

EDAMS: An Encoder-Decoder Architecture for Multi-grasp Soft Sensing Object Recognition

Oliver Shorthose¹, Alessandro Albini, Luca Scimeca, Liang He and Perla Maiolino

Abstract—The use of tactile sensing exhibits benefits over visual detection as it can be deployed in occluded environments and can provide deeper information about an object’s material properties. Soft hands have increasingly been used for tactile object identification, providing a high degree of adaptability without requiring complex control schemes. In this work, we propose a framework for identifying a range of objects in any pose by exploiting the compliance of a soft hand equipped with distributed tactile sensing. We propose EDAMS, an Encoder-Decoder Architecture for Multi-grasp Soft sensing and an ad-hoc data structure capable of encoding information on multiple grasps, while decoupling the dependency on the pose order. We train the model to map the high-dimensional multi-grasp tactile sensor data into a lower-dimensional latent space capable of achieving the geometrical separation of each object class, and enabling accurate object classification. We provide an empirical analysis of the benefit of multi-grasp perception for object identification, and show its impact on the separation of the objects in sensor space. Notably, we find the classification accuracy to change widely across the number of grasps, ranging from 47.0% for a single grasp, to 99.9% for 10 grasps.

Index Terms—Soft robotic hands, object identification, tactile sensing

I. INTRODUCTION

Object recognition stands as one of the most fundamental capabilities for autonomous robots, leading to successful grasping and manipulation. Beyond the widespread vision-based methods for object recognition, tactile sensing has the potential to provide additional information about the physical properties of objects like surface features, friction, and stiffness [1]–[5]. It can also be utilised in unstructured environments with variable lighting conditions, transparent objects and occlusions [6]. Due to their potential, tactile-based and hybrid methods have seen a rapid development in recent years for object recognition [7], [8].

The majority of works in the relevant literature make use of rigid end-effectors equipped with tactile sensors to perform object recognition. However, these grippers

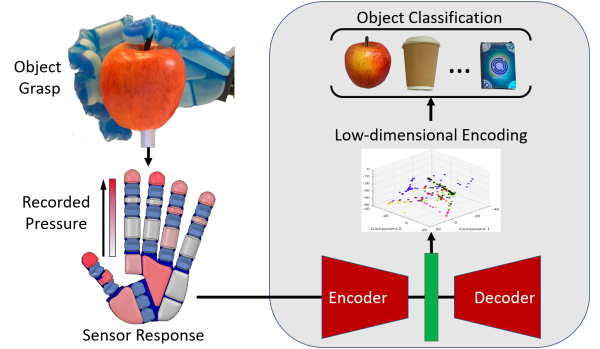


Fig. 1. Overview of the proposed model. Sensor data is recorded from multiple grasps of objects and fed into the EDAMS architecture. Classification is made on the low-dimensional clusters from the bottleneck of the encoder-decoder.

require complex control schemes to manipulate different types of objects and to guarantee accurate and safe interaction [9], [10]. On the other hand, due to their compliance, soft end-effectors can adapt to, and manipulate, a wide range of objects using simpler control strategies [11]–[13]. In this respect, several works have exploited the compliance of soft grippers to perform object recognition using proprioceptive sensing [2], [14] or a combination of proprioceptive and tactile sensors [15], [16] with simple open loop control. Indeed, the compliance allows the grasp configuration to adapt passively to the shape and stiffness of the objects, generating a different tactile response.

While the aforementioned works only grasp objects from a single posture, different approaches leverage the dexterity provided by an anthropomorphic hand to grasp objects in different orientations, achieving object recognition that is robust to pose uncertainty [17], [18]. For example, the work proposed in [17] combined flex and force sensors to discriminate between 10 different classes of objects, providing variation between the object poses when being grasped. In [18], a high resolution tactile sensor was integrated on a soft robot hand to be exploited for object recognition. The authors show that a high number of tactile sensors is fundamental in recognising object features successfully and the performance of a recognition system rapidly drops when the spatial resolution decreases. Whilst these works show the possibility of performing

We gratefully acknowledge support by EPSRC Programme Grant ‘From Sensing to Collaboration’ (EP/V000748/1)

O.Shorthose, A.Albini, L.He, and P.Maiolino are with Oxford Robotics Institute, University of Oxford, Oxford, UK. P.Maiolino is also with the University of Genoa, Liguria, IT. L.Scimeca is with Harvard University and Dana Farber, Boston, MA, USA.

