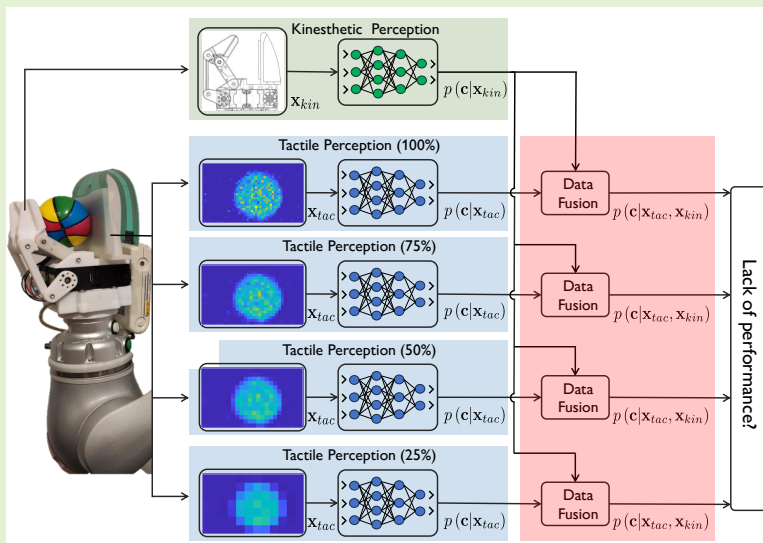


# Multimodal haptic object recognition: Can kinesthetic inference compensate for the lack of tactile sensing resolution?

Francisco Pastor, Da-hui Lin-Yang, Alfonso J. García-Cerezo, *IEEE Member*, and Juan M. Gandarias

**Abstract**—Haptic perception arises from the integration of cutaneous and kinesthetic cues, yet achieving human-level tactile performance in robotic systems remains technically demanding and economically costly. This raises a fundamental question: how much does spatial resolution truly matter for robotic touch, and can kinesthetic inference compensate when tactile resolution is limited? To investigate this, we conduct a study on haptic object recognition using an underactuated, sensorized gripper equipped with a high-resolution tactile array and joint angle sensors. We evaluate a dataset of 36 objects collected under a squeeze-and-release exploratory procedure (EP). Tactile sequences are downsampled to four resolutions (100%, 75%, 50%, and 25%) using bicubic interpolation, and ConvLSTM-based models are trained to quantify accuracy degradation as spatial detail decreases. We then use two fusion strategies from our previous work: Bayesian inference and neural-based inference, which combine tactile and kinesthetic modalities, and assess whether kinesthetic cues can offset the loss of tactile precision. The results quantify the performance drop as tactile resolution decreases and uncover promising evidence that multimodal fusion can partially recover this lost capability. Interestingly, the two fusion methods exhibit distinct behaviors in terms of robustness and consistency. As discussed comprehensively in the manuscript, these findings suggest that, under the right conditions, reliable object recognition may not depend solely on maintaining high tactile resolution, opening new possibilities for designing more scalable and cost-efficient haptic perception systems.

**Index Terms**—Deep learning in grasping and manipulation, haptic object recognition, multimodal haptic perception, tactile sensing.



## I. INTRODUCTION

HUMAN skin is the largest organ of the body and a vital interface with the external world, enabling touch sensations through cutaneous mechanoreceptors [1]. Thanks to the sense of touch, we can perceive pressure, vibration, temperature, and other tactile information. However, spatial tactile resolution is not uniformly distributed across the body: regions such as the fingertips and lips exhibit considerably higher resolution than areas such as the forearms or the back.

This research was funded by the University of Málaga, the Ministerio de Ciencia, Innovación y Universidades, Gobierno de España under grant PID2021-122944OB-I00.

Francisco Pastor, Da-hui Lin-Yang, Alfonso J. García-Cerezo, and Juan M. Gandarias are with the Robotics and Mechatronics Group, Institute for Mechatronics Engineering and Cyber-Physical Systems (IMECH.UMA), University of Málaga, 29071, Málaga, Spain (e-mail: fpastor@uma.es).

This heterogeneity is well known and has been historically assessed in several studies [2], [3].

Cutaneous information alone, however, may not be enough for many day-to-day tasks. Kinesthetic (proprioceptive) signals—arising from receptors in muscles, tendons, joints, and skin—encode limb position and effort and are essential for achieving skilled interaction and control. Haptic perception is conventionally defined as the integration of cutaneous and kinesthetic inputs within active exploration, which supports robust recognition of object properties [4].

The importance of touch sense extends to robotics as well. Endowing robots with similar tactile capabilities to humans has long been recognized as a key step toward dexterous and adaptive manipulation [5], [6]. As in humans, robots benefit from both “cutaneous” information—distributed tactile sensing on robotic artificial skins that contact the environment—

and “kinesthetic” information—joint position, velocity, and torque/force signals. Kinesthetic sensing has been standard practice in robotics since the field’s early days, typically via joint encoders and, more recently, torque/force measurement devices [7]. By contrast, large-area, high-resolution tactile skins (also known as e-skins) remain far less common, especially outside laboratories [8]. Despite decades of progress, broad deployment is hindered by complexity and cost, particularly when covering extended surfaces at high spatial resolutions, which pose challenges for sensor density, wiring, robustness, and real-time computation [9].

Prior work has shown that fusing tactile and kinesthetic information can considerably improve robotic perception [10]. In our previous work [11], we explored how the fusion of these two types of information improves performance in an object recognition task using a robotic gripper. Building on this line of research, we investigate two fundamental questions: i) How critical is high spatial tactile resolution for robust performance? ii) To what extent can kinesthetic information compensate for low tactile resolution? Providing universal answers to these questions is complex because outcomes depend on multiple factors: the types of tactile and kinesthetic sensors, the fusion algorithms, and the target application, among others. Nevertheless, to the best of our knowledge, no studies have quantified the performance trade-off between tactile spatial resolution and kinesthetic compensation in robotic object recognition.

Hence, this article presents an experimental study focused on a specific yet representative task: haptic object recognition using both tactile and kinesthetic information. Within this framework, the contributions of this work are the following:

- A quantification of the performance impact of varying tactile spatial resolution, including a comprehensive evaluation of the impact of tactile resolution on object recognition performance across a diverse dataset.
- An experimental evaluation on the capacity of kinesthetic information to recover performance when tactile resolution is deliberately reduced.
- A detailed discussion on the trade-off between tactile resolution and kinesthetic information fusion, aiming to provide empirical insights and to support further design choices for scalable, cost-effective tactile systems in robotics.

