



OPEN Tactile-GAT: tactile graph attention networks for robot tactile perception classification

Lun Chen✉, Yingzhao Zhu & Man Li

As one of the most important senses in human beings, touch can also help robots better perceive and adapt to complex environmental information, improving their autonomous decision-making and execution capabilities. Compared to other perception methods, tactile perception needs to handle multi-channel tactile signals simultaneously, such as pressure, bending, temperature, and humidity. However, directly transferring deep learning algorithms that work well on temporal signals to tactile signal tasks does not effectively utilize the physical spatial connectivity information of tactile sensors. In this paper, we propose a tactile perception framework based on graph attention networks, which incorporates explicit and latent relation graphs. This framework can effectively utilize the structural information between different tactile signal channels. We constructed a tactile glove and collected a dataset of pressure and bending tactile signals during grasping and holding objects, and our method achieved 89.58% accuracy in object tactile signal classification. Compared to existing time-series signal classification algorithms, our graph-based tactile perception algorithm can better utilize and learn sensor spatial information, making it more suitable for processing multi-channel tactile data. Our method can serve as a general strategy to improve a robot's tactile perception capabilities.

Keywords Haptics, Deep learning, Time-series analysis, Graph attention networks, Tactile perception, Tactile classification

Human beings observe the world through their eyes and experience it through touch. Tactile perception, an important way of sensing, has garnered increasing attention in recent years. Tactile perception can provide different types of information compared to visual perception, such as temperature, weight, surface texture, and the softness and hardness of objects.

Tactile perception also plays a crucial role in many applications such as robotics, minimally invasive surgery¹, advanced prosthetics², and manufacturing³. As robots are increasingly deployed in unstructured and complicated environments⁴, they need to perform manipulative tasks like grasping objects with arbitrary unknown shapes and avoiding sliding while exerting minimal force on the objects. While the visual sensor can recognize and provide images of the target, these contact-less features are limited only to the appearance information and cannot acquire the physical properties of the target. On the other hand, the tactile sensor can recognize and provide physical property information of the target. Tactile feedback helps the manipulator to perceive the hardness⁵, roughness^{6,7}, and friction⁸ of the target while grasping it, and adding tactile sensors to the manipulator can help to obtain tactile perception, which assists in estimating the stability of the target grasp and target recognition. Therefore, developing a method for robots to learn human tactile perception is an urgent scientific problem in the field of intelligent robot research.

Nowadays, deep learning has produced fruitful research results in the processing of complex data such as text, images, and voice. For instance, in tasks like object classification⁹, image segmentation¹⁰, and disease detection¹¹, it has even surpassed human-level accuracy. It has even surpassed human-level accuracy in some complex tasks. The continuous development of deep learning and sensor technology has promoted interdisciplinary research on manipulator tactile perception. Currently, most tactile perception research uses computer vision to infer the tactile properties of objects, i.e., to predict the object type, weight, and material from the object image. However, the robustness of machine vision is low when the camera is exposed to variable light sources, such as living scenes. Also, when the object is non-rigid (easy to deform), it is challenging for the robot to achieve stable grasping based solely on the single tactile attribute of machine vision. To address these issues, researchers in robotics and vision have contemplated equipping robots with supplementary tactile sensors to enable them to acquire tactile information.

China Telecom Research Institute, Guangzhou, China. ✉email: chenl103@chinatelecom.cn

The framework for robot tactile perception based on deep learning is illustrated in Fig. 1. The robot collects tactile data through sensors, such as pressure and bending sensors. Deep learning algorithms then extract meaningful features from the tactile data and fuse the features obtained from various sensors to acquire more comprehensive and precise tactile information. Finally, the robot's movements are regulated or modified based on the fused tactile information to accommodate diverse tasks and environments.

However, there are several issues with the existing tactile perception algorithm based on deep learning.

- (1) Mostly based on vision rather than sensors: Due to the limitations of size, cost, and performance of tactile sensors, most existing tactile perception algorithms focus on using visual information to make cross-modal inferences about tactile information. These algorithms, which rely on visual input, do not take into account the structural characteristics of tactile signals and their perception accuracy is easily affected by environmental factors.
- (2) Single tactile signal: Most robotic arms and tactile gloves only use pressure information collected by the pressure sensor on the device as a single tactile signal. A single pressure signal cannot fully represent multi-dimensional tactile signals. Just as when a human hand grasps an object, it will feel the deformation of the object through the pressure change of the palm and the curvature degree of the fingers, and then infer the softness and hardness of the object.
- (3) The physical spatial information among tactile sensors has not been utilized: The same and different types of tactile signals may be correlated in space. For example, when grasping an object, the pressure information of different positions of the hand is correlated, and the pressure information of the hand is also related to the curvature degree of fingers.

In this study, we address these issues by proposing a new sensor-based tactile perception framework using a graph attention mechanism, named Tactile-GAT. This framework leverages explicit and implicit relation graphs to enhance the classification performance of tactile signals.

Our contributions are outlined as follows:

1. Utilization of Spatial Relationships: Unlike most methods that process temporal signals without considering spatial adjacency relationships, our Tactile-GAT framework explicitly utilizes the spatial relationships among multi-channel and multi-type tactile information. This approach allows for a more nuanced understanding and processing of tactile data.
2. Graph Structure Networks: We demonstrate that graph structure networks are inherently more suitable for processing tactile signals compared to traditional methods. By adapting these networks to tactile perception, our framework can better capture the complex interdependencies between different tactile signals.
3. Introduction of Embedding Layers: Our model includes embedding layers that learn the connectivity and weights between different tactile signals. This innovation allows for a dynamic and context-aware interpretation of tactile data, enhancing the robot's ability to make informed decisions based on tactile input.

This paper is organized as follows: Sect. [Related work](#) gives an overview of various methods for tactile perception classification, while Sect. [Dataset](#) provides the dataset. Section [Tactile graph attention network](#) discusses the proposed framework. Section [Experiments and results](#) focuses on results and analysis, and finally, Sect. [Discussion and conclusion](#) draws conclusions and provides perspectives.

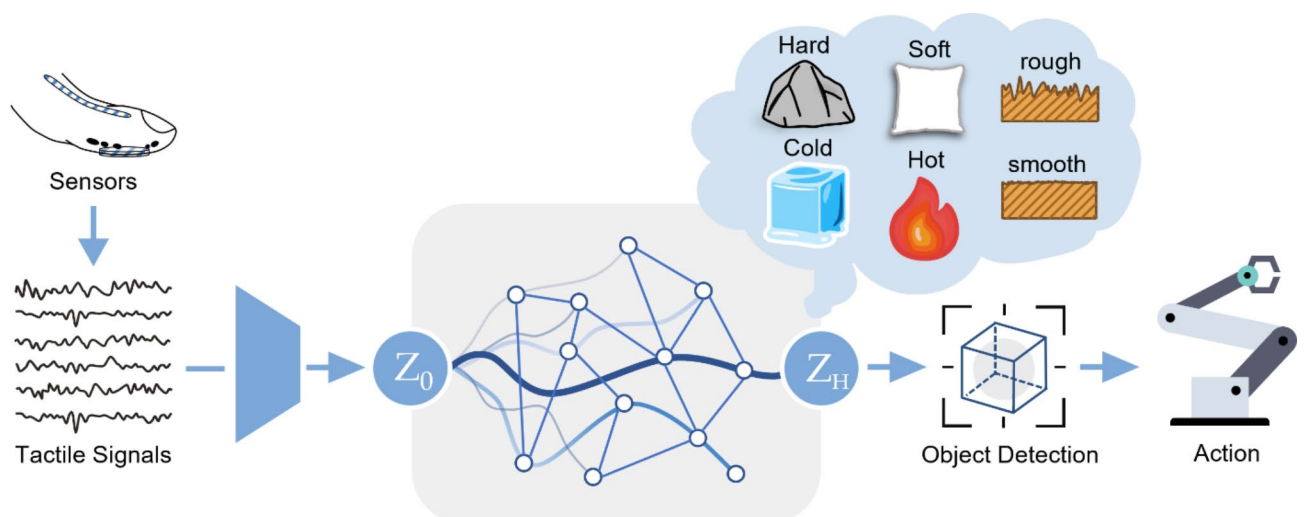


Fig. 1. The framework of robot tactile perception.

