Continuous learning of contact episodes from proprioceptive sensors in industrial assembly scenarios using Adaptive Resonance Theory

Vinzenz Bargsten[1] [0000−0001−8884−2103], Dimitrios Rakovitis[1], Vamsi Krishna Origanti[1], Adrian Danzglock[1], and Frank Kirchner[1,2]

[1] DFKI (German Research Center for Artificial Intelligence), Bremen, Germany
[2] Workgroup Robotics, University of Bremen, Germany
vinzenz.bargsten@dfki.de

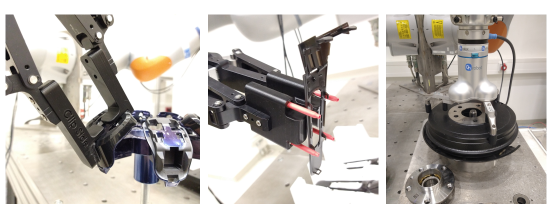
Abstract. Robotic automation plays a crucial role in modern industrial production, yet many assembly tasks still require manual intervention. Unlike standard pick-and-place operations, which can be executed with position-controlled manipulator arms, assembly tasks inherently involve physical contact between components, requiring precise force and torque management. Accurately assessing whether tightly fitting parts are correctly aligned or whether flexible components (e.g., plastic covers, cables) are properly assembled is challenging due to inevitable uncertainties in material properties and positioning. Traditional solutions rely on hand-crafted thresholds set by experts, which are costly and impractical for frequently changing product variants. To address this challenge, we present a machine learning approach based on Adaptive Resonance Theory (ART) for real-time, continuous learning and adaptation of contact episodes using proprioceptive sensor data. Our method processes joint torque and end-effector force-torque measurements, encoding these time-series signals into a frequency domain representation using Short-Time Fourier Transform (STFT). The ART-based module dynamically classifies contact patterns, identifying the most suitable learned category while detecting novel situations that deviate from prior experience, enabling adaptive control strategies. The proposed approach provides a scalable and cost-effective solution by reducing reliance on predefined heuristics and enabling online adaptation to new product configurations. The system is experimentally validated in two industrial assembly scenarios, where it demonstrates robust classification accuracy, real-time responsiveness, and adaptability compared to static threshold-based methods. Results highlight its potential for seamless integration into industrial control workflows, allowing robots to autonomously adjust assembly strategies or escalate novel contact situations for further inspection.

Keywords: FuzzyART · contact classification · robotic assembly · match-based learning · flexible manufacturing.

1 Introduction

The classification of contacts between robotic systems and their environment is still an active field of research, as it is of great relevance for robotic manipulation of objects and motion in the environment. The main topics can be roughly divided into two classes, dealing with unintentional contacts and intentional contacts. The former case often involves the detection of collisions in order to avoid damage and injury. The focus of this paper, however, is about the second case, which focuses on intentional contacts such as those that occur in industrial assembly tasks.

In the recent years, some methods have been proposed to distinguish between the two classes. For example, in [7] the rate of change in torque measurement is used to distinguish collision from interaction. Methods working in the frequency domain have advanced this concept such as in [10, 11] by analyzing the frequency magnitudes of the signal instead. Considering assembly tasks, potentially more than two states have to be distinguished. Regarding robotic assembly, in a recent work [15] a method was proposed to improve an assembly task on the kinematic level. It suggests improving the object localization accuracy by combining CAD model of the part, tactile knowledge, and a particle simulation. On the other hand, some works with a different focus than assembly show the great potential of classification using joint torque measurements [9] or acoustic vibration sensing [13]. However, these methods do not allow a continuous learning and adaption from data during operation but rely on experts for setup, data acquisition, labeling, training with each modification of the task. The aim of this study is to make robotic systems more autonomous and thus less reliant on costly adaptions by experts. For this purpose it is necessary to reduce their dependence on models, to deal with inaccuracies and to enable them to continuously learn and adapt themselves. Therefore, the method presented in this work is based on Adaptive Resonance Theory (ART). This theory originated in cognitive science and describes neural network dynamics that continuously learn without being affected by catastrophic forgetting. In a preliminary study [4] with a focus on collisions and novelty detection, a principle feasibility has already been investigated. This