## II. RELATED WORKS

Over the past decade, tactile sensors have been successfully applied in a wide range of robotic domains. In healthcare, tactile sensing has been used in assistive and rehabilitation robots [12] and to support minimally invasive surgery [13]. In the food industry, tactile arrays have been used for a wide variety of applications, including non-destructive quality control, sorting, and gentle handling [14], [15]. In search-and-rescue scenarios, tactile-equipped robots have demonstrated the ability to perform human-robot physical interaction with increased safety [16]. Nevertheless, this work does not focus on any particular application but rather on the challenge of tactile object recognition, with a special emphasis on multimodal haptic perception approaches. Below is a series of

works related to these topics that help the reader understand this article in the context of the current state of the art.

### A. Object recognition with grippers

Robotic grippers have emerged as a key platform for in-hand object recognition, enabling the acquisition of multiple sources of information during manipulation. Several strategies have been proposed to identify objects through gripper interactions, ranging from static tactile measurements [17], [18] to dynamic exploratory procedures (EPs). Among these, the squeeze-and-release EP is widely adopted because it provides reliable cues about stiffness, internal characteristics, and shape while requiring simple actuation patterns [19]. Other works have used this procedure, applying controlled forces to the object with the gripper, capturing both pressure distributions and joint kinematics over time [11].

Different approaches have been explored to process tactile data collected by grippers. High-resolution tactile arrays often produce pressure maps that can be treated as images, enabling the use of computer vision techniques such as Convolutional neural networks (CNNs) for classification [20]. More recent works extend this concept to temporal sequences of tactile images, leveraging recurrent architectures like Long Short-Term Memory (LSTMs) neural networks or Convolute-LSTM neural networks to exploit dynamic patterns during palpation [21]. Similarly, some studies consider video-like representations of tactile data, where spatio-temporal features are learned for improved recognition performance [22]. Beyond tactile sensing, gripper-based recognition frequently integrates kinesthetic information, such as joint angles or actuator positions, to complement tactile cues.

Gripper design also plays a critical role in recognition performance. Adaptive and underactuated grippers have been widely used because their mechanical compliance facilitates diverse EPs without complex control strategies [23], [24]. Some designs integrate tactile arrays into passive fingers that adapt to object geometry, improving contact coverage and data richness [25]. Other works have also explored camera-based tactile sensors (e.g., GelSight) mounted on gripper fingertips, enabling high-resolution imaging of surface textures and deformation patterns [26].

Overall, object recognition with grippers has evolved from static tactile snapshots to dynamic, image-based and video-based representations, supported by advances in sensor technology and deep learning. These developments highlight the importance of EP design, sensor resolution, and gripper adaptability in achieving robust haptic perception.

### B. Multimodal perception

Multimodal perception is a hot topic in robotics. Fusing complementary signals improves robustness and data efficiency in perception pipelines across various domains (e.g., autonomous exploration, locomotion, and localization) [27]–[30]. Early works combined contact audio (obtained with a microphone) with vision for dynamic surface discrimination [31]. Within grasping, recent trends focus on integrating visual sensing with other modalities in robotic grippers. Most

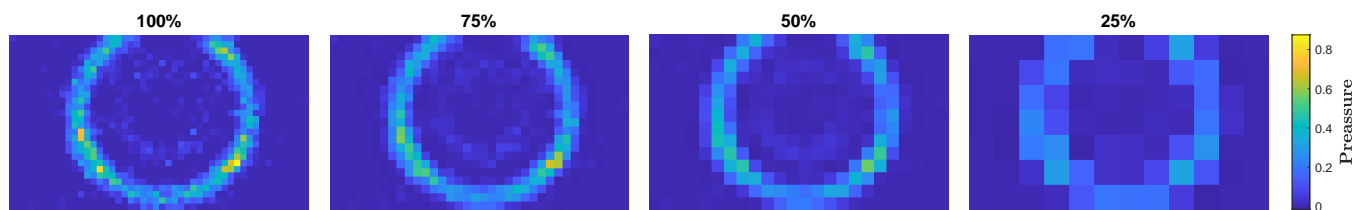


Fig. 1. Tactile frames of a 3D tactile video of a skate wheel with the different resolutions. From left to right: i) 100% dataset, with  $28 \times 48$  pixels, ii) 75% dataset, with  $21 \times 36$  pixels, iii) 50% dataset, with  $14 \times 24$  pixels, and iv) 25% dataset, with  $7 \times 12$  pixels.

common strategies fuse tactile and visual inputs to classify objects or surfaces, learning cross-modal features that are more discriminative than either modality alone [32]–[35]. Other approaches fuse joint signals with vision to estimate in-hand object pose [36]. More recent visuo–tactile grippers report further gains and improved recognition accuracy [37] but also require extra calibration and computational costs [38]. These methods can be impractical when the visual channel is degraded by clutter, occlusion, or adverse lighting, motivating alternatives that rely primarily on haptic information.

While multimodal haptic fusion offers clear advantages for object recognition, it has been explored in only a small number of studies. In [39], a recursive tactile sensing framework is developed to recognize materials using multimodal approaches. BioTac-based systems identify objects via compliance, texture, and thermal cues, illustrating the value of multimodal haptics even with a low resolution tactile sensor [40], [41]. Combining proprioceptive and tactile features has proved to improve identification accuracy from few, informative touches [42]. In this sense, kinesthetic-tactile fusion has been utilized by embedding pressure and 3D contact locations into 4D point clouds for classification [10]. Gripper-centric designs that integrate tactile arrays with passive fingers that adapt to the grasped object also enable recognition under palpation procedures [43]. A similar approach was considered in [25], where a gripper with reconfigurable underactuated fingers and tactile sensing is presented. These works highlight the potential of combining tactile and kinesthetic modalities, yet the impact of reduced tactile resolution on such fusion strategies remains unexplored.

### III. MATERIALS AND METHODS

#### A. Robotic Gripper

The experimental setup employs the underactuated, sensorized gripper previously described in [44]. The gripper consists of three fingers: two actuated fingers with two phalanges each, and a fixed thumb that integrates a high-resolution tactile array. The actuated fingers are driven by smart servos and equipped with joint angle sensors, enabling the acquisition of kinesthetic information during grasping. The tactile array embedded in the thumb provides pressure images with 1400 taxels arranged in a  $28 \times 50$  matrix, covered by a silicone pad to improve contact quality. This configuration allows the gripper to simultaneously capture tactile and kinesthetic data during squeeze-and-release exploratory procedures, providing complementary information about object stiffness, shape, and internal features.

