End-to-End Semantically Aware Tactile Generation

Mohammad Mahdi Heydari Dastjerdi
Carleton University
Abbas Akkasi
abbas.akkasi@carleton.ca

Carleton University
Hilaire Djani
Carleton University
Aatreyi Pranavbhai Mehta
Carleton University
Majid Komeili
Carleton University

Research Article

Keywords: Tactile Generation, U-Net++, Perceptual Loss, Gradient Penalty

Posted Date: November 6th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5338871/v1

Additional Declarations: No competing interests reported.

End-to-End Semantically Aware Tactile Generation

Mohammad Mahdi Heydari Dastjerdi[1], Abbas Akkasi[1*], Hilaire Djani[1], Aatreyi Pranavbhai Mehta[1], Majid Komeili[1]

[1*]School of Computer Science, Carleton University, 1125 Colonel By Drive, , Ottawa, Ontario, K1S 5B6, Canada.

*Corresponding author(s). E-mail(s): abbas.akkasi@carleton.ca;
Contributing authors: mohammadheydari@cmail.carleton.ca;
hilairedjani@cmail.carleton.ca; aatreyipranavbhaime@cmail.carleton.ca;
majid.komeili@carleton.ca;

Abstract

Tactile graphics are an essential tool for conveying visual information to visually impaired individuals. However, translating 2D plots, such as B´ezier curves, polygons, and bar charts, into an effective tactile format remains a challenge. This paper presents a novel, two-stage deep learning pipeline for automating this conversion process. Our method leverages a Pix2Pix architecture, employing a U-Net++ generator network for robust image generation. To improve the perceptual quality of the tactile representations, we incorporate an adversarial perceptual loss function alongside a gradient penalty. The pipeline operates in a sequential manner: firstly, converting the source plot into a grayscale tactile representation, followed by a transformation into a channel-wise equivalent. We evaluate the performance of our model on a comprehensive synthetic dataset consisting of 20,000 source-target pairs encompassing various 2D plot types. To quantify performance, we utilize fuzzy versions of established metrics like pixel accuracy, Dice coefficient, and Jaccard index. Additionally, a human study is conducted to assess the visual quality of the generated tactile graphics. The proposed approach demonstrates promising results, significantly streamlining the conversion of 2D plots into tactile graphics. This paves the way for the development of fully automated systems, enhancing accessibility of visual information for visually impaired individuals.

Keywords: Tactile Generation, U-Net++, Perceptual Loss, Gradient Penalty

1 Introduction

Tactile graphics, encompassing elements like pictures, diagrams, maps, and graphs, utilize raised surfaces to convey information to individuals with visual impairments. These graphics serve as a critical means for non-textual communication, translating visual elements into a tactile format. They represent a subset of accessible image formats, alongside methods like verbal descriptions, sound, and haptic feedback, which aim to improve image comprehension for visually impaired individuals [1]. Although braille writing systems and automatic text-to-speech translators have proven effective in communicating information, they lack the capability to process and interpret graphics and images. Tactile graphics can be created using an embosser, similar to braille, and can be used to help users understand information through touch.

Tactile graphics leverage a discrete, multi-level grayscale encoding scheme. Foreground information is typically depicted in black, contrasting with a white background. To convey spatial relationships, controlled variations in height are introduced. These variations directly correspond to quantized grayscale values, where darker shades map to progressively raised surfaces. Notably, the number of distinct grayscale levels is often purposefully limited to eight [2]. This design choice optimizes tactile legibility by mitigating information overload and ensuring a manageable set of perceivable height variations. Figure 1 shows a simple tactile example.

Fig. 1 A sample tactile for visually impaired students. After [2]

We propose a novel approach to expedite tactile graphic generation, particularly in time-sensitive situations requiring prompt access to tactile representations. This method focuses on producing outputs that seamlessly integrate with existing third-party SVG generation tools. This approach facilitates the utilization of established design software, such as Corel Draw, Potrace, Adobe Illustrator, or PowerPoint, for tactile graphic creation. This leverages existing user familiarity with these tools, potentially streamlining the design process compared to the current method of manual tactile graphic generation within the same software. Our long-term vision centers on the development of a deep learning pipeline capable of autonomously converting RGB images into a format optimized for tactile perception. This automation ideally minimizes the need for manual designer intervention prior to embosser processing.

While many image types hold potential for conversion to tactile formats, this work focuses on automating the conversion of specific elements within statistical representation charts. These charts are crucial for conveying information in a wide range of contexts. Specifically, we aim to convert elements like Bézier curves (used for smooth lines), scatter plots (representing data points), polygon shapes (for various data enclosures), and bar charts – all into a tactile format that can be interpreted by people with visual impairments.

The development of our image-to-tactile translation model draws inspiration from the Raster-to-Vector approach [3], for converting rasterized floorplans to vector graphics. While traditional methods often rely on image pre-processing techniques like edge detection, the Raster-to-Vector approach directly identifies key components of the floorplan and reconstructs them in a new style within a vector representation. This focus on semantic understanding, rather than solely low-level features, inspired our approach of detecting key objects within a 2D image and subsequently translating them into a tactile representation.