(a) (b) (c)

Fig. 1. (a),(b): Plastic parts being considered in one of the two industrial assembly scenarios. (c): Drum part being grasped, belonging to the second industrial scenario of automotive disc brake and drum brake assemblies

work extends the approach in the context of industrial assembly, while being part of a larger framework. The industrial assembly tasks being considered relate to two scenarios which still rely on manual work. One scenario deals with the mounting of plastic shaver parts of different variants onto a fixture prior to a lacquering process. The second scenario deals with the assembly of automotive parts for a disc brake assembly and a drum brake assembly. The parts are illustrated in Fig. 1.

The outline of the paper is as follows: Sec. 2 describes the methods, Sec. 3 presents experimental evaluation, and Sec. 4 discusses the results and concludes the study.

2 Method Description

The method is a composition of two elements, an ART neural network and a pre-processing step transforming the input data to frequency domain. These two elements are explained in the following paragraphs.

2.1 Adaptive Resonance Theory

ART's operating principle has been inspired by the theories developed in cognitive sciences. In contrast to error-based learning such as backpropagation, ART describes a match-based learning method [3, p. 87]. For this purpose, the main elements in an ART neural network are an *input representation field*, $\mathbf{F}_1$, a *category representation field*, $\mathbf{F}_2$, the *bottom–up* weight vectors from $\mathbf{F}_1$ to $\mathbf{F}_2$, the *top–down* weight vectors in opposite direction, and an *orienting subsystem*. With these elements, the general operating principle starts with the input representation layer presenting the input to the category representation field. There, all of its *nodes* that represent the actual categories (classes) begin to determine their individual activation level for the provided bottom–up weights. A *competition* is then started by the orienting subsystem. Beginning with the node having the highest activation, the node is activated and presents its expectation through the top–down weight vector to the input representation field. Here, the last step of the cycle is conducted by the orienting subsystem: a *vigilance* test based on the similarity. If that test is passed by the node, it can learn from the input by adapting its weights. Otherwise, the node is inhibited and the node having the next highest activation selected. If none of the existing nodes pass the test, the neural network grows by adding a new (uncommitted) node and committing it to store the new information.

Many variants of ART classification algorithms have been developed in the last decades [12]. The various algorithmic implementations typically differ in metrics used to measure similarity and the input encoding being applied. *FuzzyART* ([5, 6]) gained some particular interest and many extensions have been proposed. Since it uses fuzzy set operations for the similarity measure, the category representation and learning dynamics have some outstanding properties. Specifically, it starts with a real-valued input vector $\mathbf{x} \in \mathbb{R}^n,\ 0 \leq x_i \leq 1\ \forall i$

to be *complement coded* for numerical stability[3], forming a new input vector $\mathbf{I} = (\mathbf{x}, \mathbf{1} - \mathbf{x})^T$. The encoded input vector $\mathbf{I}$ is then presented through the $\mathbf{F}_1$ layer to the $\mathbf{F}_2$ layer, where the nodes compete for the highest activation according to the activation function,

$$T_j = \frac{\left\| \mathbf{I} \wedge \mathbf{w}_j \right\|_1}{\alpha + \left\| \mathbf{w}_j \right\|_1}, \tag{1}$$

where $\|\cdot\|_1$ refers to the L_1 norm, $\mathbf{w}_j$ are the weights of the category j template, and α modifies the preference for uncommitted nodes. The weight vector serves as memory and is composed of the stored weights w_{ij}, which in this ART variant are combined bottom–up and top–down weights. The operator $\wedge$ is the fuzzy set *AND* operator, i.e., an intersection, and is defined here as the element-wise minimum, $\mathbf{a} \wedge \mathbf{b} \equiv (\min(a_0, b_0), \ldots, \min(a_i, b_i), \ldots)^T$. The winner of the competition is activated by the orienting subsystem and has to overcome the resonance criterion, $M_j \geq \rho$, in order to be successfully selected, where $0 \leq \rho \leq 1$ is the vigilance parameter controlling the granularity of the categories. The match value M_j is computed by the match function,

$$M_j = \frac{\left\| \mathbf{I} \wedge \mathbf{w}_j \right\|_1}{\left\| \mathbf{I} \right\|_1}. \tag{2}$$

