Fast and Accurate Tactile Object Recognition using a Random Convolutional Kernel Transform

John Doherty[0000-0002-2167-6101], Bryan Gardiner[0000-0001-5642-6850], Emmett Kerr[0000-0002-4551-1272] and Nazmul Siddique[0000-0002-0642-2357]

Abstract— **The task of tactile object recognition is an ever-evolving research area comprising of the gathering and processing of features related to the physical interaction between a robotic system and an object or material. For a robotic system to be capable of interacting with the real-world, the ability to identify the object it is interacting with in real-time is required. Information about the object is often strongly enhanced using tactile sensing. Recent advancements in time series classifiers have allowed for the accuracy of real-time tactile object recognition to be improved, therefore generating opportunities for enhanced solutions within this field of robotics. In this paper, improvements are proposed to the state-of-the-art time series classifier ROCKET for analysis of tactile data for the purposes of object recognition. A variety of classifier heads are implemented within the ROCKET pipeline; these models are then trained and tested on the PHAC-2 tactile dataset, achieving state-of-the-art performance of 96.3% for single-modality tactile object recognition while only requiring 11 minutes to train.**

I. Introduction

Many researchers are investigating the potential of robotic systems to interact with humans and the environment in a growing manner. One way to aid in improving this interaction is to perform tasks such as object recognition and material classification. When physically interacting with objects, a robotic system can gather and utilise an extremely rich and dense collection of physical properties of the object through the use of haptic interaction and tactile sensing.

Due to the nature of gathering these physical properties at successive times, the data can be referred to as time series [1]. One of the challenges associated with gathering and processing time series data is the sheer volume of data gathered during physical contact, and the time required to physically gather this data. As such, object recognition has been a difficult task due to a lack of publicly available datasets that have sufficient data that can be utilised for training models with a high degree of accuracy.

This paper addresses a current research gap, where single-modality tactile object recognition systems are typically not trained and tested on multiple classes of objects. This work utilises the PHAC-2 dataset and is the first study to implement a variety of classifier heads into the ROCKET architecture to investigate the impact on the task of object recognition using tactile data on a dataset with a large number of objects (60 objects). This modified ROCKET implementation achieves state-of-the-art object recognition performance on the PHAC-2 dataset while only requiring a single modality to achieve peak performance, ensuring that both training and testing times are not compromised when utilising a complex dataset, enabling the proposed system to run in real-time.

II. Related Work

A. Tactile Datasets

Gathering tactile data is often an extremely time consuming and manual process [2]. As a result, many researchers opt to collect and use small-scale datasets which are subsequently not available for public usage [3]; as such, there is currently a lack of publicly available rich and dense tactile datasets.

VibTac-12 [4] is a collection of 12 texture classes of various household objects such as sandpaper, Velcro strips and rubber bands. The dataset features 20 second recordings for each texture sampled at 200Hz. The dataset also features corresponding audio data samples at 8kHz. The main limitations of this dataset when performing tactile research is that it only contains 12 classes, with each class only having one 20 second sample. This makes VibTac-12 unsuitable for any models requiring a large amount of data.

Penn Haptic Adjective Corpus 2 (PHAC-2) [5] is a visual-tactile multimodal dataset featuring a variety of data gathered on 60 different household objects. The tactile data within PHAC-2 is gathered by two individual Syntouch BioTac [6] tactile sensors mounted on a robotic gripper. Each class features 10 independent tactile readings and corresponding high-resolution images. Each class also includes related haptic adjectives which were determined by a selection of human participants. PHAC-2 is generally utilised for the training and testing of multimodal object recognition systems such as [7], where the visual data is combined with the tactile data to achieve higher performance.

B. Tactile Driven Classification

There are a variety of high performing time series classification models which are capable of functioning with a wide range of data; many of which focus on deep learning approaches. Schmitz et al. [8] developed a method utilising deep learning and dropout for the task of tactile object recognition; a multi-fingered hand combined with a shallow ANN to perform multimodal object recognition using the data from a TWENDY-ONE robotic hand [9]. Although the approach in this research performed well, there are some conditions for the performance of the model to be competitive.