As illustrated in Figure 2 , a bar chart is initially transformed into a non modifiable tactile representation. Subsequently, a second conversion separates the various components of the bar chart into distinct channels. This decomposition facilitates independent manipulation of each element. Both stages of the pipeline leverage neural network architectures, which will be thoroughly discussed in Section 3. This study offers several key contributions to the field of tactile graphics generation:

- A Novel Two-Step Pipeline: We propose a novel two-step pipeline for generating editable tactile representations of statistical data charts.
- Tailored Evaluation Metrics: We introduce a set of tailored evaluation metrics specifically designed for assessing the quality of tactile graphics. These metrics include pixel accuracy, dice score, and Jaccard coefficient.
- Domain-Specific Dataset Creation: We contribute to the domain by creating a new dataset of work-related tactile graphics.

The following sections detail the remainder of this paper. Section 2 reviews existing research relevant to our study. Section 3 describes the proposed model in detail. Section 4 presents the experimental setup and analyzes the obtained results. Finally, Section 6 summarizes the key findings and offers concluding remarks.

2 Related Work

While there are plenty of work on producing the tactiles physically and understanding them have been done previously, the filed of converting the variations of RGB images to the tactile forms has been relatively under explored [4–8].

Li et al. [9], presented a hierarchical framework for structuring tactile information in robotics applications. This framework categorizes tactile data into four levels: raw, contact, object, and action. Higher levels of information progressively build upon the data extracted from lower levels. The authors further explored the specific types of information that can be gleaned from each level within the hierarchy.

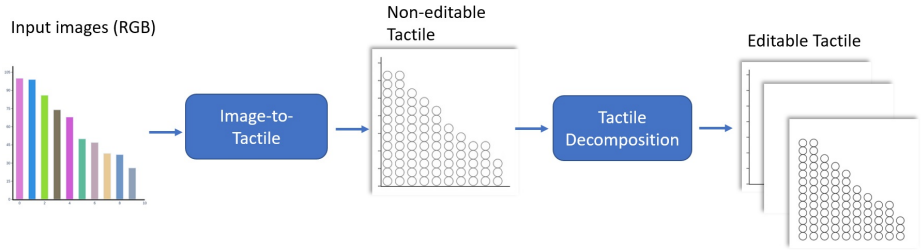

Fig. 2 The proposed pipeline for converting source images to editable tactile format.

Despite the advancements made in image-to-tactile conversion methods, these approaches can be broadly categorized into two main types: 1) **Rule-based methods:** These methods rely on pre-programmed algorithms that define how image features are translated into tactile elements. Human intervention may be necessary at various stages to ensure accurate conversion. 2) **Machine learning-based methods:** These methods leverage machine learning algorithms to learn a model that can automatically convert source RGB images into tactile representations.

The first approach relies on expert knowledge of human tactile perception to craft conversion rules. In contrast, machine learning models learn these patterns automatically by analyzing a training dataset containing source RGB images and their corresponding tactile representations.It's worth noting that machine learning-based approaches can be further subdivided into categories like image-to-image translation and image segmentation.

There has been previous attempts for automating the process of generating tactiles from images. The Tactile Graphics Assistant (TGA) is a software program developed by Ladner et al. [10] to significantly improve the efficiency and accessibility of translating visual information into tactile graphics. TGA empowers specialists to convert not only individual figures but entire textbooks of illustrations into a format suitable for blind and visually impaired individuals. TGA automates several key tasks including: image acquisition, image classification and segmentation, character recognition and tactile output generation. Štampach and Mulíčková [11], developed a rule based model which to partly automate the conversion of the map images into tactile form.

Barvir et al. [12], proposed a method for creating interactive tactile maps that are both affordable and accessible for visually impaired people. The key aspects of their model are: 1) TouchIt3D technology, this links a 3D-printed tactile map with a mobile device, providing an interactive experience. 2) OpenStreetMap data, this freely available data source is used to create the base map information.3) Semi-automated workflow,the process is designed to be efficient, reducing time and cost compared to traditional methods. This workflow likely involves automating some map design steps. Their work focuses on creating tactile maps for three purposes: Walking navigation, Public transportation use and Tourist exploration.

The study by Touya et al. [13] presents an initial exploration of adapting existing automated cartography techniques, such as map generalization, schematization, and stylization, for the creation of on-demand tactile maps. Their key contribution lies in demonstrating the potential of such an approach, while highlighting the need for further research on specific challenges outlined in a comprehensive research agenda.

Jiang et al. [14], investigate a semi-automated approach to address the limitations of manually created tactile maps for visually impaired people. The researchers developed a system that automates some design tasks involved in generating these maps, while likely leaving room for specialist input to ensure quality and customize the final product. Their initial evaluation focused on the graphic design of the produced maps, assessing clarity, information representation, and effectiveness in conveying intersection details. By involving tactile graphics professionals, the researchers ensured the design caters to the needs of visually impaired users.