#### B. Dataset

The dataset used in this work partially corresponds to the multimodal collection introduced in [11], which was specifically designed to evaluate haptic fusion strategies. It comprises 36 objects grouped into rigid, deformable, and in-bag categories, where the in-bag objects contain small elements packaged inside a plastic bag. Each object was grasped 60 times, resulting in a total of 2160 grasp sequences. For each grasp, synchronized tactile and kinesthetic data were recorded. Tactile data consists of sequences of pressure images from the Tekscan array, capturing the variation of contact forces during the squeeze-and-release exploratory procedure. Each sequence consists of 21 frames of  $28 \times 50$  pixels, forming a spatiotemporal representation of the interaction. Kinesthetic data represents time series of joint angles from the actuated and underactuated finger joints, providing information about the global configuration of the gripper and the object's geometry.

The dataset is particularly challenging because objects within the same category often share similar physical properties (e.g., sponges with and without inclusions, or bags filled with nuts of different sizes). This makes it an ideal benchmark for analyzing the effect of tactile resolution on recognition accuracy.

##### 1) Downsampling Tactile Data for Performance Analysis:

Although high-resolution tactile arrays provide rich information, their integration into robotic grippers is challenging due to hardware complexity, cost, and computational demands. Many tactile grippers reported in the literature employ sensors with fewer taxels, which reduces spatial resolution and may compromise recognition accuracy [45]. To analyze this trade-off, we artificially reduce the resolution of the tactile images by applying bicubic interpolation, a widely used resampling method in image processing that provides smoother results than bilinear or nearest-neighbor interpolation [46]. Bicubic interpolation has also been applied in tactile perception studies to downsample pressure images while preserving structural information [47], [48].

In this work, tactile images from the original dataset are resized to lower resolutions to analyze the effect of reduced tactile information on object recognition performance. Specifically, we create a new dataset composed of four sub-datasets with different resolutions:

- **100% Dataset:** Original tactile data, excluding the first and last columns for curation purposes, yielding a  $28 \times 48$  matrix.
- **75% Dataset:** Downsampled to retain 75% of the original

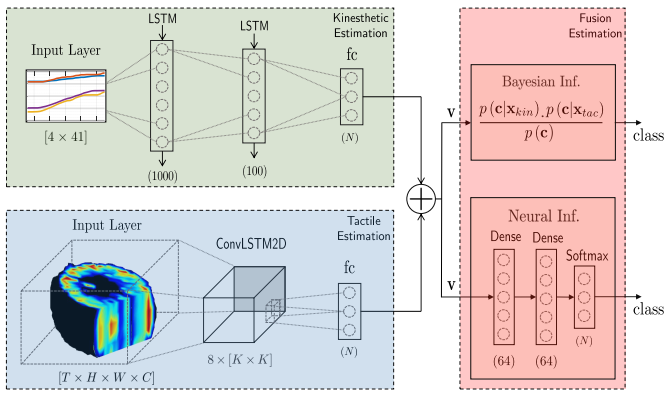


Fig. 2. The tactile network (ConvLSTM-based) and the kinesthetic network (LSTM-based) independently estimate class-probability vectors. These outputs are then fused using either an analytical Bayesian rule or a neural fusion inference to obtain the object classification vector.

pixels, resulting in a  $21 \times 36$  matrix.

- **50% Dataset:** Reduced to 50% of the original resolution, producing a  $14 \times 24$  matrix.
- **25% Dataset:** Downsampled to 25% of the original resolution, yielding a  $7 \times 12$  matrix.

A visual representation of an example of the four sub-datasets is presented in Fig. 1. The figure presents a tactile image with the different resolutions of the squeeze-and-release procedure performed on the skate wheel object.

Kinesthetic information is not reduced because the number of moving fingers is not changed, and the size of the kinesthetic data vector depends on the sensor's inherent reading capability. Unlike tactile sensors, kinesthetic sensors can typically be mounted on any gripper design without modification. Therefore, reducing kinesthetic data would not be meaningful, as the sensor configuration and functionality remain unchanged.

### C. ConvLSTM Architecture for the Tactile Object Recognition Problem

To evaluate the impact of tactile resolution on recognition performance, we train four deep learning models based on a ConvLSTM architecture [49], one for each dataset variant described in Section III-B.1. ConvLSTM networks combine convolutional layers with recurrent LSTM units, enabling the extraction of spatiotemporal features from sequences of tactile images. This design is particularly suited for squeeze-and-release exploratory procedures, where object properties are revealed progressively over time.

Figure 2 illustrates the overall architecture and input configurations for the four models. Each network receives a sequence of 21 tactile frames as input, with spatial dimensions determined by the dataset resolution:  $28 \times 48$  (100%),  $21 \times 36$  (75%),  $14 \times 24$  (50%), and  $7 \times 12$  (25%). These frames are stacked along the temporal axis, forming a 4D tensor of shape  $(T, H, W, C)$ , where  $T = 21$  is the number of frames,  $H$  and  $W$  are the height and width of the tactile matrix, and  $C = 1$  is the single-channel pressure map.

The first layer in each model is a ConvLSTM2D block, which applies convolutional kernels across both spatial and

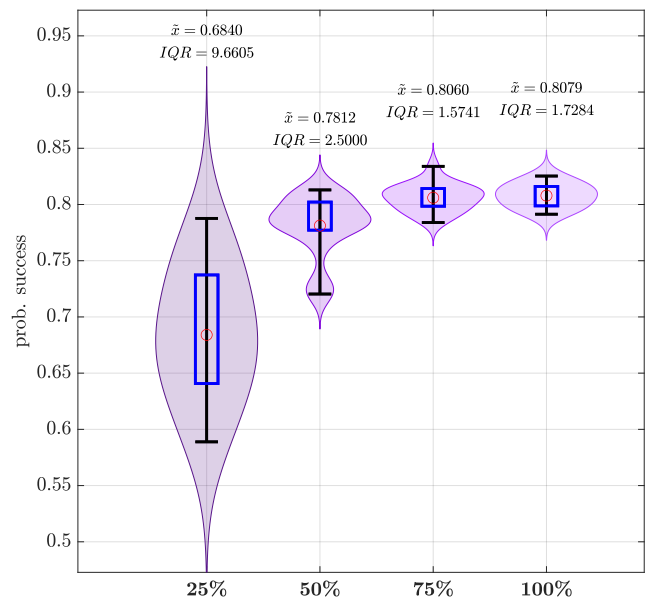


Fig. 3. Boxplots and violin plots showing accuracy distributions for tactile-only classifiers at four resolution levels (25%, 50%, 75%, and 100%) across 20 training runs.

temporal dimensions while maintaining recurrent connections to capture dynamic contact patterns. For the 100% and 75% datasets, we use kernels  $(K \times K)$  of size 5, while for the 50% and 25% datasets, smaller kernels  $(K = 3)$  are employed to match the reduced spatial resolution. All ConvLSTM layers have 8 filters, use  $\tanh$  activation and same padding to preserve spatial dimensions.