We are grateful for access to the Tier 2 High Performance Computing resources provided by the Northern Ireland High Performance Computing (NI-HPC) facility funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant No. EP/T022175/1.

John Doherty, Bryan Gardiner, Emmett Kerr and Nazmul Siddique are with the Intelligent Systems Research Centre, Ulster University, Northland Rd, Derry, BT48 7JL.

Firstly, the data from four independent grasps is utilised, with each grasp lasting 20 seconds. Secondly, the reported performance of 88% is achieved on a 20 object dataset, a third of the object classes that is available in a dataset such as PHAC-2

Pastor et al. [10] also investigated the use of convolutional neural networks (CNN) by developing TactNet3D [11]. This research utilises two parallel grippers and a modified Tekscan tactile sensor to gather dense tactile information; this information is then fed to a 3D CNN. Again, this approach scales significantly with the number of grasps, achieving a performance of ~80% with only one grasp on a custom dataset containing 24 classes.

Prado da Fonseca et al. [12] also proposed a multi-grasp method of tactile object recognition driven by exploratory procedures. Multiple machine learning methods were tested, including Support Vector Machines [13] and Random Forests [14]. This approach scales significantly with the number of grasps, with the computational cost increasing greatly with each grasp. A peak performance of ~93% is achieved, however, the dataset consists of only six objects.

Huaping et al. [15] combined extreme learning machines and kernel sparse learning. Utilising the PHAC-2 dataset and tests a variety of models including support vector machines. This research focuses on determining the adjectives associated with each tactile reading in the PHAC-2 dataset and consists of several pre-processing steps, e.g., three of the 24 adjectives from the PHAC-2 dataset were removed and the data was split into 21 independent training and testing sets. While this approach achieved a mean accuracy of ~90%, it is not comparable to other state-of-the-art approaches as each model was trained to only detect one adjective, providing a solution that is not generalisable.

C. *Multi-modal Classification*

As aforementioned, PHAC-2 is a multi-modal dataset, containing multiple tactile modalities as well as visual data relating to each object. As such, there is a variety of implementations where multiple modalities are utilised for the task of adjective prediction and object recognition. Gao et al. [16] implemented a multi-modal approach utilising the PHAC-2 dataset to predict the haptic adjectives associated with input data. This implementation, which utilised multiple tactile modalities in a CNN alongside a GoogleNet image processor achieved a peak accuracy of 85.9%. Liu et al. [17] utilised a latent pairing matrix to perform multimodal fusion for object recognition with the PHAC-2 dataset. A variety of combinations of modalities were tested, and a peak performance of ~83% was achieved when fusing visual and tactile data. Notably, experimental work was also completed on a haptic-only object recognition method, where the accuracy ranged from 65% to 79%.

Abad et al. [18] utilised a GelSight [19] sensor combined with visual images and UV markings to perform object recognition on a collection of coins. This proposed method utilises AlexNet [20] for processing of the features and achieves a performance of 93.4% when evaluating a custom dataset containing only 5 coins.

Existing research into tactile object recognition currently has some major limitations. The majority of the existing models are trained and tested on custom datasets, and many of these datasets are extremely small, making comparisons between models difficult as the data they are trained and tested on is not comparable due to the extremely large variance in the quality of the dataset. In addition to this, many of the tactile object recognition models require multiple grasps with more than one tactile sensor gathering information. This approach has two main downsides; firstly, the time required to gather the tactile data for both training and testing will grow drastically with the required number of grasps. Secondly, the computational overhead required to process the rich tactile data gathered from multiple grasps combined with multiple sensors will consequently mean that many of these approaches will be unable to run in real time. As evident by the decrease in peak performance in multi-modal approaches, the task of creating one modal which can classify all 60 objects in the PHAC-2 dataset is considerably difficult when compared to models which are tested on much smaller and less complex datasets.

III. Methodology

A. *Dataset Selection*

It was decided to utilise the PHAC-2 dataset for this research. The rich and dense data gathered from the BioTac sensors provides an excellent source of information for tactile object recognition. Furthermore, PHAC-2 it is often selected within this field as a benchmark dataset for robotic tactile object recognition, introducing rigour due to the inclusion of a large set of object classes. In contrast to other work utilising PHAC-2, it was decided that all 60 objects would be included; providing a model capable of recognising an extremely varied number of objects. It should be noted, the haptic adjectives and visual data associated with each object are not utilised in this research as the focus is placed on the physical tactile data gathered from the two BioTac sensors.