Chen and Takagi [15], proposed a pattern recognition-based method for automating the translation of hand-drawn maps into tactile graphics. This method likely utilizes algorithms capable of identifying and classifying various symbols and features within the map. By automating these processes, the authors aim to improve the efficiency and potentially the accuracy of translating hand-drawn maps into tactile formats. This could make the creation of customized tactile maps for individual needs more feasible.

Engel and Weber [16], investigated the influence of design on the readability and usability of tactile charts for data exploration. Their analysis encompassed 69 tactile charts, including various chart types (bar, line, pie, area, and scatter plots) sourced from publications, accessibility guidelines, and transcription institutes. The study focused on the design of axes and tick marks, the use of labels and legends, the visual style of chart elements, and design considerations specific to each chart type. Based on their findings, the authors propose a foundational set of design guidelines for bar charts. In the study by Choi et al. [17], the authors present a deep learning approach that automatically analyzes visualizations to identify key components. This includes the visualization type, graphical elements, labels, and legends. Most importantly, the system extracts the underlying data the visualization represents. By leveraging this extracted information, the system can provide a synthesized reading experience for visually impaired users.

Watanabe and Mizukami [18], investigated the effectiveness of tactile scatter plots in conveying relationships between two quantitative variables for blind people. The study compared three data representations: tactile graphs, tactile tables, and electronic tables. Participants were tasked with identifying the relationship between variables in each format. The results revealed that participants understood tactile graphs the fastest, followed by tactile tables, while electronic tables required the longest time. Furthermore, both tactile formats received higher subjective ratings for usability compared to the electronic tables.

Gorniak et al. [19], presented VizAbility, a groundbreaking multimodal accessibility system. This system merges keyboard navigation with standard interaction methods, empowering visually impaired users to actively explore and interact with data visualizations. VizAbility leverages an LLM (Large Language Model)-based pipeline to analyze user queries. By synthesizing information from the underlying data, chart

structure, user's focus within the visualization, and external web-based information, the system generates comprehensive responses to user queries.

Building upon the widespread adoption of machine learning models, tactile image generation can also be addressed within this framework. Two primary approaches can be leveraged: image-to-image translation and image segmentation. Image segmentation can be implemented using supervised learning techniques or even rule-based programming approach. Supervised image-to-image translation tasks can be categorized as binary or multi-domain, and tackled using supervised, semi-supervised, or unsupervised learning approaches. As a crucial area of computer vision research, image-to-image translation boasts a wealth of existing research. Popular supervised approaches include Pix2Pix [20], while CycleGAN [21], exemplifies unsupervised approaches, both of which have laid the foundation for numerous GAN-based image-to-image translation solutions. However, the primary challenge associated with supervised machine learning approaches lies in the limited availability of suitable annotated data. Acquiring such data is often an expensive endeavor.

The task of image-to-sketch translation holds relevance for the domain of image-to-tactile translation, as both sketches and tactile graphics represent a human-interpretable abstraction of an image. Several Generative Adversarial Network (GAN) models have been proposed for bidirectional image-to-sketch conversion [22]. These models incorporate features like identity preservation, sketch quality, and composition-aided generation to synthesize sketches that capture both the essence and key details of the original image.

Sketches and tactile graphics both reflect human comprehension of an image, making image to sketch translation relevant to the problem. Several GAN models have been proposed for one-to-many [23, 24] and bidirectional image to sketch conversion [22]. These models use features like identity preservation, sketch quality, and composition-aided generative adversarial networks to synthesize sketches and facial features. Key facial features are learned to be embedded in the features and mapped to real photos, and a spatial attention pooling module and dual generator training technique are used in the DeepFacePencil [25] model.

3 Proposed Method

Our proposed method for tactile generation utilizes a pipeline with two modules. The first module performs image-to-haptic translation, converting RGB images into non-editable tactile representations. The second module refines these non-editable tactiles into editable formats suitable for tactile display. This refinement process involves component segmentation, where the model identifies individual elements within the tactile representation. Each element is then mapped to a corresponding channel within the editable tactile format. Finally, inpainting techniques are employed to address any gaps or discontinuities that may arise due to the intersection of different components. A two-step pipeline offers several advantages over a single-step approach for tactile generation. One key benefit is the reduced need for channel-wise training data. Such data, where each tactile element is isolated within a specific channel, is significantly more challenging to acquire compared to grayscale tactile images. In the proposed

pipeline, the first module leverages readily available grayscale data. This module performs content pre-processing, removing unnecessary details and adjusting the style of each component to conform to tactile representation conventions. This pre-processing simplifies the subsequent task of component decomposition handled by the second module. Since the first module has already established a well-defined tactile representation, the second module can focus on the more manageable task of identifying and separating individual components within the pre-processed tactile data.

The proposed pipeline leverages a variant of Pix2Pix [20] with a PatchGAN discriminator [26] for both stages, Unlike standard discriminators that analyze entire images, PatchGAN acts as a zoomed-in critic in a GAN. It examines tiny image squares (N x N pixels) and determines if each one is a genuine part of a real image or a fabrication by the generator.