¹ corresponding author’s email: ollies@robots.ox.ac.uk

object recognition from multi-modal feedback, or from high resolution tactile sensing using a single grasp, they require different types of sensors and flexible PCBs to be embedded in the soft body, thus increasing the complexity of the fabrication and affecting the durability and compliance of the system.

Soft pneumatic tactile sensors can be exploited and embedded to provide tactile feedback, without affecting the overall compliance of the body [19]. Despite this, fabrication methods currently only allow pneumatic sensors with low spatial resolution to be equipped on soft robots, where the information acquired from a single hand grasp may be inadequate for accurate object identification. In addition, tactile sensors can only capture local information leading to discriminative ambiguities. As a consequence, we argue that multiple contacts are often required to perform accurate object recognition via the sense of touch.

This paper addresses the problem of performing tactile object recognition from multiple grasping postures using a soft hand equipped only with pneumatic tactile sensors. This is achieved by using multi-grasp tactile information and employing an Encoder-Decoder Architecture for Multi-grasp Soft sensing (EDAMS) to map the information into a low-dimensional latent space, while enforcing the geometrical separation of different objects [20]. We empirically analysed how the accuracy increases in relation to the number of grasps, further stressing the importance of multi-grasp sensory information for tactile object discrimination.

In this paper, we improve on the hand-design from [21] to increase the contact sensing area and the conformability when grasping. The compliance and kinematics provided by the hand allow for a wide range of objects to be grasped in different orientations without complex control schemes. The objects are chosen from commonly used objects with different features such as a range of stiffness and geometries.

To summarise, the main contribution of this paper is a data structure and learning architecture capable of decoupling the dependency of the grasp pose sequences from sensor data, as well as learning a low level embedding that enforces meaningful geometric separation within object subsets in sensory space (EDAMS). The paper is structured as follows: Section II details the updated design and control scheme of the hand; Section III describes the experimental setup and the data collection procedure. The testing and analysis of the object identification task are presented in Section IV; Section V provides the conclusion of our work and the future direction.

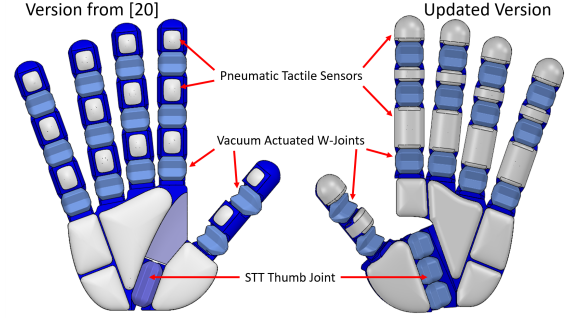


Fig. 2. Comparison of the soft hand presented in [21] (left) against the modified version (right).

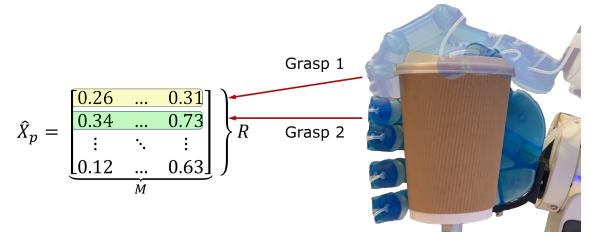


Fig. 3. Structure of the sample \hat{X}_p . Tactile responses corresponding to grasps from different postures are laid on separate rows, forming a matrix of multi-grasp sensor data. R is the number of grasps in the sample, and M is the number of sensors in the hand.

II. METHODOLOGY

A. Hand Design

The soft anthropomorphic hand's design was first presented in [21] and is shown in Fig. 2. The hand uses vacuum actuated W-joints to achieve human-like dexterity as demonstrated by the Feix taxonomy and Kapandji thumb opposition tests [22], [23]. Up to 15 degrees of freedom (DOFs) can be independently actuated, but to simplify the control for this work, the whole hand was underactuated as a single DOF.