PHAC-2 features a variety of modalities corresponding to the modalities captured by the BioTac sensors, including fingertip deformation, pressure, vibration and temperature readings. In this research, a focus was placed on the temperature and pressure readings due to the vibration and fingertip deformation modalities containing an extremely increased volume of data to process. For example, a vibration sample is 19 times the size of a temperature sample.

PHAC-2 contains tactile data gathered from two independent BioTac Sensors (BioTac-1 & BioTac-2), these sensors were mounted on a WillowGarage PR2 [21] gripper and in turn, gathered tactile data from opposite sides of the same object. There is a wide range of objects available for training, with each sensor gathering 600 total samples, generated from 10 individual samples for all 60 objects. This ensures that there is sufficient data to counteract overfitting.

Fig. 1 displays the readings generated from both BioTac-1 and BioTac-2 for two modalities, Thermal Flux (TAC) and Static Pressure (PDC). The readings across the two BioTac sensors are similar, but there are some noticeable differences. The more rounded peaks during the two points of contact indicate that BioTac-2 contacts the object slightly before BioTac-1, as well as maintaining contact for a longer period of

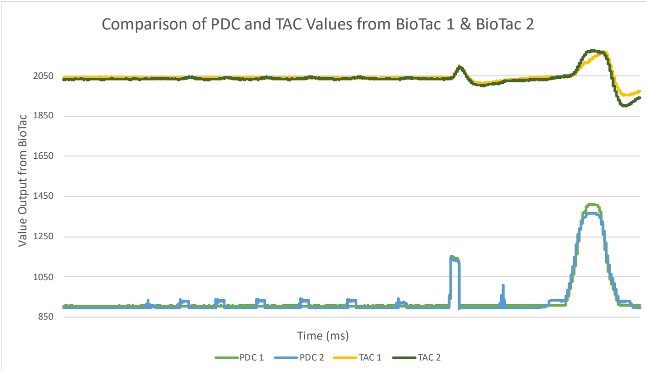

Fig 1. Comparison of PDC and TAC Values from BioTac-1 & BioTac-2

time. Alongside this, the TAC reading drops significantly more for BioTac-2 than BioTac-1 after this point of contact. The variance in the values reported from BioTac-1 and BioTac-2 outline the potential advantages which could be achieved by merging the datasets generated from these sensors, as the values are similar but not identical.

PHAC-2 was partitioned to create a 70/30 training/testing split. As each object class has 10 unique samples, the testing split was created by taking three samples from each class, ensuring that both the training and testing subsets contained perfect data balance with no overlap. This was completed independently for the samples in BioTac-1 and BioTac-2. These two independent datasets were then merged to create a third unique dataset, referred to as BioTac-1&2, containing 1200 samples, 20 from each class.

B. *Model Selection*

Due to the tactile data presented in the PHAC-2 dataset being treated as a collection of time series data, a focus was placed on finding an effective and efficient time series classifier which could be used for the task of tactile object recognition. Due to the nature of the tactile data being analysed, it was decided the Random Convolutional Kernel Transform (ROCKET) [22] family of models was chosen for integration into the proposed tactile object recognition pipeline. ROCKET is a state-of-the-art method of time series classification, utilising random convolutional kernels modelled upon the structure of CNNs. The kernels within the ROCKET model are randomised on the basis of length, weights, bias, dilation and padding [23]. Fig. 2 displays the proposed processing pipeline, where the features extracted from the random kernels are fed through to the chosen classifier, where output predictions are evaluated.

ROCKET has a variety of theoretical advantages when compared to other high performing time series classification methods. The main advantage of ROCKET is the greatly reduced computational cost, and hence reduction in both training and inference time compared to other CNN implementations. It is noted that ROCKET can train in a fraction of the time required by other state-of-the-art classifiers such as TS-CHIEF [24].