After the ConvLSTM layer, the output tensor is flattened and passed through a fully connected (fc) layer with 36 neurons, corresponding to the number of object classes in the dataset. The final layer uses a  $\text{softmax}$  activation to produce a probability distribution over all classes, enabling categorical object recognition.

All models are trained using the Adam optimizer with a learning rate of  $10^{-4}$  and categorical cross-entropy loss. Training is performed for 40 epochs with a batch size of 16.

These experiments help identify the minimum resolution required to maintain reliable object recognition without incurring unnecessary computational or hardware costs.

### D. Fusion of Tactile and Kinesthetic Information

To mitigate the performance degradation caused by reduced tactile resolution, we also evaluate fusion strategies that combine tactile and kinesthetic modalities. Both Bayesian and neural-based fusion approaches are applied to the resized tactile data and the kinesthetic sequences. This analysis allows us to quantify how multimodal fusion can compensate for the loss of tactile resolution and maintain robust recognition performance.

Kinesthetic information is presented as a time series of joint angles from the two actuated and two underactuated finger joints (4 channels over 41 time steps). The network, presented in Fig.2, consists of two LSTM layers with 1000 and 100 units respectively, both with  $\tanh$  activations, and a final fully

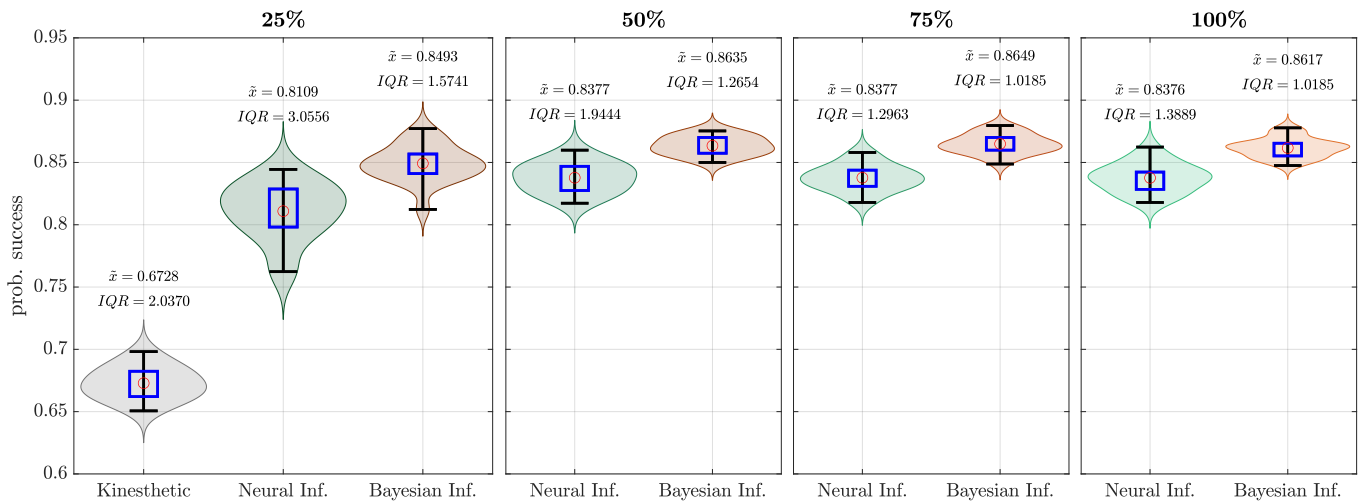


Fig. 4. Probability distributions of recognition accuracy for each tactile resolution (25%, 50%, 75%, and 100%). Each subplot displays the mean, quartiles, interquartile range (IQR), and violin plots for the Bayesian fusion and neural-based fusion classifiers computed over 20 runs. The 25% subplot also includes the results of the kinesthetic-only classifier, shown once because it does not vary with the spatial resolution of the tactile data.

connected layer that outputs a posterior distribution over the 36 object classes. The model is trained with Adam (learning rate  $10^{-5}$ ), categorical cross-entropy, batch size 16, for 700 epochs, using a 0.2 validation split and shuffled mini-batches. Prior to training, the joint signals are normalized per channel. Its configuration is identical across all experiments because the kinesthetic input dimensionality does not depend on tactile resolution.

1) *Bayesian inference*: The Bayesian fusion approach exploits the probabilistic nature of the unimodal classifiers. Both the tactile and kinesthetic networks output a categorical probability distribution over the set of object classes, denoted as  $p(\mathbf{c}|\mathbf{x}_{tac})$  and  $p(\mathbf{c}|\mathbf{x}_{kin})$ , respectively. The class labels  $\mathbf{c} \in \mathcal{Y}$  are defined prior to training, and the total number of labels ( $N$ ) is defined as 36, corresponding to the number of objects in the dataset. Assuming conditional independence between modalities given the class, the joint posterior distribution is computed as:

$$p(\mathbf{c}|\mathbf{x}_{tac}, \mathbf{x}_{kin}) \propto \frac{p(\mathbf{c}|\mathbf{x}_{tac}) \cdot p(\mathbf{c}|\mathbf{x}_{kin})}{p(\mathbf{c})}$$

where  $p(\mathbf{c})$  is the prior probability of each class, typically uniform in our experiments ( $p(\mathbf{c}) = 1/N$ ). A complete definition of this method applied to this particular problem is described in our previous work [11]. The Maximum A Posteriori (MAP) estimate is then obtained by selecting the class with the highest joint probability. This method is computationally efficient because it does not require additional training beyond the unimodal networks. It also provides theoretical guarantees under the independence assumption, making it robust when training data is limited or heterogeneous.

2) *Neural inference*: The neural-based fusion strategy adopts a data-driven approach to learn fusion rules directly from examples. Instead of applying an analytical formula, this method concatenates the probability distributions produced by the tactile and kinesthetic classifiers ( $p(\mathbf{c}|\mathbf{x}_{tac})$  and  $p(\mathbf{c}|\mathbf{x}_{kin})$ ) into a single feature vector of length  $\mathbf{v} \in \mathbb{R}^{2 \cdot N}$ . This vector

is then fed into a fully connected fusion network composed of two dense layers with 64 neurons each, followed by a softmax layer that outputs the fused classification probabilities ( $p(\mathbf{c}|\mathbf{x}_{tac}, \mathbf{x}_{kin})$ ). The neural inference leverages supervised learning to capture complex relationships and correlations that may exist between tactile and kinesthetic signals. This flexibility allows the model to potentially outperform Bayesian fusion when trained on large datasets, as it can learn patterns beyond those captured by analytical rules. However, its performance is highly dependent on the amount and diversity of training data. With limited data, neural fusion may underfit or overfit, resulting in lower accuracy compared to Bayesian methods.