Multi-material 3D printing is utilised to manufacture the hand and to achieve a fully integrated design which includes 19 pneumatic tactile sensors distributed on the fingers and the palm. Specifically, digital materials obtained from a combination of Agilus®, a compliant, rubber-like material, and VeroCyan®, a rigid plastic, are used in different proportions to achieve different shore hardness across the hand. In this respect, the soft joints are made from ShoreA 60, the pneumatic sensor chambers from ShoreA 35 and the structural part of the hand from ShoreA 95. The material proportions are available from the manufacturer.

The passive adaptability of the hand to the shape of the objects is crucial in generating the different sensor responses and simplifying the inference process for the object recognition task. Thus, several changes have been made to the original hand design presented in [21]:

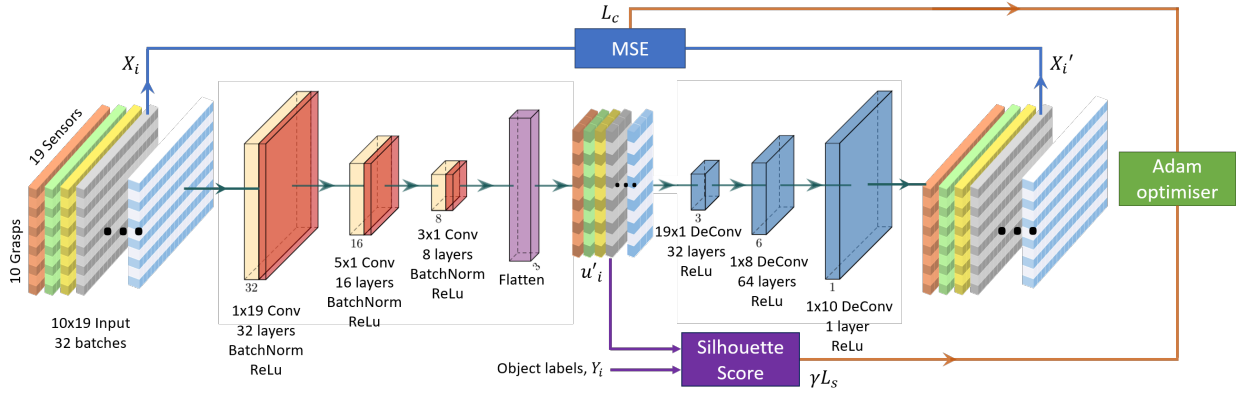


Fig. 4. Architecture of the EDAMS Encoder-Decoder Model. u'_i is the bottleneck data to be classified.

- 1) The surface area of the sensors has been increased to cover the sides of each finger and the finger tips. This was introduced to ensure that there was contact between the objects and sensors in all poses.
- 2) The lengths of each finger phalanx have been scaled to mimic the respective distances between joints in the human hand [24]. Specifically, shortening the medial phalanx reduces the contact area of the central sensor, but it increases the surface area of its neighbouring sensors and improves the grasping around objects.
- 3) The scaphotrapezotrapezoidal (STT) joint was rotated 10° out of the plane of the palm. In the previous design, we observed that the opposition of the thumb was not into the plane of the hand when the STT joint was actuated. The hand could satisfy the Feix and Kapandji tests as desired, but its grasping strength was limited by this. The increased angle was introduced to ensure the thumb grasps objects into the plane of the hand, increasing the stability of the grasp.

Fig 2 shows how the above modifications to the original design allow the hand to achieve greater conformability and improves the tactile sensors' response (with respect to the number of sensors in contact with the object and higher sensor values). Moreover, we performed a new characterisation of the sensors' responses with respect to applied forces, by following the same procedure explained in [21]. The results of the characterisation tests are presented in the supplementary video attached to the manuscript. The tactile sensor responses are linear in the applied displacement range and show good repeatability across three 3D printed fingers.

B. Model Architecture and Learning

In this Section we first explain how we organised the individual samples and the dataset to: (i) capture the

information of multiple grasp poses; (ii) make the classification robust to different sequences of grasping poses; (iii) evaluate the effect of the number of grasps on the classification accuracy. Finally, we describe the structure of the EDAMS model and the training procedure to ensure the separation of objects in the low-dimensional embedding.

