6-Wool, 4: F9-Linen, 5: F10-Silk, 6: N1-Fiber, 7: N2-Wool, 8: N3-Silk.

instructed to adjust the amplification factor and carrier wave frequency using the interactive interface. In our empirical experiments for amplification factor and frequency adjustments, we found it is not trivial to achieve considerable rendering performance by adjusting only one factor (i.e. either amplification factor or frequency), so we decide to allow users to adjust these two parameters concurrently. The participant can stop at any time when they feel the rendered tactile sensation well matches with the physical material sample on the side of the tactile display. The recorded results of the amplification factors and the frequencies are used for both ground-truth and generated signals rendering in the next phase for each participant individually.

Phase 2: Virtual Texture Comparison (20–30 min). In this phase, the participant is asked to evaluate the authenticity of the frictional signals rendered on the tactile display. Specifically, taking the real physical fabric sample as the reference, the participant needs to select the virtual tactile texture that is more likely to be rendered by the ground-truth signal within a group of two provided frictional signals (generated vs real/ground-truth). Both the generated and the ground-truth signals are scaled by the same amplification factor and modulated with the same frequencies acquired from Phase 1 for each participant for each type of fabric material.

The HapTex database [22] records the friction-related parameters with the normal pressure force between 0.8 and 1.2 N applied on the surface, falling in the range between 0.2 and 1.5 N, which could be recognized as the “light” pressure for surface texture perception [65]. Hence, we require the users to gently move their fingertips for texture perception during our experiments both in the physical and virtual texture surfaces. Since the frictional-coefficient signals in HapTex are collected mainly based on the displacements, we do not strictly control the sliding speed of the fingertip during the perception experiments. Additionally, we aim to measure the users’ texture perception in a natural process without controlling the pressing force and the sliding speed.

6.5. Task and procedure

Each experimental session involves one experimenter and one participant. Upon arrival, the participant is seated in a comfortable position. The session starts with introducing the experiment process, and the participant is asked to finish a pre-questionnaire about his/her demographic information. Before the actual perception experiment, the participant needs to wash their hands with soap and dry with a paper towel to ensure the normal tactile perception on his/her bare fingertip, and the touchscreen of the electrovibration device is also cleaned with alcohol by the experimenter. During all the experiment sessions, the room temperature was 23 °C–26 °C, and the humidity was 15%–35%.



Fig. 8. The experimental setup. The participant sits at one side of the table and explores the virtual textures; meanwhile, the experimenter sits at the other side of the table to transpose the real material samples and records the participant’s response.

The actual experimental session starts with Phase 1 for parameter determination. Specifically, the participant is instructed to adjust the amplification factor and the frequency for the pre-recorded/ground-truth frictional signals, to match the real physical materials placed aside. The experimenter presets the initial values of these two parameters according to the empirical experiments. When the tactile perception of the real material is stronger than the virtual material (i.e., the real material induces a larger frictional force), the participant needs to increase the amplification factor in the software interface (Fig. 6), as well as to increase the frequency when the real material is perceived smoother with smaller friction than the virtual material. Finally, the parameters set by each participant are saved locally and used to render both the generated and the ground-truth friction signals in Phase 2. While being satisfied with the virtual material, the participant is asked to remember the tactile sensation of the virtual material as much as possible.

There is a voluntary 5-minute break period after Phase 1, and Phase 2 begins right after the break. In Phase 2, each material is rendered using both the real/ground-truth and the generated signals on the electrovibration tactile display. The participant perceives the virtual textures and evaluates the similarity with the reference of the real physical fabric samples. To this end, the participant needs to verbally report whether the current virtual texture is rendered using the real/ground-truth signals. Thus, each participant performs a total of 2 types of signal stimuli (generated or real) \times 8 types of materials \times 2 repetitions = 32 trials. The order of fabric types is displayed randomly, and the order of the frictional signals is counterbalanced within each type of fabric material. During the experiment, the participant wears an eye-patch when he/she perceives the textures and completes the tasks. The verbal responses are recorded by the experimenter (Fig. 8).

6.6. Results

Table 3 shows the amplification factor and the frequency values recorded in Phase 1 for each fabric material, respectively. Note that the amplification factor and the carrier signal frequency vary broadly across different types of fabric materials. The amplification factor could scale the amplitude of the driven voltage. For the electrovibration perception, a texture stimulus being rendered with a high frequency could be perceived as being smooth [1]. For instance, the material#5 (F10-Silk) generally feels smoother than the others, so it might yield the highest frequency value (293.3 Hz). On the other hand, existing research shows that

Table 3

The mean and the SD of the amplification factor and the frequency recorded in the Phase 1 of the user study.