Notably, we employ U-Net++ [27] instead of the standard U-Net architecture. U-Net++ offers several advantages over U-Net, including the introduction of nested, dense skip-connections and a deep supervision scheme. These enhancements facilitate improved feature fusion, leading to superior segmentation performance and enhanced gradient flow within the network [27].

To enhance the performance of our model beyond the capabilities of the original architecture, we extend the loss function by incorporating two additional terms: adversarial perceptual loss (APL) [28] and gradient penalty (GP) [29]. APL improves the perceptual realism of the generated images by guiding them to not only resemble real data statistically (as achieved by the distance-based loss) but also to deceive a pre-trained image recognition network. This ensures the generated images capture the high-level features and details that humans perceive as important. Furthermore, GP is introduced to promote stable training. By enforcing a penalty on the gradient norm of the discriminator, GP mitigates the vanishing/exploding gradient problem, leading to a more efficient and robust training process. The combined effect of these additional loss terms fosters the generation of perceptually realistic images while ensuring a well-behaved training trajectory.

The GAN loss term is responsible for implementing adversarial training between the generator and the discriminator, playing a pivotal role in evaluating the model's performance. Equations 1 and 2 represent the formulation of this term. The criterion for the GAN loss, denoted as f_c, offers candidates such as mean squared error, binary cross-entropy, hinge loss, or Wasserstein loss. The patch scores for real data and generated data are represented by y and $\hat{y}$, respectively. To simplify calculations, we employed $\mathbf{J}|y|$ to denote a matrix of ones matching the size of y, and $\mathbf{O}|\hat{y}|$ to denote a matrix of zeros matching the size of $\hat{y}$. The discriminator module is represented by D.

$$L_{GAN}(G) = \sum_{x,z} \left(f_c \left(D\left(x, G\left(x\right)\right), \; J_{|\hat{y}|} \right) \right) \tag{1}$$

$$L_{GAN}(D) = \sum_{x,z} \left(f_c \left(D\left(x, z\right), \; J_{|y|} \right) \; + \; f_c \left(D\left(x, G\left(x\right)\right), \; O_{|\hat{y}|} \right) \right) \tag{2}$$

The distance-based loss term effectively penalizes the generator by comparing its output to the ground truth, employing L1 loss to ensure the preservation of low-frequency accuracy and fidelity. This L1 loss term can be expressed as shown in

equation 3, where x represents the input and z denotes the ground truth.

$$L_{L1}(G) = \sum_{x,z} |z - G(x)| \tag{3}$$

Building upon the insights from [30], our approach integrates the discriminator's feature maps to provide valuable perceptual feedback to the generator throughout the training process. To achieve this, we utilized the absolute difference instead of the second norm. This loss term can be viewed as an extension of the L1 loss, as it not only guides the generator based on the desired output but also takes into account the feature maps it triggers in the discriminator, promoting a more profound perceptual comprehension of input images. We refer to this loss term as 4, where ϕ_κ represents the specific feature maps selected from the discriminator immediately after passing through the activation function.

$$L_{per}(G) = \sum_{\kappa \in K} \sum_{x,z} |\phi_\kappa(G(x)) - \phi_\kappa(z)| \tag{4}$$

To mitigate the imbalanced competition between the generator and discriminator during training, we employed a modified version of gradient penalty originally proposed in [29]. By periodically penalizing the discriminator for pronounced changes induced by strong gradient signals, we ensured its stability without compromising training speed or convergence. This technique involved interpolating between the input and generator output, using a random matrix α, to generate an intermediary value that approximated the ground truth more closely. The specific formulation of this loss term can be found in equation 5.

$$L_{gp}(D) = \sum_{x,z} (||\nabla_{\hat{z}} D(x, \hat{z})||_2 - 1)^2 \tag{5}$$

The optimization problem or minimax game is formed by the collection of the above loss terms as shown in equation 6. In this formulation, the wight of each loss term is expressed as a scalar value, denoted by the symbol λ.

$$\begin{aligned} G^* = arg \min_G \max_D L_{GAN}(G, D) + \lambda_a L_{L1}(G) \\ + \lambda_{gp} L_{GP}(D) + \lambda_{per} L_{per}(G) \end{aligned} \tag{6}$$

4 Experimental Setup

This section details our dataset specifically designed for the task of converting 2D plots into a tactile format suitable for visually impaired users. We then outline evaluation metrics to assess model efficacy in achieving this objective. Furthermore, the experimental settings are detailed at the end of this section.

4.1 Dataset

Recent advancements in deep learning have facilitated the development of end-to-end models, capable of integrating multiple tasks within a single pipeline. This eliminates

the need for manual feature extraction and extensive preprocessing or postprocessing steps, streamlining the overall workflow. However, a critical factor for leveraging these advancements is access to a sufficiently large and diverse training dataset. To address this challenge, we have synthesized two comprehensive datasets encompassing a total of 5,000 samples.

The first dataset caters to a variety of 2D plot types, including Bézier curves, scatter plots, and polygons, offering a well-rounded representation of common visualizations. The second dataset focuses specifically on bar charts, providing targeted training data for this prevalent chart type. Both datasets encompass three key components:

- RGB Images: The original visual representation of the 2D plots.
- Non-Editable Tactile Images: The desired output format, representing the tactile representation for visually impaired users.
- Channel-wise Tactile Image Triplets: Intermediate representations potentially useful for model training.