1) *Dataset Structure*: We devise a data-structure that is suitable for learning, and that enables the *decoupling* of the sequence of the order of grasp poses in the data during learning. For each single grasp, we encode the tactile sensor responses as a 1D vector $\in \mathbb{R}^M$, where M is the number of sensors. We then form a matrix $\hat{X}_p \in \mathbb{R}^{R \times M}$, with $\{1, \dots, P\}$ and P the total recorded dataset size, containing the responses of the tactile system to R different grasping postures of the same object, repeated 30 times (Fig. 3).

Since the grasp data in \hat{X}_p follows the grasp pose sequence recorded during experiments, there would normally be a data dependency with respect to the order of the grasps and the object's identity. At the dataset level, we decouple this dependency with data augmentation. In particular we defined new samples X_i , with $i = \{1, \dots, N\}$, by randomly permuting, $K = 100$ times, the rows of each \hat{X}_p . In our case $P = 210$ (see Section III), thus, the total number of samples in the augmented dataset is $N = P \times K = 21000$.

To evaluate how the number of grasps affects object recognition, we replaced the respective number of lower rows in each sample X_i with zeros to simulate 1, 3, 5 and 7 grasps and retrained the model. These number of grasps were chosen to be equally distributed up to 10 grasps.

2) *Encoder-Decoder Model*: We developed an Encoder-Decoder Architecture for Multi-grasp Soft sensing (EDAMS) to be able to identify features of the grasp data, and produce a (meaningful) lower dimensional latent space that can subsequently be used to classify objects. In the encoder, three convolutional layers with *batch normalisation*

and *ReLU* activation were implemented as shown in Fig. 4. The first convolutional layer contains $1 \times M$ weight filters and can thus read the data from a single grasp at once, then stride across the row dimension to read all other grasps. This design, together with the data structure previously described, allows the model to learn *sensor adjacency* rules within the column dimension, but not the row dimension, thus achieving a decoupling of the order of grasps at the model level. This is followed by stacked convolution and reshape layers for the embedding, and de-convolutions for decoding (Fig. 4).

3) *Low-Dimensional Embedding Isolation*: For the purpose of our experiments, we wish to learn a low level embedding which enforces meaningful geometrical separation within object subsets in sensory space.

Let X_i be a general tactile data matrix, as previously defined. As per encoder-decoder literature we wish to encode X_i into a subspace u'_i , and then reconstruct the matrix into X'_i while enforcing reconstruction consistency. We implement a simple mean squared error (MSE) consistency loss of the form:

$$L_c = \sum_i \frac{(X_i - X'_i)^2}{N} = \sum_i \frac{(X_i - D(E(X_i)))^2}{N} \quad (1)$$

where E and D are respectively the encoder and decoder networks, and $u'_i = E(X_i)$. We enforce a geometrical separation between each object class within the u subspace by adding a differentiable implementation of the Silhouette Score within our function [25]. The Silhouette Score for any object class is given by:

$$L_s = \frac{a - b}{\max(a, b)} \quad (2)$$

where a is the average intra-cluster distance, and b is the average inter-cluster distance to the closest object class in the projected space [25], [26]. Finally, we train our model on a loss of the form:

$$L = L_c - \gamma L_s \quad (3)$$

where the hyper-parameter γ can modulate the model's *pressure* to find a sub-space enforcing geometrical object class separation. We use a vanilla Adam optimiser to handle the training dynamics [27], and choose to stop training when either $L_s \geq 0.99$ on the validation set (suggesting high cluster separation), or when there was no validation loss improvement for more than 100 epochs (early stopping). We further set u to be three-dimensional, to allow for easy plotting and evaluation. We find this low-dimensional sub-space to suffice in representing our set of objects.



Fig. 5. The selection of objects used for the identification tasks. The figure reports the dimensions in mm: H, height; W, width; L, length; D, max. diameter; d, min. diameter. The objects are presented in the orientation used for data acquisition.

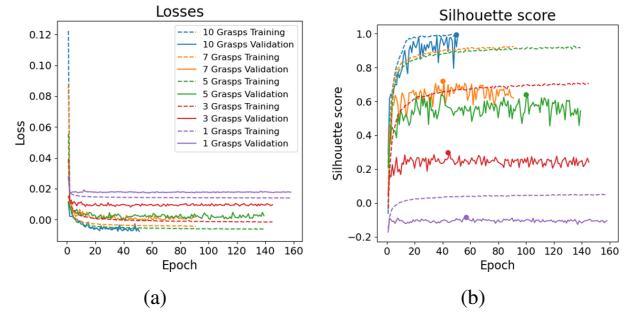


Fig. 6. Learning process for the EDAMS architecture. (a) The loss function is a summation of the MSE reconstruction loss and the silhouette score loss. (b) Silhouette score relating to the geometric cluster separation in the bottleneck.