ROCKET does have some disadvantages which may make it unsuitable for specific time series classification tasks. The current implementations of ROCKET only support one modality during training, meaning if the user wishes to avail of multiple modalities simultaneously, it may be more suitable to investigate other models. Another disadvantage is that all time series samples are required to have the same number of steps, usually requiring some pre-processing of the data before presenting it to the model.

It was required to perform initial pre-processing of the PHAC-2 data, to ensure it is compatible with ROCKET. As aforementioned, ROCKET requires the steps in each time series sample to be identical lengths. This is not the default case in PHAC-2 as the time taken to collect each sample may vary. To ensure compatibility with ROCKET, it was required to downsample each sample to the same length as the shortest sample in the dataset.

C. *Classifier Selection*

Traditionally, a standard linear classifier head is used at the final stage of the ROCKET pipeline to perform the classification task on the features generated by the randomised kernels within the model. Within the proposed tactile time-series data processing pipeline, exploration of replacing the standard linear classifier with a variety of alternative classifiers was implemented, with the aim of evaluating the impact this will have on tactile data classification. Key characteristics of each of the classifiers utilised as part of the Head of the ROCKET pipeline are detailed below.

Logistic Regression is implemented as a linear model for classification. With Logistic Regression, regularisation is implemented by default and a variety of solvers are utilised depending on whether it is a binary or multi-class classification problem.

Ridge Classifier [25] converts binary targets to -1, 1 and then treats the problem as a regression task. Ridge Classifier uses a MSE + L2 penalty loss function; the advantage of this classifier is that it can effectively shrink coefficient estimates reducing overfitting on complex datasets. Ridge Classifier CV [26] is a modified version of Ridge Classifier which features in-built leave-one-out cross-validation.

Random Search [27] implements a fit and score method, where randomised searches are performed over parameters. A

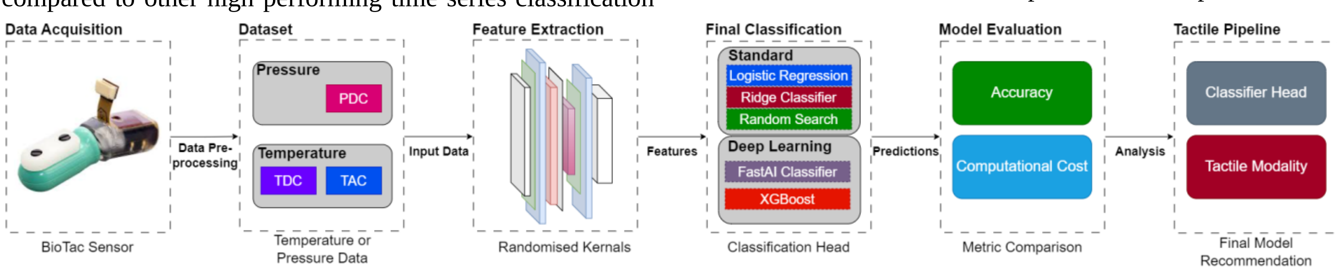

Fig 2. Proposed Processing Pipeline for Tactile Object Recognition

Random Search classifier can be implemented by adding a variety of epsilon values to a Logistic Regression classifier. Random Search can be used on functions that are not differentiable or continuous, and as such, the main advantages are the ease of tuning with little-to-no increase in computational cost.

Alongside these classifiers, two further classifiers that utilise a neural network structure to perform more significant processing of the features were selected, namely FastAI Classifier and XGBoost.

FastAI Classifier, developed by Howard et al. [28] is a neural network head which utilises linear framework and a ReLU [29] activation function. As this head is based on a neural network containing multiple layers, it is required to perform multiple epochs of training to achieve peak performance, requiring significantly more training time than the three previously selected classifiers.

XGBoost is an open-source tree boosting head developed by Chen et al. [30]. XGBoost utilises tree boosting combined with a sparsity-aware algorithm and weighted quantile sketch for tree learning. Like the FastAI Classifier, XGBoost will require training across multiple epochs which will considerably increase the training time.

IV. EXPERIMENTS & RESULTS

An individual ROCKET model was developed for all five classifier heads previously outlined: Logistic Regression (default head), Ridge Classifier CV, Random Search, FastAI Classifier and XGBoost. It should be noted that all classifiers utilise their standard variables. In the case of FastAI Classifier and XGBoost, these models require multiple epochs of training, so are trained until peak accuracy is achieved.