Fabric material	Amplification factor (SD)	Frequency (SD) [Hz]
1	216.25 (56.04)	26.25 (10.69)
2	98.33 (15.33)	45.42 (8.65)
3	133.33 (9.24)	93.58 (42.01)
4	210.00 (35.24)	120.83 (75.61)
5	150.83 (59.21)	293.33 (40.53)
6	110.42 (21.22)	57.08 (9.16)
7	146.25 (48.55)	85.83 (28.75)
8	131.67 (35.29)	173.33 (24.25)

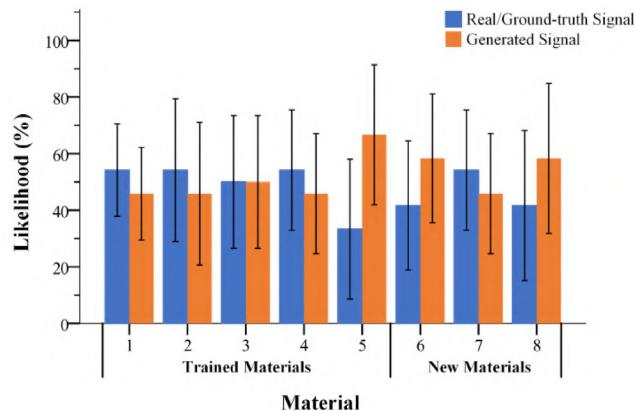


Fig. 9. Mean values and 95% confidence intervals of the likelihood of being perceived as real for the ground-truth and the generated signal.

the electrovibrational stimulus with the carrier-wave frequency above 240 Hz tends to require a higher amplitude to be perceivable by human subjects [1,56], so the material#5 (F10-Silk) obtained a large amplification factor (150.83), which is the third-largest, for the clear tactile perception of the electrovibrational stimulus. We also notice that the physical sample of material#4 (F9-Linen) involves a set of regular patterns with different weaving techniques (see Fig. 7-4), resulting in the bumpy sensation when the fingertip is sliding across the boundary between different patterns. Such bumpy sensation needs a large amplification factor value (210.00) for electrovibration rendering. Moreover, these weaving patterns result in different levels of roughness and slipperiness among different areas of the fabric sample. This can be reflected by the obtained carrier-wave frequency with high variance ($SD = 75.61$) across different participants, as we observe that different participants tend to touch and explore different regions of the physical sample for material#4 (F9-Linen) and adjust the electrovibration stimulus according to the areas they touch. Therefore, we can see that the tactile perception of the fabric texture could be largely dependent on the person who is touching the texture, and the location where the person is touching. In our experiment, we electrovibrationally render the frictional signal for each participant based on the parameters (i.e. the carrier-wave frequency and the amplification factor) that he/she adjust to match with the physical sample in Phase 1, to ensure the realness of the tactile signal rendering for each participant.

Fig. 9 shows the percentage of the real/ground-truth and the generated signals being identified as the actual signal recorded from the real physical material (i.e. being more closed to the real physical sample) for each type of fabric material. In the following discussion, we define the percentage of one particular type of signal being identified to be the actual pre-recorded/ground-truth signal as its likelihood of being perceived as real. The average likelihood of the ground-truth signal being perceived as real is

47.92% while it is 52.08% for the generated signal on average. We perform a two-way repeated-measures ANOVA on the likelihood of being perceived as real considering the factors of stimuli and material type for the trained materials and the unseen/new types of fabric materials, separately. The statistical results demonstrate that the type of stimuli (generated signal vs real signal) has a minor effect on the likelihood of being perceived as real for the trained fabric materials ($F(1, 11) = 0.044, p = 0.838, \eta_p^2 = 0.003$; the likelihood for the ground-truth signal: 49.17% vs the likelihood for the generated signal: 50.83%) and the unseen/unknown materials ($F(1, 11) = 0.508, p = 0.491, \eta_p^2 = 0.004$; the likelihood for the ground-truth signal: 45.83% vs the likelihood for the generated signal: 54.17%). This suggests that virtual textures created by the generated frictional signals are perceived to be similar to the textures created by the pre-recorded signals, with the generated signals being slightly more likely to be perceived as real. Furthermore, the two-way repeated-measures ANOVA on the overall data show no interactive effects of these two factors (stimuli \times material type: $F(7, 77) = 0.584, p = 0.767, \eta_p^2 = 0.050$). This further indicates that the material type does not significantly affect the likelihood for either the ground-truth or the generated signals being perceived as real.

We also notice three types of materials (F10-Silk, N1-Fiber and N3-Silk) rendering by generated signals yielding slightly but not significantly higher performance than pre-recorded signals in the discrimination experiment. During the experiments, some participants commented that it was difficult for them to distinguish the generated and the pre-recorded signals for F10-Silk and N3-Silk. As both of these materials are silk which yields relatively lower frictional-coefficient signals and could be considered to be smooth, the confusion in these materials could be due to the glass-made touch screen being smooth by nature. N1-Fiber also demonstrates that the participants tended to consider the generated signal slightly more likely to be the real signal. This could be due to the various prior experiences among the limited number of experiment subjects. We will involve a larger group of participants in future studies.