Related work

Tactile perception ability can effectively help us plan and adapt and change our interaction strategy with objects and the environment in real-time. For example, when we observe that the object is slippery before grasping, we instinctively increase the grip force to increase friction; Similarly, when the object is deforming, we will adjust the grip force accordingly. To endow robots with the ability to acquire tactile information and enable them to more effectively handle various objects and environments in industrial environments and our daily settings, researchers have combined deep learning with tactile perception. Deep learning¹² is a machine learning algorithm with a multi-layer structure that can extract higher-level features from the original input. By using adaptive feature extraction and classification modules, deep learning can achieve end-to-end learning and no longer rely on complex hand-crafted features. Deep learning can improve and expand existing tactile signal processing methods in multiple dimensions. It can extract features from unprocessed or minimally processed data to obtain more effective feature information and higher classification accuracy¹³.

Based on the representation forms of tactile signals, existing research methods can be divided into two categories: time-series tactile features and image-type tactile features.

Time-series tactile features

Time-series tactile features involve capturing tactile signals that vary over time, reflecting the dynamic interaction between the robot and the environment. These signals include temporal sequences of pressure, force, and other tactile data, which are crucial for recognizing textures, shapes, and other object properties.

Sundaram³¹ proposed a low-cost haptic glove with an array of 548 flexible pressure sensors. This glove collects tactile pressure maps of different objects, which are used with a deep convolutional neural network to extract features, enabling object recognition based solely on the sense of touch without the aid of vision.

Rasouli³⁴ used piezoresistive fabrics to mimic the tactile sensing units in the skin. By touching a textured surface with an artificial finger equipped with a tactile sensor array, the collected temporal tactile signals are transformed into spike patterns. In the texture recognition task, ten graded textures can be classified with 92% accuracy.

Yan²⁹ proposed a texture recognition method that uses a curved soft-tactile sensor and a long short-term memory (LSTM) model with an attention mechanism to recognize Braille characters and various fabrics. This approach processes sequential tactile data to effectively capture temporal patterns in the tactile signals.

The GTac-Hand³², which integrates advanced tactile sensors capable of detecting not only pressure but also shear forces, provides robots with detailed feedback about the objects they interact with, allowing for more nuanced manipulations in complex environments.

Satoshi³³ developed a multi-fingered robotic hand enhanced with graph convolutional networks (GCNs) and 1152 distributed tactile sensors for in-hand manipulation of diverse objects. This configuration leverages human manipulation data to adaptively adjust robot grip based on object properties such as size, shape, and hardness, enhancing the safety and robustness of handling fragile items.

The DIGIT sensor³⁰, developed by Lambeta et al., represents a significant advancement in tactile technology. It is a compact and high-resolution sensor that captures detailed tactile images of object surfaces it contacts. This capability allows the sensor to detect fine surface textures and shapes, translating these physical interactions into digital tactile maps. These maps are then processed using deep learning algorithms to infer properties such as the object's material composition, weight, and potential slipperiness.

Image-type tactile features

Image-type tactile features refer to devices or sensors that collect tactile signals directly or indirectly through cameras. Since the collected tactile information is presented in the form of images. This approach allows for the application of computer vision techniques to tactile data, and the processing algorithms differ from those used for time-series signals. It facilitates detailed analysis of surface textures, shapes, and other physical properties by leveraging image-based processing methods.

Due to the constraints of sensor size and cost, the existing touch sensors used for robotic tactile perception are mostly GelSight^{21–24}, pressure^{25–28}, bending, and friction sensors²⁹.

Calendra³⁵ proposed an action-conditioned depth model that combines a Gelsight sensor with an RGB camera. By jointly predicting manipulator grasping outcomes and planning action sequences using visual and tactile inputs, the model can improve grasping performance. Gao³⁶ used convolutional neural networks (CNN) to extract features from visual and temporal tactile signals and demonstrated that visual data can be effectively transferred to tactile classification tasks in various but related tasks. Guo³⁷ employed deep learning networks to detect proper grip rectangles for a robot when grasping objects by utilizing information from vision and multimodal tactile sensations. Experimental results indicate that integrating tactile data can improve the model's ability to extract visual information from images. Kumra¹⁹ proposed Generative Residual Convolutional Neural Network (GR-ConvNet), which uses N-channel input images to infer the grasping rectangle of an object in the image and provides a grasping strategy for unknown objects. Additionally, Purri²⁰ proposed a cross-modal framework that can learn complex mappings between visual information and tactile physical properties. The model can infer 15 tactile physical properties, including friction and texture, from multi-view images of an object's surface.

Tactile signals have the following two characteristics:

Multi-dimensionality: Tactile signals involve various features, such as pressure, friction, temperature, and humidity, which are coupled with each other. However, there is still a lack of clear mapping models to represent these signal features comprehensively.

Contextual Relevance: Tactile signals exhibit continuity in both time and space, with preceding and subsequent data showing causal changes. For example, there is a correlation in the spatiotemporal characteristics of pressure signals and a linkage relationship between multidimensional data.

In recent years, graph neural networks (GNNs)³⁵ have been developed to model complex patterns in graph-structured data. GNNs have practical applications in areas such as antibacterial discovery, physics simulations³⁶, fake news detection³⁷, traffic prediction³⁸, and recommendation systems³⁹. Within the context of robotic tactile perception, however, there remains a significant opportunity to enhance the processing and utilization of tactile signals.

Our approach, Tactile-GAT, introduces a novel application of Graph Attention Networks (GAT)⁴⁰ specifically designed to address these challenges. By integrating explicit relationship graphs, Tactile-GAT captures the physical connectivity between tactile sensors, enhancing the model's ability to interpret complex multi-channel tactile data. Furthermore, it learns latent relationships between sensors during model training, which significantly improves feature extraction and learning from tactile signals. This method not only diverges from traditional reliance on visual data in robotic systems but also offers a more refined analysis of tactile information, which is crucial for tasks requiring high precision and adaptability in dynamic environments. Through Tactile-GAT, we aim to bridge the existing gap in efficiently utilizing spatial information from tactile sensors, providing a robust framework that enhances the autonomy and decision-making capabilities of robots in real-world scenarios.

Dataset

Hardware setup

In our experiments, we used a hardware configuration comprising sixteen FSR402 force-sensitive sensors and six 2.2" flex sensors, as shown in Fig. 2. The FSR402 force-sensitive sensor is a circular flexible thin film pressure sensor with short legs, consisting of two layers that are glued together with double-sided tape. When external pressure is applied to the active area, the disconnected circuit of the lower layer is connected through the pressure-sensitive layer of the upper layer, thereby converting pressure into resistance. The output resistance decreases as pressure increases, and its pressure measuring range is from 20 g to 6 kg. One side of the flex sensor is printed with a polymer ink containing conductive particles. When the sensor is straight, the particles give the ink a resistance of about 30k Ohms. When the sensor is bent away from the ink, the conductive particles move further apart, increasing this resistance.

By combining the force and flex sensors with a static resistor to create a voltage divider, a variable voltage that can be read by a microcontroller's analog-to-digital converter (ESP32) is produced. The ESP32's ADC has a 12-bit resolution, enabling it to convert the analog signals into digital values ranging from 0 to 4095, where 0 represents 0 volts and 4095 represents the maximum input voltage of 3.3 volts. This digital conversion is crucial as it facilitates the precise mapping of pressure and bending degrees into a format that can be processed for further analysis and application.