#### IV. RESULTS AND DISCUSSION

The performance of the models was evaluated in terms of recognition accuracy. In addition to accuracy metrics, the performance was also analyzed using confusion matrices, which provide a detailed view of class-wise recognition behavior and misclassification patterns. To ensure statistically meaningful results, each experiment was repeated 20 times. For every run, the dataset was randomly split into training, validation, and test sets.

##### A. Impact of Tactile Resolution on Object Recognition

Fig. 3 summarizes the recognition performance of tactile-only models trained at the four resolution levels. These results follow an expected trend: higher tactile resolution leads to better mean classification performances. However, the differences between consecutive resolutions reveal interesting insights. The smallest gap occurs between 100% and 75%, with a mean probability difference of only 0.0019, suggesting that reducing the resolution to 75% preserves most of the object's internal characteristics. The difference between 75% and 50% is slightly larger (0.0248), while the most significant drop occurs between 50% and 25% (0.0972), indicating that severe downsampling removes critical structural information required for accurate recognition.

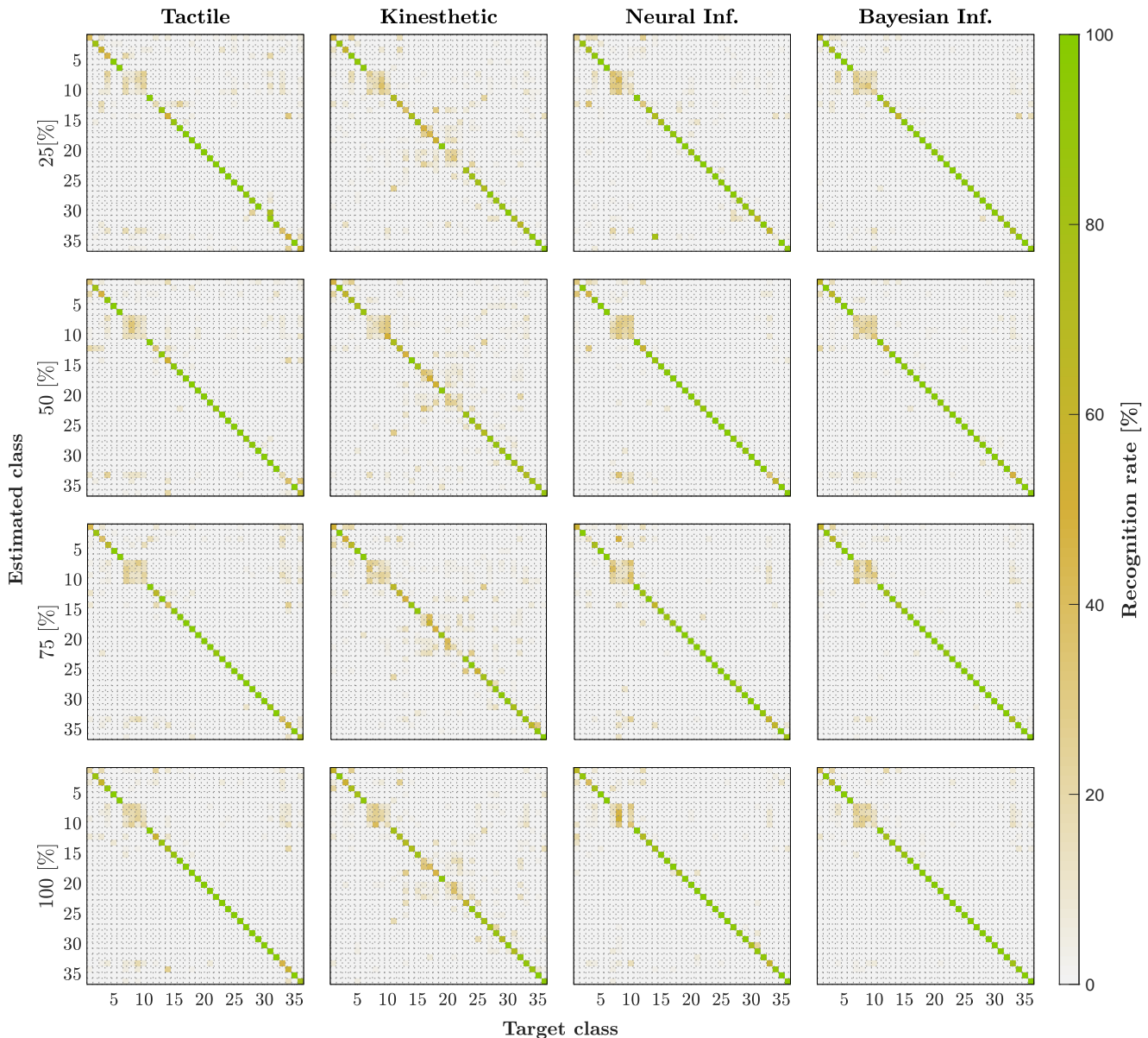


Fig. 5. Confusion matrices (CM) where the abscissa denotes the target class and the ordinate the estimated class. The color scale on the right indicates the recognition rate. The figure is organized as a 4x4 grid: rows correspond to tactile resolutions (top to bottom: 25%, 50%, 75%, 100%) and columns to classifiers, from left to right: i) tactile-only, ii) kinesthetic-only, iii) neural fusion, iv) Bayesian fusion.

Variability across runs was analyzed through the interquartile range (IQR). Both 100% and 75% datasets exhibit small IQRs, implying stable performance regardless of the specific training/validation/test split. In contrast, the 50% and especially the 25% datasets show larger IQRs, reflecting strong dependence on the training data and greater sensitivity to random splits. This effect is particularly evident at 25%, where the accuracy fluctuates significantly between runs, confirming that low-resolution tactile data compromises robustness.

The violin plots of Fig. 3 provide additional insight into the distribution of accuracies. Most resolutions exhibit distributions close to normal, except for the 50% case, where outliers appear, reducing overall performance. These outliers reinforce the observation that mid-level resolution introduces variability and dependence on the training set.

In summary, the results confirm the expected relationship between tactile resolution and recognition accuracy: more information yields better classification. Nevertheless, two noteworthy findings emerge. First, the IQR for 100% is slightly larger than for 75%, and the highest single-run accuracy was achieved by the 75% model. This suggests that the network configuration may not fully exploit the additional information at full resolution, and tuning hyperparameters could further improve performance. Second, the sharp decline at 25% resolution underscores the limitations of extreme downsampling, where essential object features are lost.