III. DATA COLLECTION

The proposed method was validated in a tactile identification task based on 7 different objects, shown in Fig. 5. They were chosen by considering daily-use objects or shapes, similar to [15]–[18], [28]. In particular, we selected close variants of the YCB dataset [29]. The cylinder and cube are rigid, the bottle, cards and cup are semi-rigid, and the sponge is compliant. The soft hand was connected to a Franka Emika robot, controlled to move the end-effector to different grasping postures. The hand was actuated by connecting each of the 15 joints to a single negative pressure supply. This supply was provided by a compressor, regulated using an analog pressure regulator (SMC IRV20-C10). When negative pressure is applied to the joints, the fingers bend as characterised in [21]. Two solenoid valves (Yosoo1210) were used to control whether the hand was connected to the pressure source or the atmosphere. The soft tactile pads were connected to 19 pressure sensors (ADP51A11 with 1-bit in 10 of noise) via 1mm diameter tubing and wired to the Arduino through a multiplexer (74HC4051) and recorded at 10Hz.

The objects were fixed in front of the robot using a

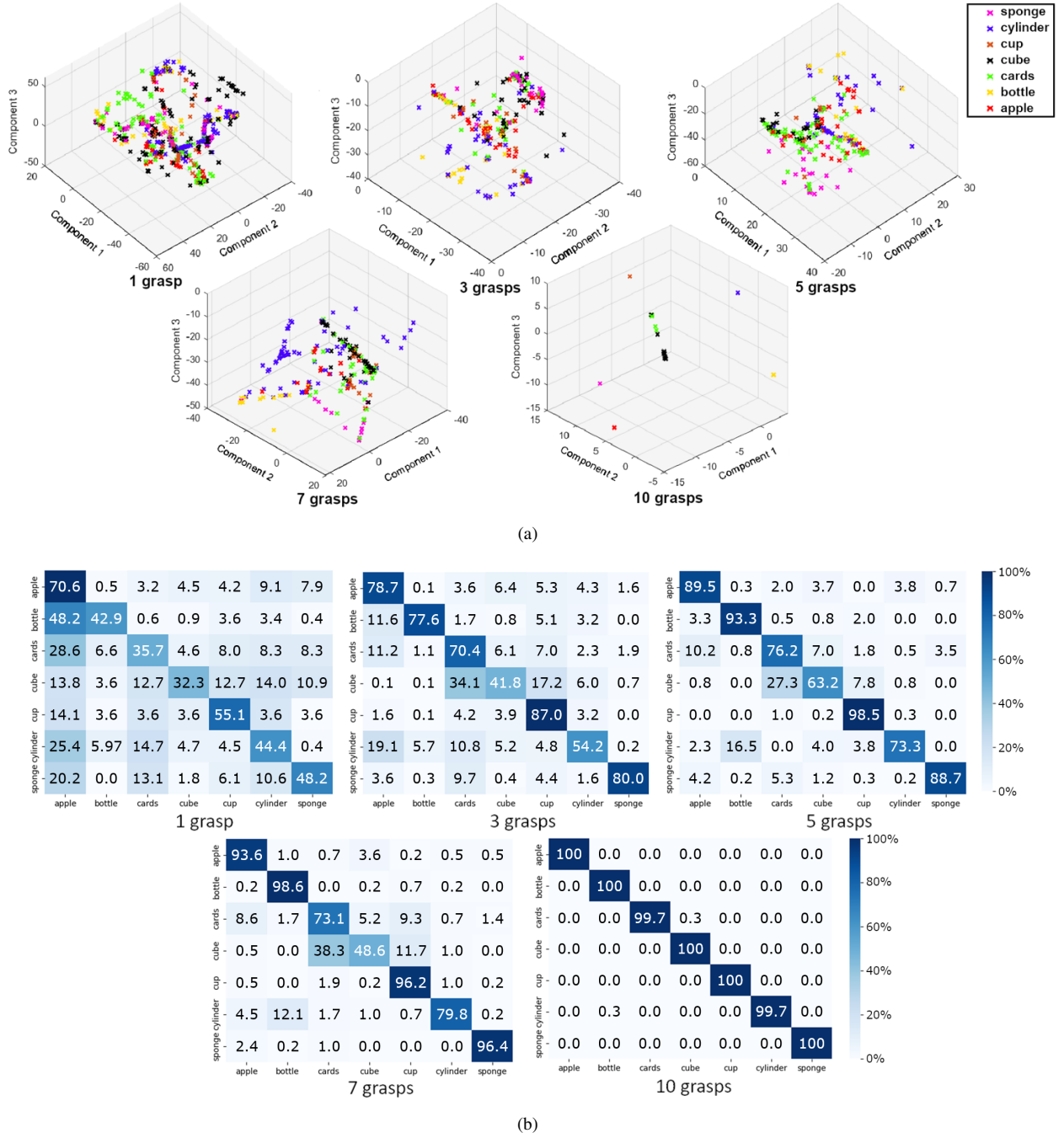


Fig. 7. Results on the test set by increasing the number of grasps (a) Object clusters from the bottleneck region in the encoder-decoder architecture. Increasing the number of grasps allows for greater separation between clusters and for the higher numbers of grasps, the datapoints are stacked in the same position. (b) Confusion matrices for the object classification. The classifier used for each number of grasps is detailed in Table I. The true values are on the y-axis and the predicted values are on the x-axis.

stand to establish a repeatable stationary position. The data sample $\hat{X}_p \in \mathbb{R}^{10 \times 19}$ was collected by grasping in 10 different postures. The grasp poses were manually chosen to approach each object from all directions and to cover its entire shape. Fig. 3 shows two of the 10 grasps for the

cup. The procedure to collect each \hat{X}_p consisted of the following steps:

- 1) Approach the object from an initial remote position.
- 2) Once the hand is in contact with the object, control the solenoid valves to reduce the pressure to -50kPa,

thus closing the hand around the object.

- 3) Hold the hand held in stationary contact with the object for 30 seconds. The final sensor reading from the steady state is then collected.