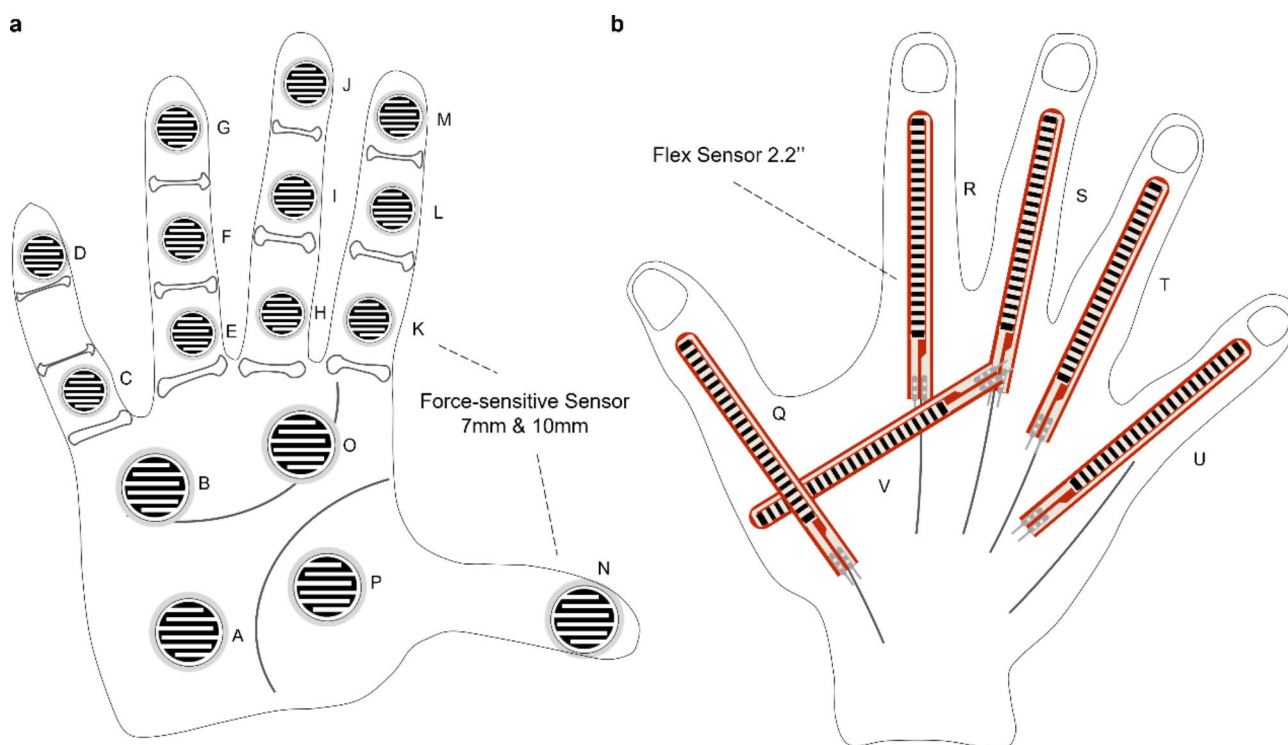


Fig. 2. The position map of pressure and flex sensors on the tactile glove. (a) 16 pressure sensors (A–N) on the front side. (b) 6 flex sensors (Q–V) on the back side.

These FSR402 sensors were installed on the palm and fingers of the glove, thoroughly recording the pressure changes on the hand contact surface when grasping objects. Simultaneously, the 2.2" Flex sensors installed on the fingers can measure bending angles from 0 degrees to nearly 180 degrees, accurately gauging the bending as the fingers move.

The physical and circuit diagram of the tactile glove are shown in Extended Data Fig. 1. This comprehensive setup not only captures detailed tactile information but also ensures that data is reliably processed and utilized in applications requiring high precision and responsiveness.

Data collection

Using the tactile glove, we recorded the pressure distribution and curvature degree of the hand when grasping 19 different types of objects, which are shown in Extended Data Fig. 2. The tactile information of each object consists of a time series of 22 channels, including 16 pressure signals and 6 flex signals. We collected tactile information from objects by touching and holding them. To ensure the validity of the collected information, the object was grasped with just enough strength to hold it without slipping and in different poses. The process of collecting tactile signals from objects is shown in Fig. 3. Figure 3 includes two phases of object tactile signal acquisition (touching and grasping) and three actions. The blue and green lines represent the average of the normalized acquisition voltage of the 16-channel pressure sensors and 6-channel bend sensors, respectively. Compared to traditional three-channel haptic datasets, this dataset contains a 22-channel time series and includes two types of tactile information instead of just a single friction signal. The data sampling frequency is about 220 Hz, which means that the microcomputer records tactile information of 22 channels 220 times per second. In the foreseeable future, as the price and volume of tactile sensors decrease, more kinds of tactile sensors will be embedded in robots. And the immediacy of feedback also puts forward higher requirements on the sampling frequency of tactile signals. This dataset meets the characteristics of high sampling rate and multiple types of future haptic signals. We will verify the performance of our tactile perception algorithm based on this dataset.

Data processing and preparation

Eliminating Outliers: During the tactile data collection process, outliers may be introduced due to sensor malfunctions, operational errors, or external disturbances. To ensure data quality, we first eliminate values that exceed thresholds set based on the physical characteristics of the sensors and experimental experience. This step helps maintain the accuracy and reliability of our data. **Data Normalization:** Given the different ranges and sensitivities of the pressure and bend sensors, we performed normalization on these sensor data. Normalization is a key step in processing data from sensors with varying ranges and sensitivities, enhancing model training efficiency and performance. **Time Window Segmentation:** To transform continuous tactile signals into discrete samples suitable for model training, it is necessary to segment the data stream into fixed-size time windows. In our experiment, we defined the window length as 400 data points, with each new window overlapping 200

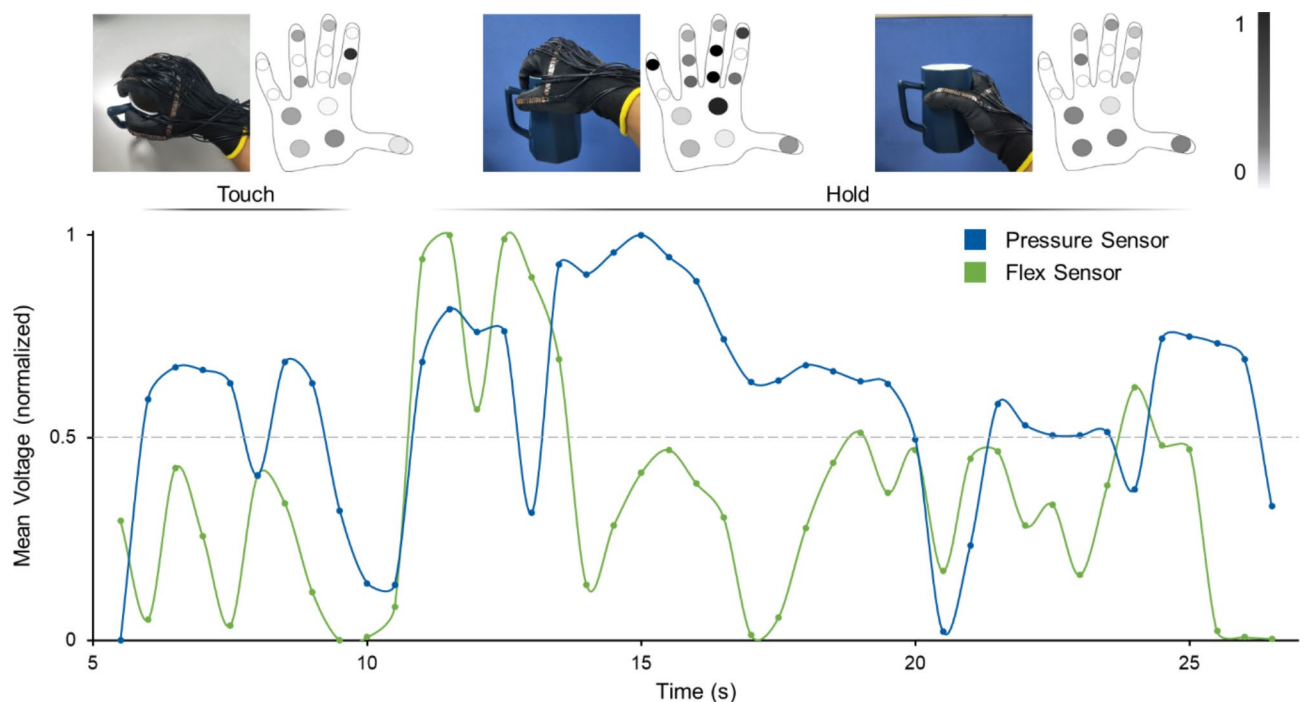


Fig. 3. The process of tactile data acquisition and examples of tactile signals. The figure includes two phases of object tactile signal acquisition (touching and grasping) and three actions. The blue and green lines represent the average of the normalized acquisition voltage of the 16-channel pressure sensors and 6-channel bend sensors, respectively.

points from the previous window. This overlap ensures the continuity of data features, while also increasing the volume of data and enhancing temporal resolution. **Training and Test Set Split:** To evaluate model performance and ensure the model can generalize to unseen data, we adopted a data split strategy for the training, validation, and test sets in the ratios of 70%, 15%, and 15%, respectively.