The data from two independent BioTac sensors were utilised within these experiments. Firstly, the models were trained and tested only on the data gathered from BioTac-1. Following this, the models' weights were reset, and they were then trained and tested on the data from BioTac-2. The final experiments involved combining the data from both BioTac-1 and BioTac-2, creating an augmented dataset consisting of 1200 samples. During training and testing, the proposed models treated these 1200 samples as if they came from the same BioTac, i.e., not processing these samples in parallel.

Table I outlines the initial experimental results using the data from BioTac-1 for training and testing. Thermal Flux (TAC) is a filtered version of the DC temperature; it can be measured as the heat flow between the heated BioTac and the object. Temperature (TDC) refers to the heat difference between the object and the BioTac but does not measure heat flow, while Fluid Pressure (PDC) refers to the average static pressure of the sensor, increasing linearly with fluid pressure.

Training time refers to the total training time including feature generation and classifier training. The training time is measured in minutes and seconds, and all models were trained on a Nvidia V100 GPU. The default ROCKET head, Logistic Regression, is italicised and all reported accuracies for TAC, TDC and PDC are the accuracies calculated when the proposed ROCKET model was tested on the entire 30% training split of the outlined BioTac data.

TABLE I. BIOTAC-1 EXPERIMENTAL RESULTS

Classifier	***TAC Accuracy***	***TDC Accuracy***	***PDC Accuracy***	***Training Time (min:sec)***
Logistic Regression	*0.783*	*0.508*	*0.867*	*5:34*
Ridge Classifier CV	0.850	0.583	**0.933**	5:20
Random Search	0.767	0.542	0.891	8:33
FastAI Classifier	0.508	0.583	0.833	24:47
XGBoost	0.758	0.458	0.783	72:01

Table I shows that a peak accuracy of 93.3% is achieved when ROCKET is combined with a Ridge Classifier CV trained and tested on PDC. On average, it is also clear that ROCKET performs significantly higher with PDC data when compared to both TAC and TDC. This is a notable improvement of 6.6% over the standard ROCKET head with no computational cost impact.

Across the five classifiers, it is evident that the neural network based approaches, i.e., the FastAI Classifier and XGBoost do not perform as well as the other selected classifiers. This is due to the number of samples not being large enough to facilitate training these networks, as there are only 600 samples available from the BioTac-1 data. Another disadvantage of the neural network approaches is the significant increase in computational overhead required. As can be seen in Table I, training times increase from 5 minutes with the top performing Ridge Classifier CV, to over 70 minutes with XGBoost, while demonstrating a loss of 15% accuracy. Based on this computational cost alongside the severe drop in accuracy, classifiers based on neural networks cannot be recommended for the processing and classification of tactile data in a relatively small dataset.

Table II presents the results from the experimental work completed on BioTac-2 data only across all proposed ROCKET heads on the PHAC-2 dataset. As can be seen in Table II, the results mirror what was discovered during the initial experiments on BioTac-1. The training times for each model across both BioTac datasets were identical, which is to be expected as the datasets are the same size.

The peak performance for the same models trained and tested on BioTac-2 are slightly lower, achieving a peak accuracy of 92.5% when combining a Ridge Classifier with the PDC data. This is marginally lower than the 93.3% that was achieved using the BioTac-1 data. The improvement achieved over the standard Logistic Regression was 7.5%, similar to the improvement seen in Table I. The two neural network based classifiers again fell short in both accuracy and training time when compared to the more heavily used standard classifiers. However the disparity between these types of classifiers was not as large as it was with data collected from BioTac-1.

The accuracies for almost all classifiers are noticeably higher on both the TAC and TDC sections of the dataset when compared with Table I. This likely indicates that the position of the BioTac-2 sensor is much more suited to gathering temperature data when compared with BioTac-1.