Fig. 5 displays 16 confusion matrices organized in four rows and four columns, corresponding to different resolution levels and each classifier. The first column shows one representative example of the tactile network results for each resolution.

These examples illustrate how reducing tactile information affects class-wise recognition behavior. At 100% resolution, the network achieves high accuracy for most objects. However, misclassifications occur among objects 7 to 10, which share very similar physical properties, and between objects 1 and 3. These errors persist as resolution decreases, but additional confusions emerge at lower resolutions. For instance, at 75%, the model begins to struggle with objects 12 and 14, as well as 34 and 36, indicating that subtle differences in shape or stiffness become harder to capture when spatial detail is reduced. At 50% resolution, the same error patterns remain but become more pronounced, with increased confusion among previously challenging classes. Finally, at 25% resolution, the degradation is most evident: objects 7 to 10, which were previously confused only among themselves, are now misclassified also as different objects. Moreover, new errors appear, such as object 30 being identified as 31, reflecting the loss of structural cues because of downsampling.

These observations confirm that reducing tactile resolution not only lowers overall accuracy but also amplifies ambiguity among objects with similar physical characteristics. At very low resolutions, the network fails to capture distinctive features, leading to widespread misclassification beyond the original error clusters.

### B. Performance of Tactile–Kinesthetic Fusion Strategies

For the multimodal experiments, the dataset was also split randomly into training, validation, and test sets across 20 independent runs to ensure statistical robustness. Crucially, tactile and kinesthetic data originating from the same grasp were always paired during these splits. This alignment ensures consistency between modalities and prevents mismatched sequences during fusion, which is essential for a fair comparison of Bayesian and neural-based strategies.

Fig. 4 summarizes the statistical results for both fusion approaches across all tactile resolutions. Each subplot shows the mean success rate, quartiles, interquartile range (IQR), and violin plots for the 20 experiments. The first subplot (25%) also includes the performance of the kinesthetic-only classifier, which remains constant across all resolutions. As expected, kinesthetic data alone achieves lower accuracy than tactile data.

The following analysis examines the behavior of the two fusion schemes, Bayesian and neural, across the different tactile resolutions. Neural inference results show that the mean success rate for 100%, 75%, and 50% resolutions is nearly identical, while the 25% case falls slightly behind, with a difference of approximately 0.0268 compared to the others. This demonstrates that kinesthetic information effectively compensates for the loss of tactile detail at moderate resolutions. However, at 25%, the fusion cannot fully recover the missing spatial information, leading to a noticeable drop in accuracy. The IQR analysis reveals a trend similar to tactile-only models: the largest variability occurs at 25%, followed by 50%, while 75% and 100% remain close and relatively small. Overall, all IQRs are smaller than those observed in tactile-only training, confirming that fusion improves consistency.

Bayesian inference exhibits a similar pattern but with slightly better performance across all resolutions. The success rates for 50%, 75%, and 100% are very close, while the 25% case again shows a larger gap. Bayesian inference consistently outperforms neural inference at every resolution, likely because the analytical approach does not depend on large training datasets to learn fusion rules. Furthermore, the IQRs for Bayesian inference are smaller than those of neural inference, reinforcing its robustness under limited data conditions.

When comparing both fusion strategies to tactile-only results, the benefits of multimodal integration become evident. The large accuracy gaps observed between resolutions in tactile-only models are substantially reduced after fusion. For example, reducing tactile data to 50% or even 25% no longer results in dramatic performance loss: Bayesian fusion achieves accuracy within roughly 0.01 of the full-resolution case, while neural fusion remains within 0.02. In contrast, tactile-only models showed a 0.12 drop between 100% and 25%. These findings confirm that combining tactile and kinesthetic modalities enables significant reductions in tactile resolution without sacrificing recognition capability.

Fig. 5, in addition to presenting the tactile network results in the first column, also includes the performance of the kinesthetic network and the two fusion strategies: Bayesian inference and neural-based inference. Each column shows how classification accuracy evolves across different data percentages (25%, 50%, 75%, and 100%), providing a complete view of the system's behavior under varying conditions. When examining the kinesthetic training column, we see that, as expected, this modality tends to misclassify more objects compared to the tactile network. Comparing both networks reveals several patterns: in many cases, both networks achieve high accuracy, while in others neither is able to classify the object correctly—particularly for objects 7 through 10. There are also situations where the tactile network succeeds while the kinesthetic network fails, such as with objects 20, 21, and 22, and, less frequently, the opposite occurs, as seen with objects 33 and 34.

Turning to the fusion-based inference columns, several interesting patterns emerge. Both neural and Bayesian inference methods demonstrate the ability to learn which source of information is more reliable for each object. This adaptability is evident in cases where only one modality (either tactile or kinesthetic) could classify the object correctly; the fusion methods also achieve correct classification in these situations. The only significant misclassifications occur when neither individual modality was able to recognize the object—again, objects 7 through 10. Overall, these results confirm that inference-based fusion strategies effectively leverage complementary information from tactile and kinesthetic sources, substantially improving robustness compared to single-modality networks.

## V. CONCLUSIONS AND FUTURE WORKS

This study provided an evaluation of how tactile resolution influences object recognition performance across a diverse set

of objects. The results showed that reducing tactile resolution consistently degraded recognition accuracy, particularly for objects with complex geometries or deformable properties. This confirmed that high-resolution tactile sensing is needed for achieving reliable classification. At the same time, the experiments demonstrated that this loss in accuracy could be mitigated through multimodal fusion strategies. Both Bayesian and neural-based inference approaches successfully combined tactile and kinesthetic information, compensating for the limitations introduced by lower tactile resolution. These fusion methods not only improved overall recognition rates but also adapted to the reliability of each modality, enabling robust performance. By quantifying these trade-offs, the work highlighted the importance of balancing sensor resolution with computational strategies for efficient robotic perception. These findings offer practical insights for designing tactile systems that optimize hardware complexity without sacrificing recognition capabilities.

Future work will focus on improving the tactile network architecture to fully exploit the information available at high resolutions, as results suggest that the current configuration does not yet capture all discriminative features provided by the 100% dataset. Investigating advanced fusion techniques, including attention-based and advanced probabilistic models, will also be a priority to improve the object recognition performance. Lastly, extending the evaluation to real manipulation tasks and considering additional sensing modalities will help assess the scalability and practical deployment of these approaches in more complex environments.