7. Discussion: Limitations and future work

The technical experiments show that our model can generate frictional signals that are visually and statistically similar to the ground-truth signal. The user studies also suggest that the participants could not distinguish the real/ground-truth signals and the generated signals in tactile simulation and rendering. We also explore the capability of generalizing our trained model to three types of unseen/new data. The signal visualization and the user-perception experiment show that our model can generate the frictional signals that are close to the ground truth for the three selected types of unseen materials. On the other hand, we observe a few limitations in our approach.

7.1. Biased signal generation

We notice the generation results of F3-Leather demonstrate the minimal overlapping with its ground-truth data in the waveform-based frictional-coefficient signal plot (i.e., the third row of Fig. 3), the largest RMSE (0.0515), and the largest ratio of RMSE/mean (10.21%). This might be due to the fact that F3-Leather yields a considerable difference from the other materials, causing the inter-class imbalance in the model training [66]. The problem of data imbalance may further bias the predictive/generative capability of the deep neural network, leading to the worse performance on the minority class than the other classes [67]. We also perform the Dynamic Time Warping (DTW) algorithm upon the ground-truth frictional-coefficient data of

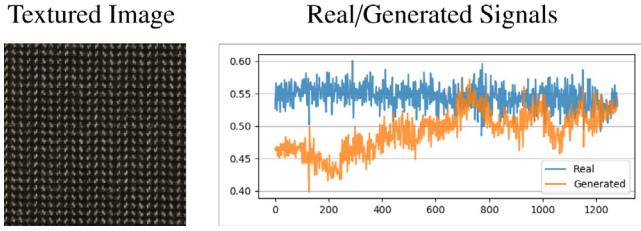


Fig. 10. The less successful signal generation for new data in the Linen category. The input textured image shows the regular weaving style with the smaller variance of frictional coefficients signals (the blue curve), but the generated result (the orange curve) shows the bumpy change on tactile properties with the similar attribute of linen fabrics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10 selected materials. The average warping distance between F3-Leather and the other materials (5.96, SD = 3.16) is larger than the average distance within the other non-F3-Leather materials (2.99, SD = 1.93), indicating a large difference between F3-Leather and the others. The t-SNE visualization (Fig. 4) also shows that the cluster of F3-Leather (i.e. the slate blue marks) is far from the other clusters. This might bias the generative model to treat F3-Leather as the minority class and negatively affect the signal generation performance for F3-Leather. On the other hand, the generated signals for F3-Leather still show a similar tendency with the ground-truth signals. This suggests that our model tends to learn the displacement-based changing trend of the frictional-coefficient data which plays an essential role in the tactile simulation of the fabric texture [12]. In addition, F9-Linen also shows the larger RMSE and MAE values (0.043 and 0.034) and the ratio of RMSE/MAE and average signals (9.23%/7.48%), we suppose it might be because of the higher variance of the frictional-coefficient signals of F9-Linen, which could cause the mismatch of the phase between the generated and ground-truth data, possibly increasing the RMSE and the MAE values.

7.2. Generalization

We observe some less successful cases of frictional signal generation for unseen materials. Fig. 10 illustrates one unsuccessful example of generating the frictional signal for the Linen category. The generated frictional signal shows a largely different trend and level compared to the ground-truth signal. One possible reason is that the textile pattern might affect the haptic properties within the same category of fabrics. Fig. 3 shows a large amplitude change on the frictional signal of F9-Linen, which might be due to the physical patterns woven on the textiles. Compared to F9-Linen, the linen material in Fig. 10 shows a more gentle variance with possibly fewer physical bumps during the weaving process. Therefore, we will improve the generative model with additional supervisions, such as the textile style, to further enhance its robustness and generalization.

We also observe that the glass surface of the electrovibration tactile display renders the smooth materials (e.g., silk) better than the rough/bumpy materials (e.g., velvet or linen). This is because the bumpy geometric patterns on certain fabric materials offer additional types of tactile sensation, such as pressure on the finger surface. Therefore, rendering realistic texture based on concurrent tactile feedback would be an exciting and challenging problem. Also, as we adopt the GAN structure, the latent space could also be interpolated to acquire new haptic textures to modify material roughness sensations for virtual texture design [68], which would be another potential future work.