- 4) The pressure is released and the hand is removed back to a remote position.
- 5) Repeat the above procedure for the 10 different grasping postures.

This procedure was repeated 30 times for each of the 7 objects, therefore $P = 210$. To ensure variation between grasping postures, we added random noise (up to 10mm) to the x, y and z directions of the grasping posture. The sensor data from each sensor were normalised by the maximum value recorded across all tests for that specific sensor.

The dataset was split 80%-10%-10% for the purposes of training, validation and testing respectively and then augmented as described in Section II. It must be noted that the data splitting was performed before the data augmentation to ensure no grasps at validation/test time were observed at training time.

A video showing the data collection procedure is provided as supplementary material.

IV. RESULTS AND DISCUSSIONS

The EDAMS model was trained with the datasets described in the previous Section. We used the validation dataset containing all 10 grasps to tune the model's hyper-parameters, and empirically set $\gamma = 0.015$. We then fixate those hyper-parameters and re-train the network from scratch for each dataset of different grasps. Fig. 6 shows the training process for the model with the training and validation losses, along with the values of L_s .

After training, the model can map any high-dimensional sensor matrix X_i into its corresponding embedding u_i . The classification process takes less than a second for each set of grasps. For $R < 10$ grasps, a random selection of the 10 recorded grasps were chosen without replacement to ensure the same grasp pose was not repeated in a sample. Fig. 7(a) shows the clusters resulted from mapping the test set into its latent space representation, across the different number of grasps considered. In order to determine the maximum classification capability for each number of grasps, we compared simple classifiers to perform object recognition in the projected u space: linear support vector machine (SVM), decision tree (DT), random forest (RF), and k-Nearest-Neighbours (k-NN.) For the classifiers, from Python's *sci-kit learn*, we used *GridSearchCV* over common ranges of each of their hyper-parameters and used a *StratifiedKFolds* cross-validator for the final model selection with 4 folds. We report the best performing models for each number of grasps in Table I.