To ensure robust model evaluation, we employ a standard data split. 90% percent (4,500 samples) of the data is allocated for training the model, while the remaining 10% percent (500 samples) is reserved for unseen testing. This approach mitigates overfitting and allows for a more accurate assessment of model generalizability on novel data (refer to Figure 3 for dataset examples).

The channel-wise output structure offers several advantages. Each channel encodes a distinct tactile component, allowing the model to treat it as an independent grayscale image. This representation effectively addresses two key challenges: 1) Quality and Class Imbalance: Compared to a single-step approach, the channel-wise format mitigates issues related to data quality and class imbalance. 2) Object Reconstruction and Inpainting: The separation of information into channels facilitates seamless object reconstruction and inpainting within the tactile image. This is particularly beneficial in scenarios where elements overlap in the input image, such as grid lines intersecting axes. By processing each component independently, the channel-wise approach avoids discontinuities and ensures a smooth, continuous representation throughout the entire image and its individual tactile components.

4.2 Evaluation Metrics

Common image quality metrics, such as Fréchet Inception Distance (FID) [31] and Inception Score (IS) [32], are not well-suited for evaluating the quality of tactile images generated by our model. These metrics rely on pre-trained image classifiers, which are designed for natural scene images and may not capture the nuances of tactile representations. For instance, FID and IS might penalize slight variations in background texture that are acceptable or even desirable in tactile graphics.

We required metrics that better reflect the segmentation nature of our task and treat pixel intensities as continuous values. Fuzzy logic provided a suitable framework for this purpose. Inspired by the concept of membership functions, we modeled pixel intensities using smooth set memberships. This allows for a more nuanced evaluation

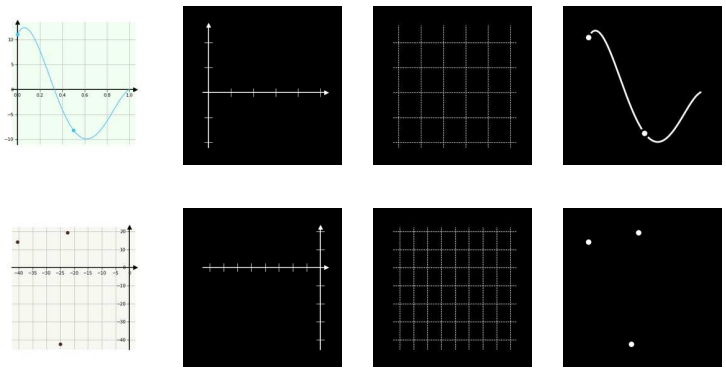

Fig. 3 Examples of the synthesized dataset. The left column shows the source domain. The other three columns show axes, gridlines, and content channels, respectively.

by incorporating a flexible range of values between 0 and 1, rather than relying solely on discrete class labels.

Consequently, we introduce fuzzy adaptations of standard segmentation metrics: pixel accuracy, Dice score, and Jaccard coefficient (formulated in Equations 7, 8, and 9 respectively), where r_i represents the pixels of the ground truth and g_i represents those of the generated output.

$$Pixel\ Accuracy\ =\ \frac{\sum_i (min\,(r_i,\ g_i))}{\sum_i (r_i)} \tag{7}$$

$$Dice\ Score\ =\ \frac{2\sum_i (r_i\ .\ g_i)}{\sum_i (r_i^2\ +\ g_i^2)} \tag{8}$$

$$Jaccard\,Coefficient\ =\ \frac{\sum_i (r_i\ .\ g_i)}{\sum_i (r_i^2\ +\ g_i^2\ -\ r_i\ .\ g_i)} \tag{9}$$

This approach offers several advantages over traditional metrics. By leveraging fuzzy logic, we can capture the inherent complexities of tactile image quality. Pixel intensities in tactile images represent continuous values, unlike the discrete classes used in natural scene images. Fuzzy logic allows us to account for this fuzziness by treating intensities as degrees of membership in fuzzy sets. This enables a more nuanced evaluation that considers partial matches between the ground truth and generated images.

Furthermore, the proposed fuzzy adaptations reward the presence of accurate content pixels in the foreground regions. Conversely, they penalize the introduction of artifacts or unwanted elements in the background areas. This comprehensive evaluation ensures the model prioritizes faithful representation of the desired tactile features while minimizing background noise.

To prioritize visual quality, we conducted qualitative evaluations alongside quantitative measures. An anonymous web application was developed for human annotators

to rank assorted images generated by different models compared to the ground truth. The ranking was based on criteria such as reduced artifacts, accurate position and color translation, and style consistency. A subset of 100 randomly selected test images per model received annotations, averaging the results. Without imposing specific priorities or weights, annotators ranked the images naturally and fairly. The findings confirmed the superior performance of our model, demonstrating higher quality tactiles compared to the base model.