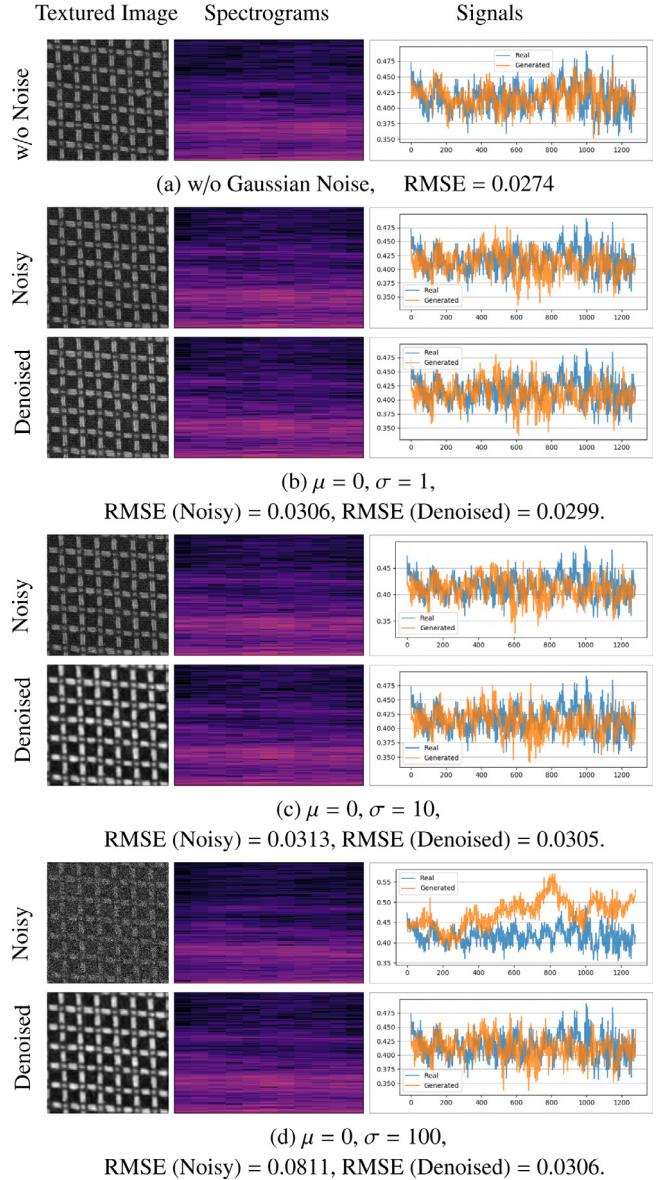


Fig. 11. The generated spectrograms and waveform-based frictional-coefficient signals of noisy or denoised fabric textured image with different noise variances. The first column represents the input fabric image and the second and the third columns show the generated spectrogram and waveform-based frictional coefficients data.

7.3. Limited dataset

The original HapTex database [22] contains both visual and tactile data among 120 types of fabric materials but we only selected 10 types for our model training. This is because some visual images were contaminated (e.g., some patterns marked by pen), and some recorded frictional signals were not valid (e.g., too short measurement distance). Such limitation in the dataset may lead to the vulnerability of our model when the input image is affected by noise. To this end, we test our model upon the images under different noise levels by adding different levels of Gaussian noise (i.e., the same mean value $\mu = 0$ with different variances $\sigma = 1/10/100$) into the visual images for Material#1 (F1-Velvet) as the inputs for our generative models. The results of the noisy image input in Fig. 11 show that our model can generate considerable quality of the tactile signal to a certain extend. We also experiment with our model by denoising the noisy images

using Gaussian filters, and the results of the denoised input in Fig. 11 shows that the common denoising approach (i.e. Gaussian filtering) could improve the generative results with more similar trends as the ground-truth data, and smaller RMSE values. To further enhance the robustness of the presented generative model, we will construct a more comprehensive database with high-quality visual and tactile data. In addition, the current database is limited by collecting the data only when users sliding their fingers across the fabric surface from left to right. In order to increase the diversity of frictional data, we intend to collect the frictional-coefficient signals upon different tools, sliding directions, and positions.