If the resonance criterion was not satisfied, the winning node is inhibited and the category with the next highest activation is tested for resonance. If no resonance occurred at all, a new category is created based on the data from the input sample. In case of passing the resonance test, the category weights $\mathbf{w}_j^{\text{old}}$ are adapted using the input sample $\mathbf{I}$ according to

$$\mathbf{w}_j^{\text{new}} = (1 - \beta)\mathbf{w}_j^{\text{old}} + \beta\, (\mathbf{I} \wedge \mathbf{w}_j^{\text{old}}), \tag{3}$$

where $0 \leq \beta \leq 1$ denotes the learning rate. In conjunction with the complement coding, the fuzzy set intersection operation $(\cdot \wedge \cdot)$ in the adaption results in characteristic properties [2, 14]: The category representation region is of a *hyperrectangular* shape. When adapting to new inputs outside its current representation region, the size of the category hyperrectangle grows according to the learning rate. Interestingly, a growth of the category hyperrectangle also results in a shrinkage of the match region. Overall, this allows to adapt to variably sized clusters with stabilizing learning dynamics. However, its hyperrectangular category shapes cannot adapt well to arbitrary cluster shapes. Therefore, in this work, the recently proposed variant *DistributedDualVigilanceFuzzyART* (DDVFA) [16, 17] has been implemented for the classification itself. It describes a multi-layered ART neural network, which is able to learn category templates represented by individual *FuzzyART*[4] neural networks as global $\mathbf{F}_2$ nodes themselves. This enables the learning of groups of classes and thus arbitrary shapes of patterns in the data can be represented.

[3] To counter the problems termed *weight erosion* and *category proliferation* in the work of Grossberg and Carpenter, arising from applying the fuzzy set AND-operator

[4] Or *GammaFuzzyART* with an additional exponential decay factor

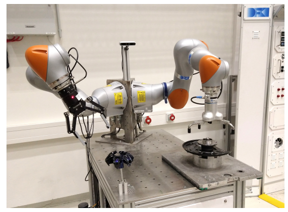


Fig. 2. The laboratory setup composed of two manipulator arms KUKA iiwa R820, Robotiq 2-finger grippers with custom fingers for the use case of consumer product manufacturing, an OnRobot 3FG25 3-finger gripper for heavy cylindrical parts in an automotive assembly use case, and end-effector force-torque sensors.

2.2 Input Data Encoding

As input data to the ART neural network, either joint torque measurements or end-effector force-torque measurements are used. However, to encode this stream of time series data samples into a suitable form for classification, the stream of raw data samples are continuously transformed into frequency domain using the method *Short Time Fourier Series* (STFT) [1, 8]. For this purpose, the time domain samples are gathered into a window on which a *Blackman* filter is applied in this implementation. The filtered window is then transformed into frequency domain by calculating the Discrete Fourier Transform (DFT). This results in a significant compression of the data in comparison to the original signal. By choosing the window size and overlap, the amount of signal history is implicitly encoded for the classification. Optionally, the discrete difference of the signal can be used for classification to enhance the higher frequency spectrum. The ART neural network then learns category templates which allow to distinguish the different episodes represented by the frequency domain data that result from the manipulator arm performing its commanded task. For an investigation of the effects of the individual choices of vigilance parameters for the global and local ART layers, sample window size and overlap, as well as frequency range we refer to the preliminary study [4]. The method has been implemented as a software node in the ROS2 framework, that outputs the assigned class as well as a possible mismatch on new input to the control layer. This node is integrated with a Learning-from-Demonstration approach, which allows users to teach new assembly motions to the system, as well as an adaptive model predictive control approach that takes into account the changes in category assignment.

3 Experimental Evaluation

3.1 Experimental Setup

The laboratory setup consists of a workbench with two manipulator arms shown in Fig. 2. In addition to the end-effector force-torque measurement, the manipulator arms also provide link-side torque measurements. The system is controlled using the ROS2 framework.