Tactile graph attention network

Given one of sensors (i.e., multivariate time series) data from N sensors of previous K timestamps $x_i = [x_i^{t-K}, \dots, x_i^{t-1}]$, $i \in \{1, 2, \dots, N\}$. $X = [x_1, x_2, \dots, x_{N-1}, x_N] \in \mathbb{R}^{N \times T}$ stands for the multivariate time-series input. In order to emphasize explicit relationships and uncover the latent relationships among multiple sensor time-series, we build explicit relationship graph \mathcal{G} and latent relationship graph \mathcal{G}' . $W \in \mathbb{R}^{N \times N}$ is the weight matrix of graph \mathcal{G}' , where $w_{ij} > 0$ indicates that there is an edge connecting nodes i and j , and the value of w_{ij} indicates the weight of this edge.

Our goal is to predict the correct label Y and recognize object in tactile signals X . The label Y can be inferred by the forecasting model F with parameter Φ and graph structures \mathcal{G} and \mathcal{G}' , where \mathcal{G}' can be input as prior or automatically inferred from data.

$$Y = F(X; \mathcal{G}; \mathcal{G}'; \Phi) \quad (1)$$

Figure 4 provides an overview of the Tactile-GAT model architecture, which consists of three main parts: the prior explicit relation graph of tactile signal channels, the latent relation graph learned through the embedding layer during training, and the graph attention networks for feature extraction.

Embedding Layer: one-dimensional embedding layer, used to learn the latent correlations between different tactile signals, and based on the correlations learned by the model, to construct a potential adjacency matrix graph among multi-channel tactile signals.

Explicit Relation Graph: We input a graph \mathcal{G} as prior based on distance and position from N pressure and flex sensors. When grasping an object, the pressure sensors on the same finger are associated, and as the bending degree of the finger changes, the pressure will change accordingly. In explicit relation graph \mathcal{G} , we connect adjacent sensors. The trained model F can informed the physical relationships \mathcal{G} among different sensor in advance and the information of adjacent channel will be passed and aggregated.

Latent Relation Graph: Latent relation graph \mathcal{G}' aims to learn the latent relationships between different sensors. To automatically infer weight matrix W of \mathcal{G}' from data X , we introduce an embedding vector for each sensor to represent its characteristics: $v_i \in \mathbb{R}^d$, for $i \in \{1, 2, \dots, N\}$, d indicates the dimension of each vector. Embedding vector is initialized randomly and updated when the model is trained. We calculate the similarity s_{ij} , the normalized dot product between the embedding vectors of sensor i and other sensors j . The large the value of s_{ij} , the more similar the embedding vectors are. A is the adjacency matrix of latent relation graph \mathcal{G}' , $A_{ij} = 1$ means that node i and node j are connected. For each node i , we select K nodes as its neighboring nodes based on the values of s_{ij} . TopK denotes top-k values among its input set.

$$i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, N\} \setminus \{i\} \quad (2)$$

$$s_{ij} = \frac{v_i^T v_j}{\|v_i\| \cdot \|v_j\|} \quad (3)$$

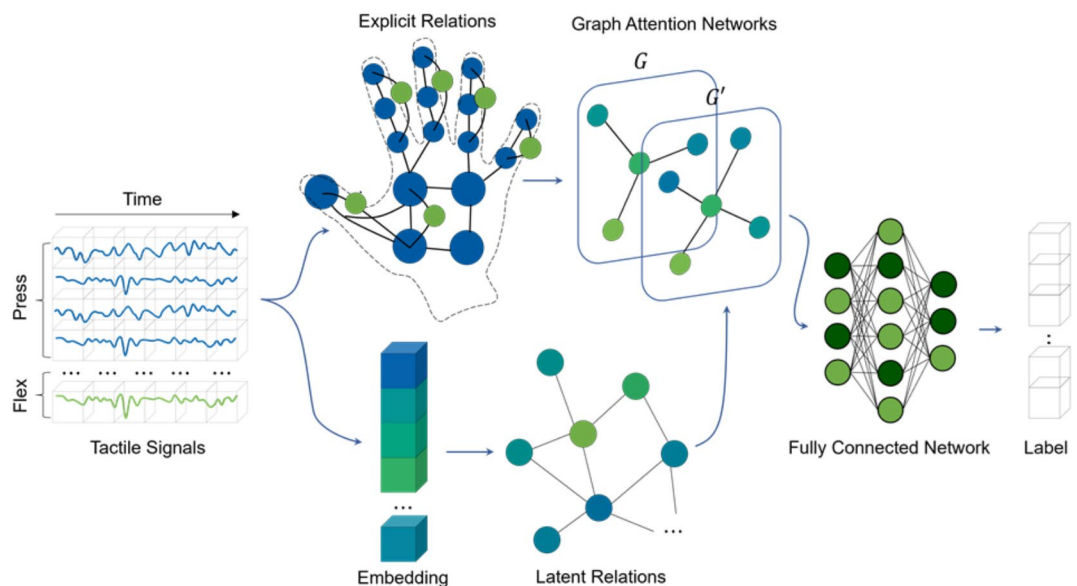


Fig. 4. Proposed framework (Tactile-GAT) for tactile signals classification.

$$A_{ij} = 1 \text{ if } s_{ij} \in \text{TopK}(s_{i1}, s_{i2}, \dots, s_{iN}) \quad (4)$$

Graph Attention Network: Graph Attention Network (GAT) serves as a pivotal tool in the realm of graph convolutional networks due to its innovative attention mechanism. Unlike traditional graph convolution methods that treat all neighbors equally, GAT introduces an attention mechanism that dynamically learns the importance of each neighbor node during aggregation. This adaptive weighting of neighbors allows GAT to effectively handle scenarios with noisy or less relevant neighbor nodes, thereby enhancing the robustness of the model. Moreover, the attention mechanism provides interpretability by highlighting which neighbors contribute most significantly to each node's representation, offering insights into the underlying relationships within the graph structure.

For explicit relation graph \mathcal{G} and latent relation graph \mathcal{G}' , W is a trainable weight matrix to represent the relationship of features. And the attention coefficients α_{ij} are computed as:³⁸

$$e_{ij} = a(Wx_i, Wx_j) = \text{LeakyReLU}(a^T[Wx_i || Wx_j]) \quad (5)$$

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{j \in \mathcal{N}_i} \exp(e_{ij})} \quad (6)$$

Where $a(\cdot)$ is a function to express the importance of node i to node j . $||$ denotes concatenation; $Wx_i || Wx_j$ concatenates the transformed features. a^T is the weight matrix connecting the layers in the neural network to each other, and $\text{LeakyReLU}(\cdot)$ function is also added to the output layer. Then $\text{softmax}(\cdot)$ normalizes e_{ij} to be attention coefficients α_{ij} , and the sum of all α_{ij} is 1. \mathcal{N}_i denotes the neighbors of node i . Combining equations above, the complete attention mechanism can be put together as follows³⁸.

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T[Wx_i || Wx_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T[Wx_i || Wx_k]))} \quad (7)$$

The normalized attention coefficients between different nodes are used to predict the output characteristics of each node x'_i . The updated information from all nodes is passed into a fully connected neural network f with parameter θ to predict the label Y of the grasping object^{38,39}.

$$x'_i = \text{ReLU}\left(\alpha_{ij}Wx_j + \sum_{j \in \mathcal{N}_i} \alpha_{ij}Wx_j\right) \quad (8)$$

$$Y = f_\theta(\hat{X}) = f_\theta([x'_1, \dots, x'_{N-1}, x'_N]) \quad (9)$$

Graph attention networks utilize explicit relation graphs and latent relation graphs to perform weighted aggregation of multi-channel tactile signal features, thereby updating the feature representation of each node.

Fully Connected Network: After the tactile node features are extracted through the graph attention mechanism, they are ultimately classified using a fully connected network.

Experiments and results

Baselines

In our study, we compare the performance of our proposed method with 12 time series classification methods spanning various neural network architectures. Each model is briefly described along with its primary applications and how it differs from our method.