TABLE II. BIOTAC-2 EXPERIMENTAL RESULTS

Classifier	***TAC Accuracy***	***TDC Accuracy***	***PDC Accuracy***	***Training Time (min:sec)***
Logistic Regression	*0.800*	*0.600*	*0.850*	*5:03*
Ridge Classifier CV	0.883	0.617	**0.925**	5:47
Random Search	0.842	0.617	0.842	8:48
FastAI Classifier	0.842	0.650	0.808	25:23
XGBoost	0.725	0.500	0.767	70:14

The temperature sensor is located near the tip of the BioTac, so it is assumed that BioTac-2 is making contact with the object at a much more suitable angle for the thermistor to collect BioTac-1 data.

Table III displays the results from the final experiments, where the data from both BioTac-1 and BioTac-2 are combined to create a unique set of data – this will be referred to as BioTac-1&2. The models will not treat this data as two independent input signals, but rather as one input signal with twice as many samples. Mirroring the results from Table I and Table II, the highest performing combination across the previous experimental work, Ridge Classifier CV and PDC, achieving an accuracy of 96.3% across all 60 objects while only requiring a training time of 11 minutes. This demonstrates state-of-the-art performance for object recognition on the PHAC-2 dataset and far exceeds other object recognition models due to the vast number of objects which can be detected at this accuracy using tactile data only. Not only does the best performing combination of ROCKET and Ridge Classifier CV see improvements, but all combinations of classifier and data type see improvements when the data from both BioTacs are used in unison. Improvements are as large as 37.5% are evident when comparing to the FastAI Classifier utilising TAC. More commonly, there are improvements in the range of 8-15% with examples across all classifier and data combinations.

One important note is that the classifiers which benefit most from the increased size of the dataset due to the amalgamation of BioTac-1 & BioTac-2 are the neural network based classifiers: FastAI Classifier and XGBoost. This is to be expected as it was theorised that these models performed worse in earlier experiments due to a lack of sufficient data when only utilising the data from one BioTac sensor. Merging BioTac-1&2 effectively doubles the amount of data available for the models to train on. Importantly, not only are the accuracies of the models trained on the culmination of BioTac-1&2 higher, but the models themselves can be assumed to be more robust as the data gathered from each BioTac occurred on separate sides of the same object, helping to alleviate overfitting. The training times for the more simplistic classifiers, Logistic Regression, Ridge Classifier CV and Random Search scale linearly with dataset size.

In the case of FastAI Classifier and XGBoost, this is not the case. The neural network classifiers, FastAI Classifier and XGBoost on average triple in training time, from 25 to 65 minutes and 70 to 220 minutes respectively. However, both classifiers see a substantial accuracy improvement of 10.8% for FastAI Classifier and 13% for XGBoost during testing

TABLE III. BIOTAC-1&2 EXPERIMENTAL RESULTS

Classifier	***TAC Accuracy***	***TDC Accuracy***	***PDC Accuracy***	***Training Time (min:sec)***
Logistic Regression	*0.900*	*0.725*	*0.921*	*10:57*
Ridge Classifier CV	0.883	0.688	**0.963**	11:09
Random Search	0.898	0.700	0.929	20:02
FastAI Classifier	0.883	0.713	0.938	65:19
XGBoost	0.792	0.642	0.875	220:27

It is believed the increase in performance achieved by implementing the Ridge Classifier CV head is due to the nature of the dataset. The data gathered from the BioTac sensors presented in the PHAC-2 dataset is relatively complex and highly correlated, demonstrating the advantages of this specific classifier. Furthermore, the current limitation of the neural network heads, FastAI Classifier and XGBoost is the number of samples. While PHAC-2 is a large dataset when compared to other publicly available tactile datasets, it is still small relative to other time-series datasets such as those in the UCR [31] collection.

Across all experimental work completed, the combination of ROCKET paired with a Ridge Classifier CV head was consistently the top performing model, achieving accuracy improvements over the standard Linear Regression head of up to 7.5% with virtually no increase to the computational overhead. The proposed single-modality model outperforms existing state-of-the-art multi-modal models for the task of object recognition, and it is assumed the computational costs are notably reduced as only one modality is utilised. Alongside this, the proposed model can be implemented on any dataset derived from the BioTac sensor. Therefore, based on the findings of this research, researchers seeking to perform tactile object recognition are advised to consider utilising the proposed ROCKET pipeline.