As shown in Table I, by increasing the number of grasps, we observe an increase in L_s , suggesting an

N grasps	L_s	Acc %	Classifier	Parameters
10	0.995	99.9	Linear SVM	$C = 0.001$
7	0.581	87.6	k-NN	$k=9$, weighted
5	0.547	83.2	Linear SVM	$C = 0.01$
3	0.275	70.0	Linear SVM	$C = 0.1$
1	0.014	47.0	Decision tree	Depth = 11

TABLE I. Silhouette score and mean classification accuracy computed on the test sets. For each number of grasps we reported the models and corresponding hyper-parameters that best performed on the validation sets.

increased ability for the trained embedded model to appropriately encode the information belonging to each object class, and achieve higher object class separation. In turn, that allows even simple classifiers to achieve increasingly more accurate object classification, doubling the performance from 47% to 99.9%. A more detailed view showing the separation of clusters in the low dimensional space is shown in Fig. 7(a).

Fig. 7(b) shows the confusion matrices for each number of grasps: the cube was most commonly misclassified as the deck of cards. The misclassification is likely due to local geometric similarities between the objects, considering the geometric profiles of the largest side of the cards compared with each of the cube's sides. The cause of the misclassification can also be traced back to the embedding. In Fig. 7(a), in fact, the representation in 3D u space of both the cube and the cards is overlapping across all number of grasps, only achieving some degree of separation at 10 grasps. We further observe how single grasps are indeed largely unreliable (Table I). In Fig. 7(b) we observe that beyond the misclassification between the cube and deck of cards, under the sensory system of a single grasp, other objects show a high degree of misclassification. This is for example the case for the bottle, misclassified as an apple $\approx 48\%$ of the times, and the cylinder, misclassified as an apple $\approx 25\%$ of the times. These objects, in fact, present similar local curvature information, leading to perceptual ambiguities hindering accurate tactile-based object recognition, and supporting our initial hypothesis. Finally, we observe how, although the deck of cards and the sponge are similar in geometry, the identification is capable of reliably distinguishing the two apart, suggesting success in this framework for discriminating between object stiffness.

V. CONCLUSION

This paper presents a framework to perform tactile object classification with a soft robotic hand and distributed tactile sensing by combining information from multiple grasps. We implemented a data pre-processing and encoder-decoder learning architecture to allow for accurate embedding and identification of objects without

reliance on the grasp pose sequence. Notably, we use a weighted silhouette score-based loss to find a low-dimensional encoding of the sensor data promoting the geometric separation of the objects. We applied machine learning classifiers to the encoded data in order to classify the objects and we showed the importance of multi-grasp sensory perception for object discrimination.

Future work will look to incorporate a Bayesian framework to instruct the hand where to grasp to confirm prior beliefs in the classification of a grasped object. This will reduce the number of grasps that are required to achieve a high classification accuracy, without restricting the ability to discern objects from any pose.