CNNs-based NNs: Fully Convolutional Network (FCN, Long, 2015)⁴⁰: Originally designed for semantic segmentation, FCN's ability to process entire images in one go makes it distinct for segmentation tasks but less so for sequence classification where temporal dynamics are crucial⁴¹; Residual Network (ResNet, He, 2015)⁴²: Known for its deep architecture enabled by residual connections that combat vanishing gradients, ResNet is primarily used in classification and object detection, providing a strong baseline for performance comparison in feature learning⁴¹; An Explainable Convolutional Neural Network (XCM, Fauvel, 2021)⁴³: Focused on transparency, XCM offers insights into convolutional networks' decision-making process, contrasting with typical black-box approaches⁴⁴; ResCNN (Zou, 2019)⁴⁵: Enhances traditional CNNs by integrating residual learning, improving training stability and accuracy in both classification and detection scenarios⁴⁶; InceptionTime (Fawaz, 2019)^{47,48} (Rahimian, 2019): Combines Inception modules to better handle time series data by capturing complex features at various scales, providing a robust comparison for temporal data analysis⁴⁹; Omni-Scale 1D-CNN (OmniScale, Tang, 2020)⁵⁰: Adapts convolutional approaches to capture features across multiple scales in a single-dimensional input, such as time series, enhancing detection and classification capabilities⁵¹.

RNN-CNNs-based NNs: Long Short-Term Memory Fully Convolutional Network (LSTM-FCN, Karim, 2017)⁵²: Merges LSTM's temporal processing abilities with FCN's spatial feature extraction, making it suitable for tasks where both time and space dimensions are key⁵³; Gated Recurrent Unit Fully Convolutional Network (GRU-FCN, Elsayed, 2018)⁵⁴: Utilizes GRU for handling temporal dependencies, combined with FCN for efficient spatial processing, ideal for sequential data that requires contextual understanding over time⁵⁵.

Wavelet-based NNs: Multilevel wavelet decomposition network (mWDN, Wang, 2018)⁵⁶: Employs wavelet analysis to decompose time series data, facilitating feature extraction at multiple resolutions, advantageous for classification tasks involving non-linear and non-stationary signals.

MLP-base NNs: Gated Multilayer Perceptron (gMLP, Liu, 2021)⁵⁷ : Integrates gating mechanisms to control information flow within MLP layers, enhancing the network’s ability to focus on relevant features for classification.

Each model showcases unique strengths in handling time series data. Our proposed method primarily evaluates different multi-channel time-series classification models based on their performance in classifying multi-category tactile signals. This comparative analysis not only demonstrates how our model excels in specific application areas but also highlights its practical value in processing complex tactile data tasks. Through this approach, we can more accurately assess and showcase the performance of different models in handling complex tactile signals.

Experimental setup

Model Architecture: Our model is designed to classify multi-category tactile signals effectively. It begins with an embedding layer that transforms the 22-dimensional input features into a 64-dimensional embedding space, facilitating higher-level feature representation. A batch normalization layer with 128 features follows to stabilize and accelerate the training process by normalizing inputs.

We employ two Graph Neural Network (GNN) layers to capture the relationships between different tactile inputs. Each GNN layer includes a graph convolutional layer with 200 input features and 64 output features, utilizing a single attention head. Batch normalization and activation functions (ReLU and LeakyReLU) are incorporated within each GNN layer to introduce non-linearity and maintain stable input distributions.

An output layer with a linear transformation maps the learned features to the final output. To prevent overfitting, a dropout layer with a probability of 0.2 is included. The architecture concludes with a final linear layer that projects the input features into a 20-dimensional output space corresponding to the number of classes.

Training Configuration: The model is trained using the Adam optimizer with a learning rate of 0.01 and a weight decay parameter specified in our configuration settings. Training is conducted over 50 epochs with a batch size of 256, balancing computational efficiency and model convergence. The cross-entropy loss function is used to measure the discrepancy between predicted and true class labels.

Performance comparison

The 12 baseline methods we compared in this study employed different network structures and algorithms for time series classification. Among them, the CNN-based methods mainly utilized convolutional neural networks to extract the feature information from the time series, while RNN-CNN-based methods combined the advantages of recurrent neural networks and convolutional neural networks to better capture the temporal features of the time series. The wavelet-based methods used wavelet decomposition for preprocessing the time series to obtain more representative features. MLP-based methods utilized a multilayer perceptron model for classification, which is advantageous for its simplicity and ease of implementation.

Our experimental results significantly demonstrate the effectiveness of our proposed method. On various tactile datasets, our Tactile-GAT model achieved an accuracy of 89.58%, with precision, recall, and F1 scores of 89.20%, 89.36%, and 0.8939, respectively, all markedly superior to other baseline methods. These evaluation results are detailed in Table 1, and the confusion matrix for Tactile-GAT is presented in Fig. 5.

Overall, Tactile-GAT exhibits superior performance in the task of classifying temporal tactile signals compared to existing time-series classification algorithms. Particularly, it excels with only 0.11 M in model parameters, demonstrating outstanding efficiency and effectiveness. The reduced number of model parameters not only indicates a simpler structure but also implies that our model can achieve faster training and inference speeds, as well as better generalization capabilities in practical applications. This makes Tactile-GAT especially

Method	Accuracy	Precision	Recall	F1 Score	Params
MLP	70.12	69.13	70.41	0.7941	20.51
XCM	76.34	76.67	75.87	0.7560	10.29
gMLP	79.94	80.20	79.93	0.7994	2.67
GRU-FCN	85.89	85.72	85.89	0.8574	0.54
xLSTM-FCN	86.75	86.66	86.76	0.8665	0.57
GCN	86.87	86.53	86.43	0.8648	0.36
mWDN	86.90	86.72	86.36	0.8650	0.60
FCN	87.20	86.60	87.21	0.8683	0.44
InceptionTime	87.49	87.55	87.58	0.8756	0.49
OmniScaleCNN	87.65	87.93	87.68	0.8777	0.39
XceptionTime	88.10	88.18	88.20	0.8819	0.43
xresnet1d34	88.55	88.61	88.64	0.8862	7.26
ResCNN	88.55	88.63	88.65	0.8864	0.36
ResNet	88.70	88.73	88.78	0.8875	0.58
Tactile-GAT (ours)	89.58*	89.20*	89.36*	0.8939*	0.11*

Table 1. The classification accuracy (%), macro precision (%), macro recall (%), macro F1 score and model parameter count (M) of various models on the tactile datasets.

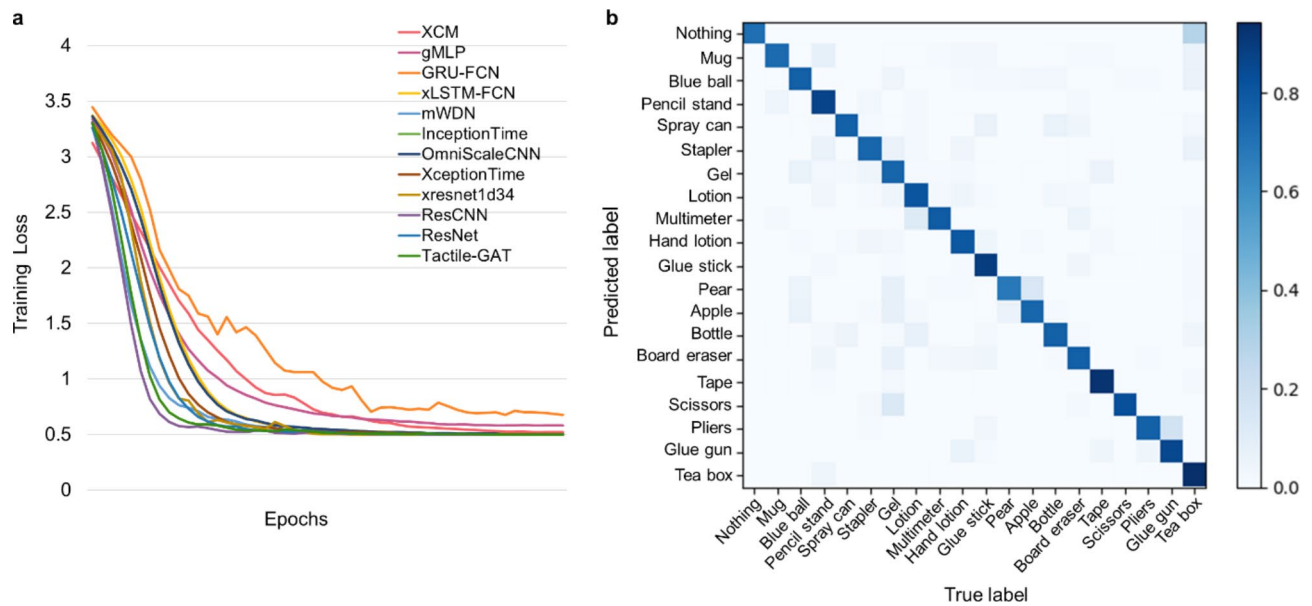


Fig. 5. Training loss and confusion maps of Tactile-GAT in tactile classification task.

suitable for resource-constrained devices, while also reducing the risk of overfitting, providing an efficient and reliable solution for tactile signal processing.