4.3 Experimental settings

To improve model generalizability and robustness to input variations, we incorporated a data augmentation strategy within our data loader pipeline. This strategy involved a series of random transformations applied to the input data with pre-defined probabilities: **Horizontal Flipping**: Randomly flipped the image horizontally with a 50% chance, augmenting the dataset with mirrored versions and enhancing the model's ability to learn rotation-invariant features. **Shifting**: Randomly shifted the image content within a range of 10% of its original dimensions in both horizontal and vertical directions. This simulates potential misalignments during data acquisition and improves the model's tolerance to slight spatial variations. **Scaling**: Randomly scaled the image size up or down within a 20% range of its original dimensions. This introduces variations in object size and helps the model learn features at different scales. **Rotation**: Randomly rotated the image by up to 15 degrees in either direction. This augments the dataset with rotated versions and improves the model's ability to recognize objects regardless of their orientation. **Partial Occlusion**: Randomly applied partial occlusion to the image with a 50% chance. This can simulate scenarios where parts of the input data might be obscured and strengthens the model's ability to handle incomplete information.

5 Results and Ablation Analysis

To facilitate a performance comparison, we employed Pix2Pix as the baseline model. To evaluate the contribution of individual components within our method, we conducted an ablation analysis. This involved systematically replacing or removing elements and observing the impact on performance.

Firstly, we investigated the limitations of the original generator architecture in handling channel-wise outputs. We replaced it with a modified version to address these limitations. Subsequently, a gradient penalty loss term was introduced to improve the stability of the training process. Finally, a perceptual loss term was incorporated to further enhance the quality of the generated outputs.

The quantitative results of the comparison are presented in Tables 1 and 2. Table 1 displays the results on 2D plots including Bézier curves, scatter plots, and polygons, while Table 2 utilizes bar charts for visualization. There are two modules in the pipeline. We trained each module using different loss components. The *+upp* on the left denotes the base model but U-Net++ instead of U-Net as the generator of the first module. The *+upp* on the top row denotes the base model but U-Net++ instead of U-Net as the generator of the second module. The *+gp* on the left denotes the base

model with U-Net++ and gradient penalty in the first module. The *+gp* on the top row denotes the base model with U-Net++ and gradient penalty in the second module. Likewise, *+per* denotes the base model but with U-Net++ and gradient penalty and adversarial perceptual loss. The three values reported in each cell are pixel accuracy (PA), dice (DS), and Jaccard coefficient (JC), respectively. The results demonstrate the effectiveness of each additional component in the proposed method. Furthermore, in the last column, we report the result of a single-step approach for directly converting RGB source images to editable tactiles. It is evident that the proposed two-step pipeline outperforms the single step approach.

Our evaluations (Tables 1 & 2) demonstrate that directly converting RGB images to editable tactiles outperforms the baseline model across all metrics. Further improvements are achieved by replacing the U-Net with a U-Net++ architecture in the first module. Additionally, incorporating a gradient penalty term in the first module leads to further performance gains.

Interestingly, adding the adversarial perceptual loss to the first module while using the base configuration for the second module shows no significant impact on the 2D plots but does affect the bar chart outputs. Replacing the U-Net with a U-Net++ architecture in the second module while maintaining the base configuration for the first module leads to a substantial performance increase. Notably, the best performance is achieved when the U-Net++ architecture, gradient penalty, and adversarial perceptual loss are all incorporated in the second module, alongside the base configuration in the first module.

While using both modules with all proposed modifications (U-Net++, gradient penalty, adversarial perceptual loss) yields improved performance, our results suggest that omitting the adversarial perceptual loss from the first module achieves the optimal outcome.

1 \ 2		base	+upp	+gp	+per	direct
base	PA	73.96	76.15	87.31	88.06	86.43
	DS	91.81	94.03	95.10	93.93	92.85
	JC	85.23	88.98	90.90	88.84	87.25
+upp	PA	74.10	75.65	87.16	87.94	85.09
	DS	91.93	93.72	95.04	95.38	92.83
	JC	95.50	88.55	90.90	91.45	87.33
+gp	PA	76.45	78.46	90.37	**91.26**	87.63
	DS	94.94	95.86	97.24	**97.66**	87.33
	JC	88.96	92.19	94.75	**95.49**	78.14
+per	PA	73.41	74.82	87.33	88.50	91.05
	DS	91.58	93.14	94.72	95.19	96.19
	JC	84.88	87.53	90.24	91.01	93.31

Table 1 Ablation analysis on 2D plots. +upp: U-Net++ instead of U-Net in the base module, +gp: plus gradient penalty added to +upp, +per: adversarial perceptual loss added to +gp. 1 and 2 stand for module1 and module2 respectively.