V. CONCLUSION

The detailed experimental work concluded during this research allows for a variety of conclusions to be drawn:

- PHAC-2 is an excellent dataset to be used for the task of tactile object recognition training and testing.
- If utilising the PHAC-2 dataset, it is beneficial to combine the samples from BioTac-1 & BioTac-2 to create a dataset twice the size discouraging overfitting of models while improving accuracy.
- ROCKET is an effective algorithm for processing time series tactile data to generate features in an efficient manner which can be used to train a variety of classifiers.
- Standard classifiers are still measurably more efficient and effective at the task of tactile object recognition when paired with ROCKET to generate features than with the use of neural network based classifiers. The use of neural network and deep learning-based classifiers is not recommended without a suitably rich

and dense dataset for training, or the use of a suitable transfer learning approach.

- ROCKET, paired with a Ridge Classifier CV achieves state-of-the-art tactile object recognition performance, achieving an accuracy of 96.3% on the PHAC-2 dataset while training for approximately 10 minutes on a V100 GPU.

A variety of future work is planned, continuing the research into the ROCKET family of models for tactile research. Firstly, research will continue on the implementation of multimodal tactile object recognition. PHAC-2 is a rich dataset consisting of many modalities which would pair well together such as a combination of the TAC and PDC data. Secondly, the development of a novel classifier which pairs well with a variety of these ROCKET-based models will be investigated. From this research it appears that neural network approaches will not pair well with the randomised backend of ROCKET without an adequate dataset to support this which is not yet publicly available.

REFERENCES

[1] H. Orii, S. Tsuji, T. Kouda, and T. Kohama, “Tactile texture recognition using convolutional neural networks for time-series data of pressure and 6-axis acceleration sensor,” in *2017 IEEE International Conference on Industrial Technology (ICIT)*, IEEE, Mar. 2017, pp. 1076–1080. doi: 10.1109/ICIT.2017.7915511.

[2] J. Schill, J. Laaksonen, M. Przybylski, V. Kyrki, T. Asfour, and R. Dillmann, “Learning continuous grasp stability for a humanoid robot hand based on tactile sensing,” in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, IEEE, Jun. 2012, pp. 1901–1906. doi: 10.1109/BioRob.2012.6290749.

[3] H. Liu, Y. Wu, F. Sun, and D. Guo, “Recent progress on tactile object recognition,” *Int J Adv Robot Syst*, vol. 14, no. 4, p. 172988141771705, Jul. 2017, doi: 10.1177/1729881417717056.

[4] O. Kursun and A. Patooghy, “An Embedded System for Collection and Real-Time Classification of a Tactile Dataset,” *IEEE Access*, vol. 8, pp. 97462–97473, 2020, doi: 10.1109/ACCESS.2020.2996576.

[5] V. Chu *et al.*, “Robotic learning of haptic adjectives through physical interaction,” *Rob Auton Syst*, vol. 63, pp. 279–292, Jan. 2015, doi: 10.1016/j.robot.2014.09.021.

[6] J. Fishel, N. Wettels, and Z. Su, “Multi-modal synergistic tactile sensing,” in *IEEE-RAS International Conference on Humanoid Robots*, 2009.

[7] M. Bednarek, P. Kicki, and K. Walas, “On Robustness of Multi-Modal Fusion—Robotics Perspective,” *Electronics (Basel)*, vol. 9, no. 7, p. 1152, Jul. 2020, doi: 10.3390/electronics9071152.

[8] A. Schmitz, Y. Bansho, K. Noda, H. Iwata, T. Ogata, and S. Sugano, “Tactile object recognition using deep learning and dropout,” in *2014 IEEE-RAS International Conference on Humanoid Robots*, IEEE, Nov. 2014, pp. 1044–1050. doi: 10.1109/HUMANOIDS.2014.7041493.

[9] H. Iwata and S. Sugano, “Design of anthropomorphic dexterous hand with passive joints and sensitive soft skins,” in *2009 IEEE/SICE International Symposium on System Integration (SII)*, IEEE, Nov. 2009, pp. 129–134. doi: 10.1109/SI.2009.5384542.