Notably, fewer parameters may limit the model's capability to express complex functionalities, potentially hindering the model's ability to capture all complex patterns and relationships in dynamic or complex datasets. Although models with fewer parameters are generally less prone to overfitting, if the model is overly simplistic, its generalization capability could be compromised, as it might not have learned enough to adapt to new, unseen data. Moreover, models with fewer parameters might be more sensitive to noise in the input data due to their limited learning and adaptation capabilities. Therefore, while preserving the key structure of the Tactile-GAT model, appropriately increasing the model's complexity could help enhance its performance in classifying different types of complex multi-channel tactile signals, thus providing a more efficient and reliable solution for tactile signal processing.

To further validate the effectiveness of Tactile-GAT, we conducted a significance analysis using paired t-tests to compare Tactile-GAT with other time-series classification algorithms, as detailed in Extended Data Table 1. The results showed that Tactile-GAT exhibited high significance ($p < 0.001$) across multiple comparative methods, particularly when compared to multilayer perceptrons (MLP) and other complex models (such as XCM and gMLP). This indicates that Tactile-GAT significantly outperforms existing algorithms in classification performance. This analysis not only reinforces the advantages of Tactile-GAT but also provides a reference for future improvements in model complexity while maintaining efficiency, enabling better handling of complex multi-channel tactile signal classification tasks.

Ablation analysis

To better understand the effectiveness of different components in Tactile-GAT, we designed two model variants and conducted an ablation study on pressure signal only, flex signal only, and both tactile signal datasets. The results are summarized in Table 2.

All three models demonstrated good performance on the dataset containing only pressure signals. This is mainly because the pressure signals from 16 channels provide more comprehensive tactile information than the bending signals from just 6 channels. Pressure signals can capture fine differences in the contact between the object's surface and the sensor, which are details that bending signals struggle to provide.

Conversely, the performance of the models on the dataset containing only bending signals was poor. This shortfall is due to bending sensors primarily measuring the degree of finger bending, which is not sensitive enough to changes in weight. This makes it difficult to distinguish between objects with similar sizes or shapes, such as apples and pears, using bending data alone. However, pressure sensors can effectively differentiate objects based on variations in the force exerted during grasping, even if the objects are similar in size and shape.

When both pressure and bending signals were used for classification, all three models significantly outperformed those that used only one type of signal. This enhanced performance underscores that bending signals provide supplementary information not captured by pressure signals alone, thereby enriching tactile feedback and improving classification accuracy. For instance, while pressure signals provide information about object hardness and weight, bending signals add valuable data about object shape and the hand's grasping posture. This multimodal sensory approach more closely mimics human tactile perception and object recognition, thereby increasing the accuracy of classification tasks.

Among the three models, the ones using both explicit and latent graph module ($\mathcal{G} + \mathcal{G}' + GAT$) exhibited better classification performance than the model that used a single graph. This result indicates that incorporating

Method	Tactile signal		Accuracy
	Press	Flex	
$\mathcal{G} + GAT$	✓		86.89
		✓	24.66
	✓	✓	88.63
$\mathcal{G}' + GAT$	✓		85.17
		✓	48.24
	✓	✓	88.46
$\mathcal{G} + \mathcal{G}' + GAT$ (Tactile-GAT)	✓		87.67
		✓	45.02
	✓	✓	89.58

Table 2. The results of ablation experiments of the Tactile-GAT model on different tactile datasets.

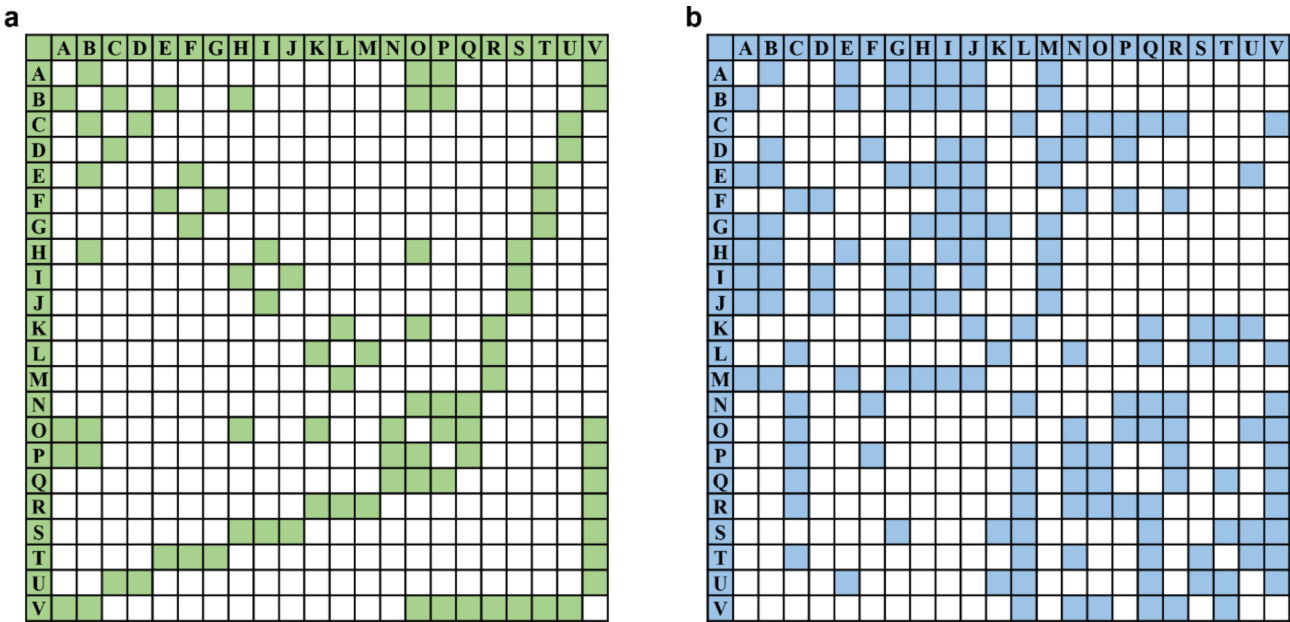


Fig. 6. Adjacency map in explicit and latent relation graphs, where A-V represent individual sensors. **(a)** The adjacency map of explicit relation graph \mathcal{G} . **(b)** the adjacency map of latent relation graph \mathcal{G}' .

both explicit relation graphs (\mathcal{G}) and latent relation graphs (\mathcal{G}') in the models, by informing the model of the physical adjacency between sensors and allowing it to learn potential relations between different sensors, can effectively improve the accuracy of multi-channel, multi-category tactile signal classification, making full use of the comprehensive tactile information provided by combining pressure and flex signals.

Tactile relation graph

Figure 6a shows the physical connectivity of the pressure and flex sensors on the tactile glove. This is also the adjacency matrix of the explicit relation graph in the Tactile-GAT model. Figure 6b displays the adjacency matrix of the latent relation graph learned by the model during training, which represents the underlying relationships between the signal channels that the model has learned.

By comparing the two graphs, we can see that the latent graph is a complement to the explicit graph. It can automatically learn the connections between channel signals and infer these connections through the intrinsic structure, similarity, and other features of the data, rather than being directly given. Additionally, the latent relation graph can also remove channel connections that are physically connected but have small correlations or redundant information through training.

The latent graph is a directed graph that can provide directionality for the reference, dependency, and control relationships between tactile signal channels. For example, in the graph, there is a connection form the pressure sensor D at the tip of the little finger to the pressure sensor N on the thumb, but not the other way around. This may be because the use of the thumb is not always necessary when grasping an object, but the use of the little finger often requires the use of the thumb as well.

Discussion and conclusion

Discussion

This study introduces a graph-based framework for tactile perception, Tactile-GAT, aimed at enhancing robots' ability to process and understand multi-channel tactile information. We developed a tactile glove equipped with pressure and bending sensors and constructed a 22-channel dataset to validate the effectiveness of the Tactile-GAT framework. Compared to existing multi-channel temporal signal algorithms, our method not only shows superior performance in tactile signal classification tasks but also operates with fewer parameters, making it more suitable for practical applications.