7.4. Random noise in the latent space

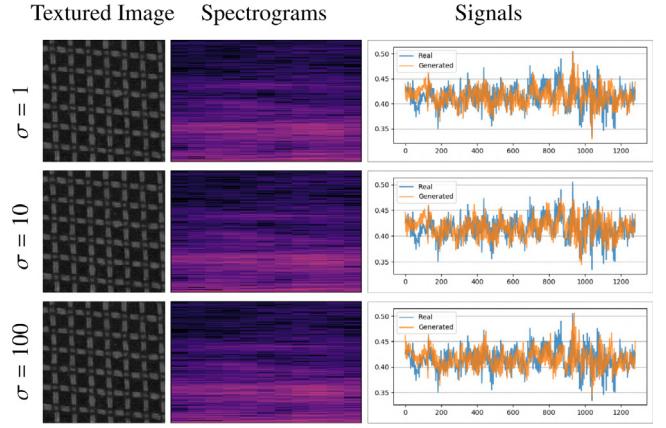
We add a randomized 50-dimensional noise vector in the latent space when training our GAN model, which could introduce specific variances in the latent representation and reinforce the robustness of the decoder. This structure could also improve the performance of the generative model [69]. In our work, we also tested the influence of the randomized noise vector in our generative model. Specifically, we added different levels of random noise (i.e., the same mean value $\mu = 0$ with different variances $\sigma = 1/10/100$) to the latent space in our model for two different fabric materials (F1-Velvet) and F2-Cotton). Fig. 12 demonstrates that inducing different levels of random noise into the latent space has no significant effect on the generated spectrograms and the values of RMSE.

7.5. Parameter determination

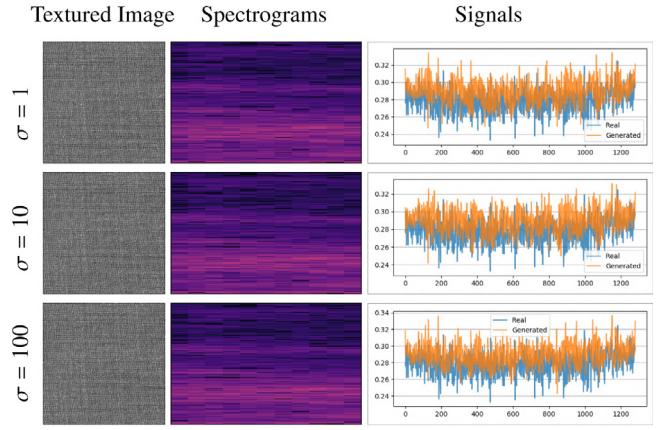
While our experiments showed that our generative models could generate the frictional-coefficient signals rendered on the electrovibrational surface to simulate the touch feeling of fabric textures, the process still requires the user to manually adjust the rendering parameters (e.g., the amplification factor or carrier wave frequency). However, it might be tedious to acquire these necessary parameters for different textures and users. During the experiments, we observe that the tactile perception of electro-vibrationally rendered fabric texture partially depends on the finger-contact location and the individual haptic sensitivity of the participant. Therefore, it would be optimal to dynamically control the amplification factor and the carrier-wave frequency to match the texture of physical fabric samples. In the future, we will investigate the design of the regression model for personal prediction of the rendering parameters for different locations and types of materials.

7.6. Gender, prior experiences, and handedness

Our user study aims to explore users' tactile perception of fabric texture simulation based on our generated and pre-recorded frictional-coefficient signals. We involved female and male participants with different prior experiences and right-/left-handedness. The previous research [70] shows that the tactile roughness discrimination threshold was unrelated to age or gender. Our user study involves 5 females and 7 males, and the average likelihoods of the ground-truth signal being perceived as real are 46.42% (males) and 50.00% (females), separately. One-way ANOVA also shows that there is no significant difference between the two genders ($F(1, 10) = 0.253, p = 0.626$). In addition, although two of the participants had prior experience on electro-vibration tactile surfaces, they reported that the usage was just 1 or 2 times and had no previous experience on texture simulation of the electrovibration tactile display. These prior experiences may not yield significant effects on the tactile perception in



(a) Velvet: All $\mu = 0$, RMSE ($\sigma = 1$) = 0.0230,
RMSE ($\sigma = 10$) = 0.0234, RMSE ($\sigma = 100$) = 0.0239.



(b) Cotton: All $\mu = 0$, RMSE ($\sigma = 1$) = 0.0193,
RMSE ($\sigma = 10$) = 0.0201, RMSE ($\sigma = 100$) = 0.0202.

Fig. 12. The generated spectrograms and waveform-based frictional-coefficient signals when adding with different noise variances. (a)-Velvet and (b)-Cotton. The first column represents the input textured image and the second and the third columns show the generated spectrogram and waveform-based frictional coefficients data.

our experiment. Lastly, Ardila et al. [71] also show that the right/left-handedness might not place a significant effect on the humans' roughness estimation. Therefore, we can assume that gender, prior experiences and right/left-handedness could not significantly affect our experimental results. On the other hand, involving a large number of participants further increases the reliability of the experimental results.

8. Conclusion

In this paper, we present a GAN-based framework to generate the frictional signals from the fabric texture images, for the tactile simulation on the electrovibration tactile display. The frictional-signal-generation model takes the visual fabric images as the input and generates the waveform of the displacement-based frictional coefficients of the fabric surface. The experimental results show that our model could generate the amplitude spectrograms and the waveforms of the frictional signals similar to the ground-truth data pre-recorded from the corresponding real physical materials. We also conduct user studies to validate the usage of the generated frictional signals on tactile simulation of fabric material. The results demonstrate that the users could not discriminate the generated and the real frictional signals being

rendered on the electrovibration tactile display, suggesting the effectiveness of our model for tactile material simulation.