## REFERENCES

- [1] W. Montagna, *The structure and function of skin*. Elsevier, 2012.
- [2] A. L. Dellon, S. E. Mackinnon, and P. M. Crosby, "Reliability of two-point discrimination measurements," *The Journal of hand surgery*, vol. 12, no. 5, pp. 693–696, 1987.
- [3] J. Tong, O. Mao, and D. Goldreich, "Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment," *Frontiers in human neuroscience*, vol. 7, p. 579, 2013.
- [4] U. Proske and S. C. Gandevia, "The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force," *Physiological reviews*, 2012.
- [5] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—from humans to humanoid," *IEEE transactions on robotics*, vol. 26, no. 1, pp. 1–20, 2009.
- [6] C. Bartolozzi, L. Natale, F. Nori, and G. Metta, "Robots with a sense of touch," *Nature materials*, vol. 15, no. 9, p. 921, 2016.
- [7] B. Siciliano, L. Villani, G. Oriolo, A. De Luca *et al.*, "Foundations of robotics," *Advanced Textbooks in Control and Signal Processing*, pp. 1–719, 2025.
- [8] R. Dahiya, N. Yogeswaran, F. Liu, L. Manjakkal, E. Burdet, V. Hayward, and H. Jörmell, "Large-area soft e-skin: The challenges beyond sensor designs," *Proceedings of the IEEE*, vol. 107, no. 10, pp. 2016–2033, 2019.
- [9] R. S. Dahiya, P. Mittendorf, M. Valle, G. Cheng, and V. J. Lumelsky, "Directions toward effective utilization of tactile skin: A review," *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4121–4138, 2013.
- [10] S. Luo, W. Mou, K. Althoefer, and H. Liu, "iCLAP: Shape recognition by combining proprioception and touch sensing," *Autonomous Robots*, vol. 43, no. 4, pp. 993–1004, 2019.
- [11] F. Pastor, J. García-González, J. M. Gandarias, D. Medina, P. Closas, A. J. García-Cerezo, and J. M. Gomez-de Gabriel, "Bayesian and neural inference on lstm-based object recognition from tactile and kinesthetic information," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 231–238, 2020.
- [12] A. Trujillo-León, W. Bachtá, J. Castellanos-Ramos, and F. Vidal-Verdú, "Assistive handlebar based on tactile sensors: control inputs and human factors," *Sensors*, vol. 18, no. 8, p. 2471, 2018.
- [13] J. Colan, A. Davila, and Y. Hasegawa, "Tactile feedback in robot-assisted minimally invasive surgery: A systematic review," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 20, no. 6, p. e70019, 2024.
- [14] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Design of a flexible tactile sensor for classification of rigid and deformable objects," *Robotics and Autonomous Systems*, vol. 62, no. 1, pp. 3–15, 2014.
- [15] Z. Zhang, J. Zhou, Z. Yan, K. Wang, J. Mao, and Z. Jiang, "Hardness recognition of fruits and vegetables based on tactile array information of manipulator," *Computers and Electronics in Agriculture*, vol. 181, p. 105959, 2021.
- [16] F. Vidal-Verdú, M. J. Barquero, J. Serón, and A. García-Cerezo, "Large area smart tactile sensor for rescue robot," in *2009 IEEE International Workshop on Robotic and Sensors Environments*. IEEE, 2009, pp. 6–10.
- [17] S. Pohongkam and J. Srinonchat, "Object recognition for humanoid robots using full hand tactile sensor," *IEEE Access*, vol. 11, pp. 20284–20297, 2023.
- [18] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel Tactile-SIFT Descriptor for Object Shape Recognition," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5001–5009, 2015.
- [19] A. M. Okamura and M. R. Cutkosky, "Feature detection for haptic exploration with robotic fingers," *The International Journal of Robotics Research*, vol. 20, no. 12, pp. 925–938, 2001.
- [20] M. Meribout, N. A. Takele, O. Derege, N. Rifiki, M. El Khalil, V. Tiwari, and J. Zhong, "Tactile sensors: A review," *Measurement*, vol. 238, p. 115332, 2024.
- [21] B. Zapata-Impata, P. Gil, F. Torres, B. S. Zapata-Impata, P. Gil, and F. Torres, "Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction Of Slip Detection," *Sensors*, vol. 19, no. 3, p. 523, 1 2019.
- [22] M. Madry, L. Bo, D. Kragic, and D. Fox, "ST-HMP: Unsupervised Spatio-Temporal feature learning for tactile data," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 2262–2269.
- [23] I. Hussain, F. Renda, Z. Iqbal, M. Malvezzi, G. Salvietti, L. Seneviratne, D. Gan, and D. Prattichizzo, "Modeling and prototyping of an underactuated gripper exploiting joint compliance and modularity," *IEEE Robotics and automation letters*, vol. 3, no. 4, pp. 2854–2861, 2018.
- [24] H. Hua, X. Wu, C. Zhao, Z. Wu, J. Song, and Z. Liao, "Design and evaluation of a pneumatic underactuated robotic gripper with in-hand multi-mode self-perception ability," *Journal of Mechanical Science and Technology*, pp. 1–11, 2025.
- [25] Z. Lu, H. Guo, W. Zhang, and H. Yu, "Gtac-gripper: A reconfigurable under-actuated four-fingered robotic gripper with tactile sensing," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7232–7239, 2022.
- [26] R. Li and E. H. Adelson, "Sensing and recognizing surface textures using a gelsight sensor," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1241–1247.
- [27] S. Gou, X. Chen, and W. Zhang, "Multimodal perception based autonomous exploration with active camera control in unknown environments," in *2022 34th Chinese Control and Decision Conference (CCDC)*. IEEE, 2022, pp. 5577–5582.
- [28] R. M. Claro, D. B. Silva, and A. M. Pinto, "Artuga: A novel multimodal fiducial marker for aerial robotics," *Robotics and Autonomous Systems*, vol. 163, p. 104398, 2023.
- [29] W. Hu, G. Z. Lum, M. Mastrangeli, and M. Sitti, "Small-scale soft-bodied robot with multimodal locomotion," *Nature*, vol. 554, no. 7690, pp. 81–85, 2018.
- [30] D. Medina, C. Lass, E. Pérez-Marcos, R. Ziebold, P. Closas, and J. García, "On GNSS Jamming Threat from the Maritime Navigation Perspective," in *IEEE International Conference on Information Fusion (FUSION)*, 2019, pp. 2–5.
- [31] O. Kroemer, C. H. Lampert, and J. Peters, "Learning dynamic tactile sensing with robust vision-based training," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 545–557, 2011.
- [32] P. Falco, S. Lu, A. Cirillo, C. Natale, S. Pirozzi, and D. Lee, "Cross-modal visuo-tactile object recognition using robotic active exploration," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 5273–5280.
- [33] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data," in *IEEE International Conference on Robotics and Automation (ICRA)*, 6 2016, pp. 536–543.
- [34] H. Liu, Y. Yu, F. Sun, and J. Gu, "Visual-Tactile Fusion for Object Recognition," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 996–1008, 2017.
- [35] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and

touch: Learning multimodal representations for contact-rich tasks,” *IEEE Transactions on Robotics*, 2020.