The results demonstrate that the Tactile-GAT framework can effectively utilize spatial relationships between multi-channel tactile signals, significantly enhancing robots' tactile perception capabilities. This improvement is critical for applications requiring real-time tactile feedback, such as robotic grasping and manipulation tasks in unstructured environments.

Despite its promising performance, Tactile-GAT has some limitations. Due to sensor size and material constraints, we currently integrate only a limited set of sensors in the tactile glove, which restricts the model's application potential in more complex environments. Additionally, the fewer parameters, while beneficial for avoiding overfitting and running on constrained devices, might limit the model's ability to capture all complex patterns in highly dynamic datasets.

Future directions

Enhanced Sensor Integration: As Tactile-GAT continues to be validated and applied successfully, a significant area for future research is the integration of a broader array of sensors into the tactile glove. Currently, the glove incorporates basic pressure and bending sensors. Looking ahead, we plan to incorporate additional types of sensors, such as temperature, humidity, and vibration sensors, to provide more comprehensive tactile feedback. This multimodal sensor integration will enable the model not only to perceive the hardness and shape of objects but also to detect more complex properties like temperature and humidity, enhancing the glove's sensory capabilities.

Cross-Modal Learning: Another promising direction for future research involves cross-modal learning. By integrating data from visual, auditory, and tactile sensors, a more comprehensive perceptual system can be developed. For example, integrating visual and tactile data could lead to more accurate identification of material properties or changes in an object's state, thereby improving the robot's ability to adapt to complex environments. This approach would leverage the strengths of each sensory modality to provide a richer, more robust understanding of the surroundings.

Real-World Application Testing: We also aim to focus on deploying the Tactile-GAT framework in real robotic systems and conducting necessary algorithm optimizations to ensure real-time performance. This includes simplifying and accelerating the framework to ensure it operates effectively on resource-constrained devices while maintaining high accuracy and responsiveness. By conducting deployment tests in real environments, we can not only validate the performance and stability of the model but also compare it with existing tactile processing technologies to accurately assess the advantages and potential applications of Tactile-GAT.

Conclusion

The proposed Tactile-GAT framework represents an innovative approach for processing tactile signals in robotics. By leveraging graph-based attention networks, our method demonstrates improved accuracy in tactile classification tasks, with an approximate 1.0% increase over existing state-of-the-art models such as ResNet. Tactile-GAT's capability to effectively integrate multi-channel tactile information allows robots to perceive and respond to their environment more accurately and efficiently.

As robotics technology continues to evolve, we anticipate that Tactile-GAT will significantly enhance the precision and efficiency of robotic operations across diverse practical applications, providing new methodologies and insights into the field of robot tactile perception.

Data availability

The datasets used in this study are available upon request from the corresponding author (chenl103@chinatelecom.cn). All code for data cleaning and analysis associated with the current submission will be available at <https://github.com/alanchancl/Tactile-GAT>. Any updates will also be published on GitHub.

Received: 23 August 2024; Accepted: 4 November 2024

Published online: 11 November 2024

References

- Li, Y. et al. Optical-Waveguide based Tactile sensing for Surgical instruments of minimally invasive surgery. *Front. Robot. AI*. **8**, 773166. <https://doi.org/10.3389/frobt.2021.773166> (2022).
- Wu, Y. et al. A skin-inspired tactile sensor for smart prosthetics. *Sci. Robot.* (2018).
- Palli, G. & Pirozzi, S. A Tactile-based wire manipulation system for manufacturing applications. *Robotics*, vol. 8, no. 2, Art. no. 2. (2019). <https://doi.org/10.3390/robotics8020046>
- Zhao, Y. J., Wen, C., Zhang, Y. D. & Zhang, H. Needle tip pose estimation for ultrasound-guided steerable flexible needle with a complicated trajectory in soft tissue. *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 11705–11712. (2022). <https://doi.org/10.1109/LRA.2022.3196465>
- Zhang, Z. et al. Hardness recognition of fruits and vegetables based on tactile array information of manipulator. *Comput. Electron. Agric.* **181**, 105959. <https://doi.org/10.1016/j.compag.2020.105959> (Feb. 2021).
- Liu, Y. et al. Centrosymmetric- and axisymmetric-patterned flexible Tactile Sensor for Roughness and Slip Intelligent Recognition. *Adv. Intell. Syst.* **4**(1), 2100072. <https://doi.org/10.1002/aisy.202100072> (2022).