REFERENCES

- [1] B. Shih, D. Shah, J. Li, T. G. Thuruthel, Y.-L. Park, F. Iida, Z. Bao, R. Kramer-Bottiglio, and M. T. Tolley, "Electronic skins and machine learning for intelligent soft robots," *Science Robotics*, vol. 5, no. 41, apr 2020.
- [2] H. Wang, M. Totaro, and L. Beccai, "Toward Perceptive Soft Robots: Progress and Challenges," sep 2018.
- [3] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, no. August, pp. 54–67, 2017.
- [4] J. Huang and A. Rosendo, "Variable Stiffness Object Recognition with a CNN-Bayes Classifier on a Soft Gripper," *Soft Robotics*, vol. 00, no. 00, pp. 1–12, mar 2022.
- [5] L. Scimeca, P. Maiolino, and F. Iida, "Soft morphological processing of tactile stimuli for autonomous category formation," in *2018 IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 2018, pp. 356–361.
- [6] L. Scimeca, J. Hughes, P. Maiolino, L. He, T. Nanayakkara, and F. Iida, "Action augmentation of tactile perception for soft-body palpation," *Soft robotics*, vol. 9, no. 2, pp. 280–292, 2022.
- [7] C. Jiao, B. Lian, Z. Wang, Y. Song, and T. Sun, "Visual-tactile object recognition of a soft gripper based on faster Region-based Convolutional Neural Network and machining learning algorithm," *International Journal of Advanced Robotic Systems*, vol. 17, no. 5, pp. 1–13, 2020.
- [8] 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.
- [9] C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi, "A Century of Robotic Hands," *Robotics, and Autonomous Systems Annu. Rev. Control Robot. Auton. Syst.* 2019, vol. 22, pp. 1–32, 2019.
- [10] Z. Kappasov, J. A. Corrales, and V. Perdureau, "Tactile sensing in dexterous robot hands - Review," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, 2015.
- [11] J. Hughes, U. Culha, F. Giardina, F. Guenther, A. Rosendo, and F. Iida, "Soft manipulators and grippers: A review," p. 1, nov 2016.
- [12] S. Abondance, C. B. Teeple, and R. J. Wood, "A Dexterous Soft Robotic Hand for Delicate In-Hand Manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5502–5509, oct 2020.
- [13] J. Zhou, Y. Chen, D. C. F. Li, Y. Gao, Y. Li, S. S. Cheng, F. Chen, and Y. Liu, "50 Benchmarks for Anthropomorphic Hand Function-based Dexterity Classification and Kinematics-based Hand Design," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, oct 2020, pp. 9159–9165.
- [14] Y. Yan, C. Cheng, M. Guan, J. Zhang, and Y. Wang, "Texture Identification and Object Recognition Using a Soft Robotic Hand Innervated Bio-Inspired Proprioception," *Machines*, vol. 10, no. 3, p. 173, feb 2022.
- [15] L. Chin, J. Lipton, M. C. Yuen, R. Kramer-Bottiglio, and D. Rus, "Automated recycling separation enabled by soft robotic material classification," *RoboSoft 2019 - 2019 IEEE International Conference on Soft Robotics*, pp. 102–107, 2019.
- [16] B. S. Homberg, R. K. Katzschmann, M. R. Dogar, and D. Rus, "Robust proprioceptive grasping with a soft robot hand," *Autonomous Robots*, vol. 43, no. 3, pp. 681–696, 2019.
- [17] P. M. Khin, J. H. Low, M. H. Ang, and C.-h. Yeow, "In-Hand Object Recognition for Sensorized Soft Hand," in *International Conference on Intelligent Autonomous Systems*, ser. Lecture Notes in Networks and Systems, vol. 412. Springer International Publishing, 2022, pp. 351–364.
- [18] T. J. Pannen, S. Puhlmann, and O. Brock, "A Low-Cost, Easy-to-Manufacture, Flexible, Multi-Taxel Tactile Sensor and its Application to In-Hand Object Recognition," in *2022 International Conference on Robotics and Automation (ICRA)*, no. 390523135. IEEE, may 2022, pp. 10939–10944.
- [19] C. Tawk, H. Zhou, E. Sariyildiz, M. In Het Panhuis, G. Spinks, and G. Alici, "Design, Modeling and Control of a 3D Printed Monolithic Soft Robotic Finger with Embedded Pneumatic Sensing Chambers," *IEEE/ASME Transactions on Mechatronics*, vol. 4435, no. c, pp. 1–1, jul 2020.
- [20] M. Polic, I. Krajacic, N. Lepora, and M. Orsag, "Convolutional Autoencoder for Feature Extraction in Tactile Sensing," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3671–3678, oct 2019.
- [21] O. Shorthose, A. Albin, L. He, and P. Maiolino, "Design of a 3D-Printed Soft Robotic Hand With Integrated Distributed Tactile Sensing," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3945–3952, apr 2022.
- [22] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The GRASP Taxonomy of Human Grasp Types," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, 2016.
- [23] A. Kapandji, "Cotation clinique de l'opposition et de la contre-opposition du pouce," *Annales de Chirurgie de la Main*, vol. 5, no. 1, pp. 67–73, jan 1986.
- [24] A. Buryanov and V. Kotiuk, "Proportions of Hand Segments," *International Journal of Morphology*, vol. 28, no. 3, pp. 755–758, 2010.
- [25] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. C, pp. 53–65, nov 1987.
- [26] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy*, vol. 23, no. 6, p. 759, jun 2021.
- [27] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, dec 2014.
- [28] R. Zuo, Z. Zhou, B. Ying, and X. Liu, "A soft robotic gripper with anti-freezing ionic hydrogel-based sensors for learning-based object recognition," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 12 164–12 169.
- [29] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set," *IEEE Robotics & Automation Magazine*, vol. 22, no. 3, pp. 36–52, sep 2015.