- [36] M. Pfanne, M. Chalon, F. Stulp, and A. Albu-Schaffer, “Fusing joint measurements and visual features for In-Hand object pose estimation,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3497–3504, 2018.
- [37] W. Fan, H. Li, W. Si, S. Luo, N. Lepora, and D. Zhang, “Vitactip: Design and verification of a novel biomimetic physical vision-tactile fusion sensor,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 1056–1062.
- [38] W. Fan, H. Li, and D. Zhang, “Magicgripper: A multimodal sensor-integrated gripper for contact-rich robotic manipulation,” *arXiv preprint arXiv:2505.24382*, 2025.
- [39] A. Gomez Eguiluz, I. Rano, S. Coleman, and T. McGinnity, “Multi-modal Material Identification through Recursive Tactile Sensing,” *IEEE Robotics and Autonomous Systems*, vol. 106, pp. 130–139, 8 2018.
- [40] J. A. Fishel and G. E. Loeb, “Bayesian exploration for intelligent identification of textures,” *Frontiers in neurorobotics*, vol. 6, p. 4, 2012.
- [41] D. Xu, G. E. Loeb, and J. A. Fishel, “Tactile identification of objects using bayesian exploration,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 3056–3061.
- [42] M. A. Schiefer, E. L. Graczyk, S. M. Sidik, D. W. Tan, and D. J. Tyler, “Artificial tactile and proprioceptive feedback improves performance and confidence on object identification tasks,” *PLoS one*, vol. 13, no. 12, 2018.
- [43] N. Gorges, S. E. Navarro, D. Göger, and H. Wörn, “Haptic object recognition using passive joints and haptic key features,” in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 2349–2355.
- [44] F. Pastor, J. M. Gandarias, A. J. García-Cerezo, and J. M. Gómez-Gabriel, “Using 3d convolutional neural networks for tactile object recognition with robotic palpation,” *Sensors*, vol. 19, no. 24, p. 5356, 2019.
- [45] E. Kirby, R. Zenha, and L. Jamone, “Comparing single touch to dynamic exploratory procedures for robotic tactile object recognition,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4252–4258, 2022.
- [46] M. Jahnavi, D. R. Rao, and A. Sujatha, “A comparative study of super-resolution interpolation techniques: Insights for selecting the most appropriate method,” *Procedia Computer Science*, vol. 233, pp. 504–517, 2024.
- [47] S. Pohongkam and J. Srinonchat, “Tactile object recognition for humanoid robots using new designed piezoresistive tactile sensor and dcnm,” *Sensors*, vol. 21, no. 18, p. 6024, 2021.
- [48] R. Gao, W. Yuan, and J.-Y. Zhu, “Controllable visual-tactile synthesis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 7040–7052.
- [49] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in *Advances in neural information processing systems*, 2015, pp. 802–810.



**Francisco Pastor** received his Ph.D. in Mechatronics from the University of Málaga in 2023. He works in the Robotics and Mechatronics Group, where he investigates haptic perception and pHRI for applications such as limb manipulation, rehabilitation, and Search-and-Rescue operations. His work includes the development of tactile-kinesthetic data fusion models, underactuated grippers, and autonomous systems for sensor placement on victims. He has participated in key national research projects,

including TRUST-ROB and FIRST-ROB, focused on robotics for search-and-rescue missions. He is the author of more than 20 journal and conference publications and a co-inventor of an international patent on controllable robotic fastening systems.



**Da-Hui Lin-Yang** received his B.Sc. degree in Mechatronics Engineering in 2018 and his M.Sc. degree in Mechatronics Engineering in 2020 from the University of Málaga, Spain, where he is currently pursuing a Ph.D. degree. He works in the Robotics and Mechatronics Group, where he investigates model predictive control and trajectory planning methods for real-time autonomous systems applied to unmanned aerial vehicles, including applications in Search-and-Rescue operations.



**Alfonso J. García-Cerezo** received the Industrial Electrical Engineer and Doctoral Engineer degrees from the Escuela Técnica Superior de Ingenieros Industriales of Vigo in 1983 and 1987, respectively. From 1983 to 1991, he held academic positions at the University of Santiago de Compostela, serving as Associate Professor and later Assistant Professor in the Department of Electrical Engineering, Computers, and Systems. Since 1992, he has been Professor of Systems Engineering and Automation at the University of Málaga (UMA). He served as Dean of the School of Industrial Engineering at UMA from 1993 to 2004 and has also served as Director of the Department of Systems Engineering and Automation. He leads the Robotics and Mechatronics Group and played a key role in establishing the University Research Institute in Mechatronic Engineering and Cyberphysical Systems. He has been the principal investigator of 86 research projects, holds 15 patents, and has authored more than 240 publications, including journal articles, conference papers, book chapters, and technical reports. He has supervised 16 Ph.D. theses. His research focuses on mobile robots and autonomous vehicles, surgical robotics, mechatronics, search and rescue robotics, and intelligent control.



**Juan M. Gandarias** is an Assistant Professor in the Systems Engineering and Automation Department at the University of Malaga, Spain. He received his B.Sc. in Industrial Engineering and his M.Sc. and Ph.D. in Mechatronics Engineering from the University of Málaga in 2015, 2017, and 2020, respectively. He is the Principal Investigator of the STIFF-JAM and TYRELL projects. He has supervised one Ph.D. thesis, one M.Sc., and 18 B.Sc. theses. He has been a senior postdoctoral researcher in the Human-

Robot Interfaces and Interaction group at the Istituto Italiano di Tecnologia. He collaborated as scientific advisor in the Real-Move start-up project, awarded with Proof-of-Concept funding from the ERC and winner of the SmartCup Liguria Award (industrial category). He has participated in European, Spanish, and Italian projects as a predoctoral and postdoctoral researcher. He has authored 40 publications in scientific journals and flagship conferences in the fields of robotics, automation, and artificial intelligence. He was a finalist for the best paper award in Human-Robot Interaction at the ICRA 2024 conference and in mobile manipulation at the IROS 2022 conference. He is the inventor of an international patent. He served as Editor of several Special Issues and Conferences. During his PhD, he spent four months as a Ph.D. visitor at the Soft Haptics & Robotics Lab in the Department of Mechanical Engineering at the University College London. He has taught approximately 300 hours for various Engineering degrees.