7. Yi, Z., Xu, T., Guo, S., Shang, W. & Wu, X. Tactile surface roughness categorization with multineuron spike train distance, *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1835–1845, Oct. (2021). <https://doi.org/10.1109/TASE.2020.3021742>
8. Willemet, L. *The Biomechanics of the Tactile Perception of Friction*. In *Springer Series on Touch and Haptic Systems* (Springer, 2022). <https://doi.org/10.1007/978-3-031-16053-0>
9. Deepshikha, K., Yelleni, S. H., Sriji, P. K. & Mohan, C. K. Monte Carlo dropout for modelling uncertainty in object detection, Aug. 08, *arXiv: arXiv:2108.03614*. Accessed: Apr. 22, 2024. [Online]. Available: (2021). <http://arxiv.org/abs/2108.03614>
10. Sunnetci, K. M. & Alkan, A. Biphasic majority voting-based comparative COVID-19 diagnosis using chest X-ray images. *Expert Syst. Appl.* **216**, 119430. <https://doi.org/10.1016/j.eswa.2022.119430> (2023).
11. Sunnetci, K. M., Kaba, E., Celiker, F. B. & Alkan, A. Deep network-based comprehensive parotid gland tumor detection. *Acad. Radiol.* **31**(1), 157–167. <https://doi.org/10.1016/j.acra.2023.04.028> (2024).
12. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition, *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324 (1998). <https://doi.org/10.1109/5.726791>
13. Lee, H. et al. Jan., Predicting the force map of an ERT-based tactile sensor using simulation and deep networks. *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 1, pp. 425–439. (2023). <https://doi.org/10.1109/TASE.2022.3156184>
14. Liu, H., Yu, Y., Sun, F. & Gu, J. Visual–tactile fusion for object recognition. *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 996–1008. (2017). <https://doi.org/10.1109/TASE.2016.2549552>
15. Ren, S., He, K., Girshick, R., Sun, J. & Faster, R-C-N-N. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149. (2017). <https://doi.org/10.1109/TPAMI.2016.2577031>
16. Redmon, J. & Farhadi, A. YOLOv3: An Incremental Improvement.
17. Liu, W. et al. SSD: Single shot MultiBox detector. In *Computer Vision – ECCV 2016*, vol. 9905, (eds Leibe, B., Matas, J., Sebe, N. & Welling, M.) in *Lecture Notes in Computer Science*, vol. 9905., Cham: Springer International Publishing, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2. (2016).
18. Gu, Y., Zhu, B. & Du, Y. Robot Grasp Detection using Inverted Residual Convolutional Neural Network, In Review, preprint. (2022). <https://doi.org/10.21203/rs.3.rs-2099144/v1>
19. Kumra, S., Joshi, S. & Sahin, F. Antipodal robotic grasping using generative residual convolutional neural network. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 9626–9633. (2020). <https://doi.org/10.1109/IROS45743.2020.9340777>
20. Purri, M. & Dana, K. Teaching cameras to feel: Estimating tactile physical properties of surfaces from images, in *Computer Vision – ECCV 2020*, vol. 12372, (eds Vedaldi, A., Bischof, H., Brox, T. & Frahm, J. M.) in *Lecture Notes in Computer Science*, vol. 12372., Cham: Springer International Publishing, 1–20. https://doi.org/10.1007/978-3-030-58583-9_1. (2020).
21. Li, R. et al. Localization and manipulation of small parts using GelSight tactile sensing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2014, pp. 3988–3993. doi: (2014). <https://doi.org/10.1109/IROS.2014.6943123>
22. Yuan, W., Mo, Y., Wang, S. & Adelson, E. H. Active clothing material perception using tactile sensing and deep learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 4842–4849. (2018). <https://doi.org/10.1109/ICRA.2018.8461164>
23. Li, Y., Zhu, J. Y., Tedrake, R. & Torralba, A. Connecting touch and vision via cross-modal prediction, Jun. 14, *arXiv: arXiv:1906.06322*. (2019). <https://doi.org/10.48550/arXiv.1906.06322>
24. Yuan, W., Wang, S., Dong, S. & Adelson, E. Connecting look and feel: Associating the visual and tactile properties of physical materials. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 4494–4502. doi: (2017). <https://doi.org/10.1109/CVPR.2017.478>
25. Kerzel, M., Ali, M., Ng, H. G. & Wermter, S. Haptic material classification with a multi-channel neural network. In *International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 439–446. (2017). <https://doi.org/10.1109/IJCNN.2017.7965887>
26. Falco, P. et al. Cross-modal visuo-tactile object recognition using robotic active exploration. In *IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 5273–5280. (2017). <https://doi.org/10.1109/ICRA.2017.7989619>
27. Zhang, X. et al. Target classification method of tactile perception data with deep learning. *Entropy*. **23**(11), 1537. <https://doi.org/10.3390/e23111537> (Nov. 2021).
28. Bimbo, J., Luo, S., Althoefer, K. & Liu, H. In-Hand object pose estimation using covariance-based tactile to geometry matching. *IEEE Robot Autom. Lett.* **1**(1), 570–577. <https://doi.org/10.1109/LRA.2016.2517244> (Jan. 2016).
29. Yan, Y., Hu, Z., Shen, Y. & Pan, J. Surface texture recognition by deep learning-enhanced tactile sensing. *Adv. Intell. Syst.* **4**(1), 2100076. <https://doi.org/10.1002/aisy.202100076> (2022).
30. Lambeta, M. et al. Jul., DIGIT: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3838–3845. (2020). <https://doi.org/10.1109/LRA.2020.2977257>
31. Sundaram, S. et al. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*. **569**(7758), 698–702. <https://doi.org/10.1038/s41586-019-1234-z> (May 2019).
32. Lu, Z. & Yu, H. GTac-Hand: A robotic hand with integrated tactile sensing and extrinsic contact sensing capabilities. *IEEEASME Trans. Mechatron.*, vol. 28, no. 5, pp. 2919–2929. (2023). <https://doi.org/10.1109/TMECH.2023.3264650>
33. Funabashi, S. et al. Multi-fingered in-hand manipulation with various object properties using graph convolutional networks and distributed tactile sensors. *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2102–2109 (2022). <https://doi.org/10.1109/LRA.2022.3142417>
34. Rasouli, M., Chen, Y., Basu, A., Kukreja, S. L. & Thakor, N. V. An Extreme learning machine-based neuromorphic tactile sensing system for texture recognition. *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 2, pp. 313–325. (2018). <https://doi.org/10.1109/TBCAS.2018.2805721>
35. Calandra, R. et al. Oct., More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3300–3307. (2018). <https://doi.org/10.1109/LRA.2018.2852779>
36. Gao, Y., Hendricks, L. A., Kuchenbecker, K. J. & Darrell, T. Deep learning for tactile understanding from visual and haptic data. In *IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 536–543. (2016). <https://doi.org/10.1109/ICRA.2016.7487176>
37. Guo, D. et al. A Hybrid Deep Architecture for Robotic Grasp Detection.
38. Veličković, P. et al. Graph Attention Networks, Feb. 04, *arXiv: arXiv:1710.10903*. Accessed: Feb. 27, 2023. [Online]. Available: (2018). <http://arxiv.org/abs/1710.10903>
39. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks, Feb. 22, *arXiv: arXiv:1609.02907*. (2017). <https://doi.org/10.48550/arXiv.1609.02907>
40. Long, J., Shelhamer, E. & Darrell, T. Fully Convolutional networks for semantic segmentation.
41. Wang, Z., Yan, W. & Oates, T. Time series classification from scratch with deep neural networks: A strong baseline, Dec. 14, 2016, *arXiv: arXiv:1611.06455*. <https://doi.org/10.48550/arXiv.1611.06455>
42. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition, Dec. 10, *arXiv: arXiv:1512.03385*. Accessed: Apr. 22, 2024. [Online]. Available: (2015). <http://arxiv.org/abs/1512.03385>
43. Fauvel, K., Lin, T., Masson, V., Fromont, É. & Termier, A. XCM: An explainable convolutional neural network for multivariate time series classification. *Mathematics*, vol. 9, no. 23, p. 3137. (2021). <https://doi.org/10.3390/math923137>
44. Freeborough, W. & van Zyl, T. Investigating explainability methods in recurrent neural network architectures for financial time series data. *Appl. Sci.*, vol. 12, no. 3, Art. no. 3 (2022). <https://doi.org/10.3390/app12031427>

45. Zou, X., Wang, Z., Li, Q. & Sheng, W. Integration of residual network and convolutional neural network along with various activation functions and global pooling for time series classification, *Neurocomputing*, vol. 367, pp. 39–45. (2019). <https://doi.org/10.1016/j.neucom.2019.08.023>
46. Wen, T., Xiao, Y., Wang, A. & Wang, H. A novel hybrid feature fusion model for detecting phishing scam on Ethereum using deep neural network. *Expert Syst. Appl.* **211**, 118463. <https://doi.org/10.1016/j.eswa.2022.118463> (Jan. 2023).
47. Fawaz, H. I. et al. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **34**(6), 1936–1962. <https://doi.org/10.1007/s10618-020-00710-y> (2020).
48. Rahimian, E., Zabihi, S., Atashzar, S. F., Asif, A. & Mohammadi, A. XceptionTime: A novel deep architecture based on depthwise separable convolutions for hand gesture classification, Nov. 09, 2019, *arXiv: arXiv:1911.03803*. Accessed: Feb. 27, 2023. [Online]. Available: <http://arxiv.org/abs/1911.03803>
49. Li, Y., Yang, G., Su, Z., Li, S. & Wang, Y. Human activity recognition based on multienvironment sensor data, *Inf. Fusion*, vol. 91, pp. 47–63 (2023). <https://doi.org/10.1016/j.inffus.2022.10.015>
50. Tang, W. et al. Omni-Scale CNNs: A simple and effective kernel size configuration for time series classification, *arXiv.org*. Accessed: Feb. 27, 2023. [Online]. Available: <https://arxiv.org/abs/2002.10061v3>
51. Teng, X., Zhang, X. & Luo, Z. Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction. *Neurocomputing*. **505**, 92–100. <https://doi.org/10.1016/j.neucom.2022.07.016> (Sep. 2022).
52. Karim, F., Majumdar, S., Darabi, H. & Chen, S. LSTM fully convolutional networks for time series classification. *IEEE Access*. **6**, 1662–1669. <https://doi.org/10.1109/ACCESS.2017.2779939> (2018).
53. A multibranch. CNN-BiLSTM model for human activity recognition using wearable sensor data | The Visual Computer. Accessed: Apr. 22, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s00371-021-02283-3>
54. Elsayed, N., Maida, A. S. & Bayoumi, M. Deep gated recurrent and convolutional network hybrid model for univariate time series classification. *Int. J. Adv. Comput. Sci. Appl.* **10**(5). <https://doi.org/10.14569/IJACSA.2019.0100582> (2019).
55. Internet traffic prediction with deep. neural networks - Jiang - 2022 - Internet Technology Letters - Wiley Online Library. Accessed: Apr. 22, 2024. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1002/itl2.314>
56. Wang, J., Wang, Z., Li, J. & Wu, J. Multilevel wavelet decomposition network for interpretable time series analysis. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London United Kingdom: ACM, Jul. pp. 2437–2446. (2018). <https://doi.org/10.1145/3219819.3220060>
57. Liu, H., Dai, Z., So, D. R. & Le, Q. V. Pay Attention to MLPs, Jun. 01, *arXiv: arXiv:2105.08050*. Accessed: Feb. 27, 2023. [Online]. Available: (2021). <http://arxiv.org/abs/2105.08050>

Author contributions

The contributions of each author to this work have been as follows: Lun Chen: Conceived and designed the research, conducted experiments, analyzed data, and wrote the manuscript. Assisted in designing the research, contributed to data analysis, and reviewed and edited the manuscript. Provided technical expertise, assisted in designing experiments, and reviewed the manuscript. Yingzhao Zhu: Contributed to data collection, analysis, and interpretation, and reviewed the manuscript. Supervised the research, provided critical insights, and reviewed and edited the manuscript. Man Li: Contributed to, analysis, and reviewed the manuscript. Supervised the research, provided critical insights, and reviewed and edited the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-78764-x>.

Correspondence and requests for materials should be addressed to L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024