CRediT authorship contribution statement

Shaoyu Cai: Conceptualization, Methodology, Software, Experiments, Analysis, Writing – original draft. **Lu Zhao:** Hardware, User study, Validation. **Yuki Ban:** Investigation, Writing – review & editing. **Takuji Narumi:** Investigation, Writing – review & editing. **Yue Liu:** Supervision. **Kening Zhu:** Conceptualization, Investigation, Supervision, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Prof. Dangxiao Wang from BUAA for providing the physical fabric materials in our experiment. This research was partially supported by the Young Scientists Scheme of the National Natural Science Foundation of China (Project No. 61907037), the Applied Research Grant (Project No. 9667189), and the Centre for Applied Computing and Interactive Media (ACIM) of School of Creative Media, City University of Hong Kong. This work was also partially supported by the National Key Research and Development Program of China (Project No. 2019B010149001), the National Natural Science Foundation of China (Project No. 61631010, No. 61960206007, No. 62172346), the Guangdong Basic and Applied Basic Research Foundation (Project No. 2021A1515011893), the 111 Project, China (B18005), and JSPS KAKENHI Grants (Project No. 20K21801 and No. 21H03478).

References

- [1] Bau O, Poupyrev I, Israr A, Harrison C. TeslaTouch: Electrovibration for touch surfaces. In: Proceedings of the 23rd annual ACM symposium on user interface software and technology. 2010. p. 283–92.
- [2] Galambos Péter. Vibrotactile feedback for haptics and telemanipulation: survey, concept and experiment. *Acta Polytechnica Hungarica* 2012;9(1):41–65.
- [3] Bochereau S, Sinclair S, Hayward V. Perceptual constancy in the reproduction of virtual tactile textures with surface displays. *ACM Trans Appl Percept* 2018;15(2):1–12.
- [4] Zhu K, Perrault S, Chen T, Cai S, Peiris RL. A sense of ice and fire: Exploring thermal feedback with multiple thermoelectric-cooling elements on a smart ring. *Int J Hum-Comput Stud* 2019;130:234–47.
- [5] Nasser A, Keng K-N, Zhu K. ThermalCane: Exploring thermotactile directional cues on cane-grip for non-visual navigation. In: Proceedings of the 22nd international ACM SIGACCESS conference on computers and accessibility. 2020. p. 1–12.
- [6] Chen T, Wu Y-S, Zhu K. Investigating different modalities of directional cues for multi-task visual-searching scenario in virtual reality. In: Proceedings of the 24th ACM symposium on virtual reality software and technology. 2018. p. 1–5.
- [7] Zhu K, Chen T, Han F, Wu Y-S. HapTwist: Creating interactive haptic proxies in virtual reality using low-cost twistable artefacts. In: Proceedings of the 2019 CHI conference on human factors in computing systems. 2019. p. 1–13.
- [8] Cai S, Ke P, Jiang S, Narumi T, Zhu K. Demonstration of thermairglove: A pneumatic glove for material perception in virtual reality through thermal and force feedback. In: SIGGRAPH Asia 2019 emerging technologies. 2019. p. 11–2.
- [9] Cai S, Ke P, Narumi T, Zhu K. ThermAirGlove: A pneumatic glove for thermal perception and material identification in virtual reality. In: 2020 IEEE conference on virtual reality and 3D user interfaces. 2020.
- [10] Shin S, Osgouei RH, Kim K-D, Choi S. Data-driven modeling of isotropic haptic textures using frequency-decomposed neural networks. In: 2015 IEEE world haptics conference. IEEE; 2015. p. 131–8.
- [11] Israr A, Poupyrev I. Tactile brush: Drawing on skin with a tactile grid display. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2011. p. 2019–28.
- [12] Bueno M-A, Lemaire-Semail B, Amberg M, Giraud F. A simulation from a tactile device to render the touch of textile fabrics: A preliminary study on velvet. *Text Res J* 2014;84(13):1428–40.
- [13] Zhao L, Liu Y, Ye D, Ma Z, Song W. Implementation and evaluation of touch-based interaction using electrovibration haptic feedback in virtual environments. In: 2020 IEEE conference on virtual reality and 3D user interfaces. 2020. p. 239–47.