1 \ 2		base	+upp	+gp	+per	direct
base	PA	55.76	57.09	60.62	84.00	59.79
	DS	74.96	74.97	77.94	92.57	78.96
	JC	63.86	63.66	68.06	87.23	69.58
+upp	PA	58.19	59.75	63.90	85.71	62.41
	DS	78.62	79.00	81.98	93.42	82.85
	JC	67.29	67.75	71.96	88.63	73.27
+gp	PA	67.20	68.62	74.04	**92.45**	72.55
	DS	89.00	88.64	92.57	**97.39**	93.41
	JC	81.17	80.66	86.84	**95.23**	88.27
+per	PA	67.77	68.76	74.58	90.49	82.80
	DS	89.55	88.79	93.17	96.70	94.81
	JC	83.15	83.03	90.76	94.17	90.96

Table 2 Ablation analysis on bar charts. +upp: U-Net++ instead of U-Net in the base module, +gp: plus gradient penalty added to +upp, +per: adversarial perceptual loss added to +gp. 1 and 2 stand for module1 and module2 respectively.

To complement the quantitative analysis, we conducted a qualitative evaluation on a randomly selected subset of 100 samples from the test set. In a double-blind experiment, two users were asked to rank the outputs of the proposed model and baseline models without knowledge of their corresponding labels. The results consistently indicated that the proposed model generated tactiles of superior quality, demonstrating a significant performance advantage over the baseline models. Tables 3 and 4 provide a detailed summary of our evaluation for 2D plots and bar charts respectively. To address the challenges associated with ranking a vast number of images, we employed a fixed U-Net++ architecture with gradient penalty in the first module of our proposed model. We then explored the impact of various configurations in the second module on image ranking performance.

	Base model (Pix2Pix)	Base model, U-Net++	Base model, U-Net++, GP	Base model, U-Net++, GP, Perceptual
Avg rank (1-4)	2.89	3.11	2.31	1.69

Table 3 Qualitative Evaluation (Ranks) of Tactile Generator Models on 2D Plots.

	Base model (Pix2Pix)	Base model, U-Net++	Base model, U-Net++, GP	Base model, U-Net++, GP, Perceptual
Avg rank (1-4)	3.51	3.26	2.05	1.18

Table 4 Qualitative Evaluation (Ranks) of Tactile Generator Models on Bar Charts

As depicted in Table 3 and 4, the proposed model, with its second module incorporating U-Net++, gradient penalty, and perceptual loss, consistently achieved the highest ranking (lowest value) for both chart types.

6 Conclusion and Future Work

In this study we explored the generation of tactile graphics using deep generative models for diverse 2D plots, including curves, polygons, scatter plots, and bar charts. We propose a novel approach that modifies the Pix2Pix model by employing a U-Net++ generator, incorporating an enhanced loss function, and generating multi-channel outputs for individual plot component access. Additionally, we introduce tailored evaluation metrics including pixel accuracy, dice score, and Jaccard coefficient. Our results demonstrate significant improvements over the baseline model across all plot categories.

Our proposed method significantly improved the generation of tactile graphics across various plot categories. The results demonstrate a notable improvement in the model's ability to accurately represent these plots. Pixel accuracy increased by 4.83 percentage points (from **86.43**% to **91.26**%), indicating better classification of individual pixels in the tactile representation. Additionally, the dice coefficient rose from 92.85 to 97.66, signifying a significant improvement in capturing the overall structure and spatial relationships within the plots. This is further corroborated by the Jaccard coefficient, which increased from 87.25 to 95.49, indicating better agreement between the model's output and the desired tactile representation.

Similar improvements were observed for bar charts. Pixel accuracy saw a dramatic increase from **59.79**% to **92.45**%, suggesting a significant enhancement in generating the specific tactile elements of bar charts. This includes the bars themselves and potentially additional information like axes or labels. The dice coefficient (78.96 to 97.39) and Jaccard coefficient (69.58 to 95.23) also showed substantial improvements, aligning with the findings for the combined category. These results indicate a better match between the model's generated tactile bar charts and the desired representations.

However, while our approach demonstrates promising results for these core plot categories, further exploration is necessary to expand the model's capabilities. Future work should investigate the generation of additional plot types, such as pie charts, maps, and molecule diagrams. Additionally, the current dataset comprised of synthesized examples necessitates further development for increased diversity. A more diverse dataset encompassing real-world tactile graphics will be crucial for ensuring the model's generalizability to a wider range of scenarios.

References

[1] McCallum, D., Ahmed, K., Jehoel, S., Dinar, S., Sheldon, D.: The design and manufacture of tactile maps using an inkjet process. Journal of engineering design **16**(6), 525–544 (2005)

[2] Sasing, B.G., See, A.R., Advincula, W.D., Chen, Y.-J.: A mobile application-based learning aid developer for teaching visually impaired students. In: 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1–2 (2021). IEEE

[3] Liu, C., Wu, J., Kohli, P., Furukawa, Y.: Raster-to-vector: Revisiting floorplan transformation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2195–2203 (2017)

[4] Xu, J., Kim, S., Chen, T., Garcia, A.R., Agrawal, P., Matusik, W., Sueda, S.: Efficient tactile simulation with differentiability for robotic manipulation. In: Conference on Robot Learning, pp. 1488–1498 (2023). PMLR

[5] Gu, Y., Demiris, Y.: Vttb: A visuo-tactile learning approach for robot-assisted bed bathing. IEEE Robotics and Automation Letters (2024)