3.2 Experimental Results

Due to the limited space available, only one isolated sub-task will be presented here to serve as an example. In particular, the drum of the brake shown in Fig. 1c is repeatedly lifted and lowered down on wheel hub and brake shoe assembly by the manipulator arm. During this process, there are only small tolerances and the brake shoes may grind on the inside of the drum. The aim is to distinguish this normal operation from the situation in which the parts become jammed, thus avoiding excessive forces by starting a new trial. Such a case has been introduced at time 204 s in the data shown in Fig. 3, resulting in a higher change of the force signal than in the previous and following motions. Specifically, the forces at the end-effector have been measured and the L_2 norm of the force vector has been computed and differentiated. This time domain signal and its resulting frequency domain counterpart obtained from the windowed

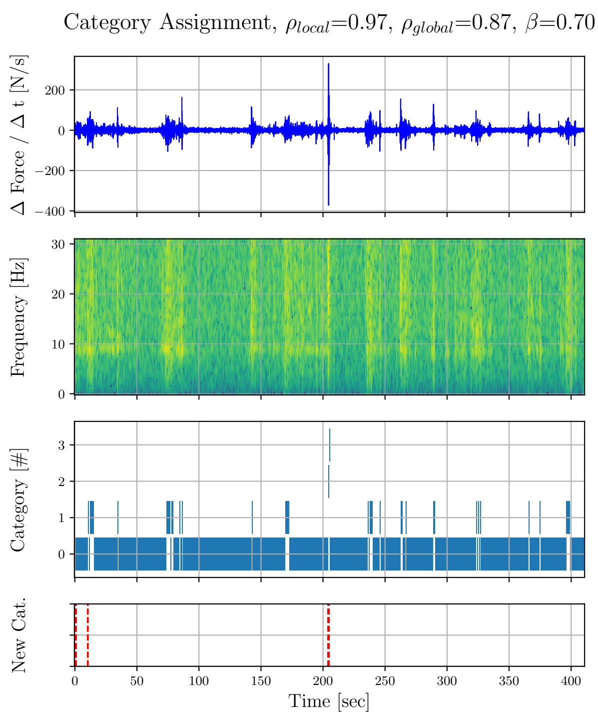


Fig. 3. Experimental results showing from top to bottom the time series signal, the windowed frequency domain representation of the signal, the category template assigned to the frequency vectors, and an indication of new category creation (data did not match previously learnt category template). The data correspond to about ten repeated lifting/lowering movements of the drum onto the brake hub. In one trial at time 204 s the parts do not align and begin to get stuck, which is detected by the ART neural network because of the dissimilarity of the data to the categories learnt so far.

Fourier transformation is shown in the first and second plots of Fig. 3. The third plot then shows the assigned category, resembling the episodic pattern in the frequency spectrum, while the lines in the bottom plot indicate commitment of a new node when no existing category template matches the input. The parameters have been determined through a prior grid search.

4 Discussion and Conclusion

A combination of a continuous machine learning method and frequency domain pre-processing of proprioceptive sensory data of a robotic manipulator arm has been proposed for finding episodes in industrial assembly scenarios. The employed ART neural network allows to incrementally learn new data and adapt existing category templates in a two-staged mode of competition and matching expectation against inputs. In contrast to threshold-based methods, the use of a vigilance parameter allows to control the level of similarity in the learnt classes. The experimental results indicate its capability to learn and classify force measurements, as well as its ability to capture rare events. Though in the case presented the larger spike in the signal is directly visible, the method has also shown its potential in cases where that is less clear. A successful transfer to a commercial manipulator arm considering actual industrial use cases has been shown. If precise joint torque measurements are available, they can equally be used for classification. Since many industrial systems do not provide such measurements the method has been validated to work with force-torque sensors. However, there is a limitation in that only short episodes (and no sequences) can be registered through the implicit coding of time series data to the frequency domain.

As the ART-based neural network inherently allows online learning, new data can be integrated in real-time and thus does not need prior data collection for an offline training. Therefore, it is a promising candidate to allow robotic applications to become more aware and adaptive to changes in tasks on their own, reducing the need for re-programming or training by experts.

An ongoing work focuses on integration of the method with an adaptive model predictive control approach for contact-rich assembly tasks. Here, the ART-based classification provides and additional measure of uncertainty which is used to modify the control behavior.