- [14] Benko H, Holz C, Sinclair M, Ofek E. Normaltouch and texturetouch: High-fidelity 3D haptic shape rendering on handheld virtual reality controllers. In: Proceedings of the 29th annual symposium on user interface software and technology. 2016. p. 717–28.
- [15] Whitmire E, Benko H, Holz C, Ofek E, Sinclair M. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In: Proceedings of the 2018 CHI conference on human factors in computing systems. 2018. p. 1–12.
- [16] Choi I, Ofek E, Benko H, Sinclair M, Holz C. Claw: A multifunctional handheld haptic controller for grasping, touching, and triggering in virtual reality. In: Proceedings of the 2018 CHI conference on human factors in computing systems. 2018. p. 1–13.
- [17] Degraen D, Zennner A, Krüger A. Enhancing texture perception in virtual reality using 3D-printed hair structures. In: Proceedings of the 2019 CHI conference on human factors in computing systems. 2019. p. 1–12.
- [18] Sun Y, Yoshida S, Narumi T, Hirose M. Pacapa: A handheld VR device for rendering size, shape, and stiffness of virtual objects in tool-based interactions. In: Proceedings of the 2019 CHI conference on human factors in computing systems. 2019. p. 1–12.
- [19] Jiao J, Zhang Y, Wang D, Visell Y, Cao D, Guo X, et al. Data-driven rendering of fabric textures on electrostatic tactile displays. In: 2018 IEEE haptics symposium. IEEE; 2018. p. 169–74.
- [20] Ilkhani G, Aziziaghdam M, Samur E. Data-driven texture rendering on an electrostatic tactile display. *Int J Hum-Comput Interact* 2017;33(9):756–70.
- [21] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst* 2014;2672–80.
- [22] Jiao J, Zhang Y, Wang D, Guo X, Sun X. Haptex: A database of fabric textures for surface tactile display. In: 2019 IEEE world haptics conference. IEEE; 2019. p. 331–6.
- [23] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1125–34.
- [24] Cai S, Ban Y, Narumi T, Zhu K. FrictGAN: Frictional signal generation from fabric texture images using generative adversarial network. In: International conference on artificial reality and telexistence & eurographics symposium on virtual environments. The Eurographics Association; 2020. p. 11–5.
- [25] Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9(11).
- [26] Meyer DJ, Peshkin MA, Colgate JE. Fingertip friction modulation due to electrostatic attraction. In: 2013 world haptics conference. IEEE; 2013. p. 43–8.
- [27] Xu C, Israr A, Poupyrev I, Bau O, Harrison C. Tactile display for the visually impaired using TeslaTouch. In: Extended abstracts on human factors in computing systems. 2011. p. 317–22.
- [28] Kim S-C, Israr A, Poupyrev I. Tactile rendering of 3D features on touch surfaces. In: Proceedings of the 26th annual ACM symposium on user interface software and technology. 2013. p. 531–8.
- [29] Romano JM, Kuchenbecker KJ. Creating realistic virtual textures from contact acceleration data. *IEEE Trans Haptics* 2012;5(2):109–19.
- [30] Ilkhani G, Aziziaghdam M, Samur E. Data-driven texture rendering with electrostatic attraction. In: International conference on human haptic sensing and touch enabled computer applications. Springer; 2014. p. 496–504.
- [31] Osgouei RH, Shin S, Kim JR, Choi S. An inverse neural network model for data-driven texture rendering on electrovibration display. In: 2018 IEEE haptics symposium. IEEE; 2018. p. 270–7.
- [32] Osgouei RH, Kim JR, Choi S. Data-driven texture modeling and rendering on electrovibration display. *IEEE Trans Haptics* 2020;13(2):298–311.
- [33] Zhao L, Liu Y, Ma Z, Wang Y. Design and evaluation of a texture rendering method for electrostatic tactile display. In: Extended abstracts of the 2019 CHI conference on human factors in computing systems. 2019. p. 1–6.
- [34] Culbertson H, Delgado JJJ, Kuchenbecker KJ. One hundred data-driven haptic texture models and open-source methods for rendering on 3D objects. In: 2014 IEEE haptics symposium. IEEE; 2014. p. 319–25.
- [35] Ujijitoko Y, Ban Y. Vibrotactile signal generation from texture images or attributes using generative adversarial network. In: International conference on human haptic sensing and touch enabled computer applications. Springer; 2018. p. 25–36.