[6] Schirmer, A., Croy, I., Ackerley, R.: What are c-tactile afferents and how do they relate to "affective touch"? Neuroscience & Biobehavioral Reviews, 105236 (2023)

[7] Xu, B., Zhong, L., Zhang, G., Liang, X., Virtue, D., Madan, R., Bhattacharjee, T.: Cushsense: Soft, stretchable, and comfortable tactile-sensing skin for physical human-robot interaction. arXiv preprint arXiv:2405.03155 (2024)

[8] Liu, H., Guo, D., Zhang, X., Zhu, W., Fang, B., Sun, F.: Toward image-to-tactile cross-modal perception for visually impaired people. IEEE Transactions on Automation Science and Engineering **18**(2), 521–529 (2021) https://doi.org/10.1109/TASE.2020.2971713

[9] Li, Q., Kroemer, O., Su, Z., Veiga, F.F., Kaboli, M., Ritter, H.J.: A review of tactile information: Perception and action through touch. IEEE Transactions on Robotics **36**(6), 1619–1634 (2020)

[10] Ladner, R.E., Ivory, M.Y., Rao, R., Burgstahler, S., Comden, D., Hahn, S., Renzelmann, M., Krisnandi, S., Ramasamy, M., Slabosky, B., *et al.*: Automating tactile graphics translation. In: Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 150–157 (2005)

[11] Štampach, R., Mulíčková, E.: Automated generation of tactile maps. Journal of Maps **12**(sup1), 532–540 (2016)

[12] Barvir, R., Vondrakova, A., Brus, J.: Efficient interactive tactile maps: A semi-automated workflow using the touchit3d technology and openstreetmap data. ISPRS International Journal of Geo-Information **10**(8), 505 (2021)

[13] Touya, G., Christophe, S., Favreau, J.-M., Ben Rhaiem, A.: Automatic derivation of on-demand tactile maps for visually impaired people: first experiments and research agenda. International Journal of Cartography **5**(1), 67–91 (2019)

[14] Jiang, Y., Lobo, M.-J., Jouffrais, C., Christophe, S.: Producing accessible intersection maps for people with visual impairments: an initial evaluation of a semi-automated approach. Cartography and Geographic Information Science, 1–21 (2024)

[15] Chen, J., Takagi, N.: A pattern recognition method for automating tactile graphics translation from hand-drawn maps. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4173–4178 (2013). IEEE

[16] Engel, C., Weber, G.: Analysis of tactile chart design. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 197–200 (2017)

[17] Choi, J., Jung, S., Park, D.G., Choo, J., Elmqvist, N.: Visualizing for the non-visual: Enabling the visually impaired to use visualization. In: Computer Graphics Forum, vol. 38, pp. 249–260 (2019). Wiley Online Library

[18] Watanabe, T., Mizukami, H.: Effectiveness of tactile scatter plots: comparison of non-visual data representations. In: Computers Helping People with Special Needs: 16th International Conference, ICCHP 2018, Linz, Austria, July 11-13, 2018, Proceedings, Part I 16, pp. 628–635 (2018). Springer

[19] Gorniak, J., Ottiger, J., Wei, D., Kim, N.W.: Vizability: Multimodal accessible data visualization with keyboard navigation and conversational interaction. In: Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–3 (2023)

[20] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)

[21] 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, pp. 2223–2232 (2017)

[22] Zhu, M., Li, J., Wang, N., Gao, X.: A deep collaborative framework for face photo–sketch synthesis. IEEE transactions on neural networks and learning systems **30**(10), 3096–3108 (2019)

[23] Osahor, U., Kazemi, H., Dabouei, A., Nasrabadi, N.: Quality guided sketch-to-photo image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 820–821 (2020)

[24] Lin, Y., Ling, S., Fu, K., Cheng, P.: An identity-preserved model for face sketch-photo synthesis. IEEE Signal Processing Letters **27**, 1095–1099 (2020)

[25] Li, Y., Chen, X., Yang, B., Chen, Z., Cheng, Z., Zha, Z.-J.: Deepfacepencil: Creating face images from freehand sketches. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 991–999 (2020)

[26] Demir, U., Unal, G.: Patch-based image inpainting with generative adversarial networks. arXiv preprint arXiv:1803.07422 (2018)

[27] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, pp. 3–11 (2018). Springer

[28] Qu, X., Wang, X., Wang, Z., Wang, L., Zhang, L.: Perceptual-dualgan: perceptual losses for image to image translation with generative adversarial nets. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2018). IEEE

[29] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. Advances in neural information processing systems **30** (2017)

[30] Wang, C., Xu, C., Wang, C., Tao, D.: Perceptual adversarial networks for image-to-image transformation. IEEE Transactions on Image Processing **27**(8), 4066–4079 (2018)

[31] Yu, Y., Zhang, W., Deng, Y.: Frechet inception distance (fid) for evaluating gans. China University of Mining Technology Beijing Graduate School (2021)

[32] Barratt, S., Sharma, R.: A note on the inception score. arXiv preprint arXiv:1801.01973 (2018)