[10] F. Pastor, J. M. Gandarias, A. J. García-Cerezo, and J. M. Gómez-De-gabriel, “Using 3D Convolutional Neural Networks for Tactile Object Recognition with Robotic Palpation,” *Sensors 2019, Vol. 19, Page 5356*, vol. 19, no. 24, p. 5356, Dec. 2019, doi: 10.3390/S19245356.

[11] K. N. Bachus, A. L. DeMarco, K. T. Judd, D. S. Horwitz, and D. S. Brodke, “Measuring contact area, force, and pressure for bioengineering applications: Using Fuji Film and TekScan systems,” *Med Eng Phys*, vol. 28, no. 5, pp. 483–488, Jun. 2006, doi: 10.1016/j.medengphy.2005.07.022.

[12] V. Prado Da Fonseca, · Xianta Jiang, E. M. Petriu, · Thiago, and E. Alves De Oliveira, “Tactile object recognition in early phases of grasping using underactuated robotic hands,” *Intell Serv Robot*, vol. 15, pp. 513–525, 2022, doi: 10.1007/s11370-022-00433-7.

[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers”.

[14] L. Breiman, “Random forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.

[15] H. Liu, J. Qin, F. Sun, and D. Guo, “Extreme Kernel Sparse Learning for Tactile Object Recognition,” *IEEE Trans Cybern*, vol. 47, no. 12, pp. 4509–4520, Dec. 2017, doi: 10.1109/TCYB.2016.2614809.

[16] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, “Deep learning for tactile understanding from visual and haptic data,” in *2016 IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, May 2016, pp. 536–543. doi: 10.1109/ICRA.2016.7487176.

[17] H. Liu, Y. Wu, F. Sun, B. Fang, and D. Guo, “Weakly Paired Multimodal Fusion for Object Recognition,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 784–795, Apr. 2018, doi: 10.1109/TASE.2017.2692271.

[18] A. C. Abad and A. Ranasinghe, “Low-cost GelSight with UV Markings: Feature Extraction of Objects Using AlexNet and Optical Flow without 3D Image Reconstruction,” *Proc IEEE Int Conf Robot Autom*, pp. 3680–3685, May 2020, doi: 10.1109/ICRA40945.2020.9197264.

[19] M. K. Johnson and E. H. Adelson, “Retrographic sensing for the measurement of surface texture and shape,” pp. 1070–1077, Mar. 2010, doi: 10.1109/CVPR.2009.5206534.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Accessed: Jun. 27, 2023. [Online]. Available: http://code.google.com/p/cuda-convnet/

[21] S. Cousins, “ROS on the PR2,” *IEEE Robot Autom Mag*, 2010, doi: 10.1109/MRA.2010.938502.

[22] A. Dempster, F. Petitjean, and G. I. Webb, “ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels,” *Data Min Knowl Discov*, vol. 34, no. 5, pp. 1454–1495, Sep. 2020, doi: 10.1007/s10618-020-00701-z.

[23] “Demo of ROCKET transform,” *SKTime*, Feb. 13, 2023. https://www.sktime.org/en/v0.10.1/examples/rocket.html (accessed Feb. 19, 2023).

[24] A. Shifaz, C. Pelletier, F. Petitjean, and G. I. Webb, “TS-CHIEF: A Scalable and Accurate Forest Algorithm for Time Series Classification,” Jun. 2019, doi: 10.1007/s10618-020-00679-8.

[25] “sklearn.linear_model.RidgeClassifier,” *SKLearn*, Feb. 13, 2023. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeClassifier.html (accessed Feb. 19, 2023).

[26] “sklearn.linear_model.RidgeClassifierCV,” *SKLearn*, Mar. 01, 2023. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeClassifierCV.html (accessed Mar. 01, 2023).

[27] “sklearn.model_selection.RandomizedSearchCV,” *SKLearn*, Feb. 12, 2023. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html (accessed Feb. 19, 2023).

[28] J. Howard, “fastai,” *Github*, 2018. https://github.com/fastai (accessed Feb. 19, 2023).

[29] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” Mar. 2018.

[30] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[31] H. A. Dau *et al.*, “The UCR Time Series Classification Archive,” 2019. www.cs.ucr.edu/~eamonn/time_series_data/ (accessed Feb. 19, 2023).