Acknowledgments This work is conducted as part of the HARTU [5] project, funded from the European Union's research and innovation programme Horizon Europe under the grant agreement No. 101092100. We thank Tofaş Türk Otomobil Fabrikası A.Ş. (TOFAS) and Philips Consumer Lifestyle BV (PCL) for their support with the industrial scenarios and for providing the product samples. The authors declare that there is no conflict of interest regarding the publication of this paper.

[5] HARTU – Handling with AI-enhanced Robotic Technologies for Flexible Manufacturing https://www.hartu-project.eu/

References

1. Allen, J., Rabiner, L.: A unified approach to short-time Fourier analysis and synthesis **65**(11), 1558–1564. https://doi.org/10.1109/PROC.1977.10770
2. Anagnostopoulos, G.C., Georgiopoulos, M.: Category regions as new geometrical concepts in Fuzzy-ART and Fuzzy-ARTMAP **15**(10), 1205–1221. https://doi.org/10.1016/S0893-6080(02)00063-1
3. Arbib, M.A. (ed.): The Handbook of Brain Theory and Neural Networks. MIT Press, 2nd ed edn.
4. Bargsten, V., Kirchner, F.: Actuator-level motion and contact episode learning and classification using adaptive resonance theory . https://doi.org/10.1007/s11370-023-00481-7
5. Carpenter, G., Grossberg, S., Rosen, D.: Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns. In: IJCNN-91-Seattle International Joint Conference on Neural Networks. vol. ii, pp. 411–416. IEEE. https://doi.org/10.1109/IJCNN.1991.155368
6. Carpenter, G.A., Grossberg, S., Reynolds, J.H.: ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network **4**(5), 565–588. https://doi.org/10.1016/0893-6080(91)90012-T
7. Cho, C.N., Kim, J.H., Kim, Y.L., Song, J.B., Kyung, J.H.: Collision Detection Algorithm to Distinguish Between Intended Contact and Unexpected Collision **26**(16), 1825–1840. https://doi.org/10.1080/01691864.2012.685259
8. Griffin, D., Jae Lim: Signal estimation from modified short-time Fourier transform **32**(2), 236–243. https://doi.org/10.1109/TASSP.1984.1164317
9. Iskandar, M., Albu-Schäffer, A., Dietrich, A.: Intrinsic sense of touch for intuitive physical human-robot interaction **9**(93), eadn4008. https://doi.org/10.1126/scirobotics.adn4008
10. Kouris, A., Dimeas, F., Aspragathos, N.: Contact distinction in human-robot cooperation with admittance control. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 001951–001956. IEEE. https://doi.org/10.1109/SMC.2016.7844525
11. Kouris, A., Dimeas, F., Aspragathos, N.: A Frequency Domain Approach for Contact Type Distinction in Human–Robot Collaboration **3**(2), 720–727. https://doi.org/10.1109/LRA.2017.2789249
12. Leonardo Enzo Brito da Silva, Elnabarawy, I., Wunsch II, D.C.: A Survey of Adaptive Resonance Theory Neural Network Models for Engineering Applications
13. Liu, J., Chen, B.: SonicSense: Object Perception from In-Hand Acoustic Vibration. https://doi.org/10.48550/arXiv.2406.17932
14. Meng, L., Tan, A.H., Wunsch, D.C.: Adaptive Scaling of Cluster Boundaries for Large-Scale Social Media Data Clustering **27**(12), 2656–2669. https://doi.org/10.1109/TNNLS.2015.2498625
15. Pankert, J., Hutter, M.: Learning Contact-Based State Estimation for Assembly Tasks. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5087–5094. IEEE. https://doi.org/10.1109/IROS55552.2023.10342219
16. Brito da Silva, L.E., Elnabarawy, I., Wunsch, D.C.: Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence **121**, 208–228. https://doi.org/10.1016/j.neunet.2019.08.033
17. Brito da Silva, L.E., Elnabarawy, I., Wunsch, D.C.: Dual vigilance fuzzy adaptive resonance theory **109**, 1–5. https://doi.org/10.1016/j.neunet.2018.09.015