- [36] Liu H, Guo D, Zhang X, Zhu W, Fang B, Sun F. Toward image-to-tactile cross-modal perception for visually impaired people. *IEEE Trans Autom Sci Eng* 2020.
- [37] Cai S, Zhu K, Ban Y, Narumi T. Visual-tactile cross-modal data generation using residue-fusion GAN with feature-matching and perceptual losses. *IEEE Robot Autom Lett* 2021;6(4):7525–32.
- [38] Wang T, Sun X. Electrostatic tactile rendering of image based on shape from shading. In: 2014 international conference on audio, language and image processing. IEEE; 2014, p. 775–9.
- [39] Wu S, Sun X, Wang Q, Chen J. Tactile modeling and rendering image-textures based on electrovibration. *Vis Comput* 2017;33(5):637–46.
- [40] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2223–32.
- [41] Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. 2016, arXiv preprint [arXiv:1605.05396](https://arxiv.org/abs/1605.05396).
- [42] Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 5907–15.
- [43] Chen L, Srivastava S, Duan Z, Xu C. Deep cross-modal audio-visual generation. In: proceedings of the thematic workshops of ACM multimedia 2017. 2017. p. 349–57.
- [44] Hao W, Zhang Z, Guan H. Cmcgan: A uniform framework for cross-modal visual-audio mutual generation. In: Proceedings of the AAAI conference on artificial intelligence, 32, (1). 2018.
- [45] Li X, Liu H, Zhou J, Sun F. Learning cross-modal visual-tactile representation using ensembled generative adversarial networks. *Cogn Comput Syst* 2019;1(2):40–4.
- [46] Strese M, Schuwerk C, Iepure A, Steinbach E. Multimodal feature-based surface material classification. *IEEE Trans Haptics* 2016;10(2):226–39.
- [47] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4681–90.
- [48] Chen S-Y, Su W, Gao L, Xia S, Fu H. DeepFaceDrawing: Deep generation of face images from sketches. *ACM Trans. Graph.* 2020;39(4):72:1–72:16.
- [49] Aytar Y, Vondrick C, Torralba A. Soundnet: Learning sound representations from unlabeled video. 2016, arXiv preprint [arXiv:1610.09001](https://arxiv.org/abs/1610.09001).
- [50] Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics. 2016, arXiv preprint [arXiv:1609.02612](https://arxiv.org/abs/1609.02612).
- [51] Mirza M, Osindero S. Conditional generative adversarial nets. 2014, arXiv preprint [arXiv:1411.1784](https://arxiv.org/abs/1411.1784).
- [52] Griffin D, Lim J. Signal estimation from modified short-time Fourier transform. *IEEE Trans Acoust Speech Signal Process* 1984;32(2):236–43.
- [53] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015, p. 234–41.
- [54] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks. 2017, arXiv preprint [arXiv:1701.04862](https://arxiv.org/abs/1701.04862).
- [55] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. 2017, arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875).
- [56] Vardar Y, Güçlü B, Basdogan C. Effect of waveform on tactile perception by electrovibration displayed on touch screens. *IEEE Trans Haptics* 2017;10(4):488–99.
- [57] Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. *Adv Neural Inf Process Syst* 2019;32.
- [58] Heravi N, Yuan W, Okamura AM, Bohg J. Learning an action-conditional model for haptic texture generation. In: 2020 IEEE international conference on robotics and automation. IEEE; 2020, p. 11088–95.
- [59] McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E, et al. Librosa: Audio and music signal analysis in python. Citeseer; 2015.
- [60] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [61] Glorot X, Bengio Y. Understanding the difficulty of training deep feed-forward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings; 2010, p. 249–56.
- [62] Jin Y, Zhang J, Li M, Tian Y, Zhu H, Fang Z. Towards the automatic anime characters creation with generative adversarial networks. 2017, arXiv preprint [arXiv:1708.05509](https://arxiv.org/abs/1708.05509).
- [63] Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. *Adv Neural Inf Process Syst* 2019;32:5508–18.
- [64] Zhang D, Sun Y, Eriksson B, Balzano L. Deep unsupervised clustering using mixture of autoencoders. 2017, arXiv preprint [arXiv:1712.07788](https://arxiv.org/abs/1712.07788).
- [65] Zophoniasson H, Bolzmacher C, Anastassova M, Hafez M. Electrovibration: Influence of the applied force on tactile perception thresholds. In: 2017 zooming innovation in consumer electronics international conference. IEEE; 2017, p. 70–3.
- [66] Sampath V, Maurua I, Martín JJ A, Gutierrez A. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data* 2021;8(1):1–59.
- [67] Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw* 2018;106:249–59.
- [68] Asano S, Okamoto S, Matsuura Y, Yamada Y. Toward quality texture display: Vibrotactile stimuli to modify material roughness sensations. *Adv Robot* 2014;28(16):1079–89.
- [69] Kingma DP, Welling M. Auto-encoding variational Bayes. 2013, arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [70] Libouton X, Barbier O, Plaghki L, Thonnard J-L. Tactile roughness discrimination threshold is unrelated to tactile spatial acuity. *Behav Brain Res* 2010;208(2):473–8.
- [71] Ardila A, Uribe BE, Angel ME. Handedness and psychophysics: Weight and roughness. *Int J Neurosci* 1987;36(1–2):17–21.