POLITECNICO DI TORINO
Repository ISTITUZIONALE

Monocular Estimation of Connector Orientation: Combining Deformable Linear Object Priors and Smooth Angle Classification

Original
Monocular Estimation of Connector Orientation: Combining Deformable Linear Object Priors and Smooth Angle Classification / Caporali, Alessio; Galassi, Kevin; Berselli, Giovanni; Palli, Gianluca. - (2024), pp. 799-804. (IEEE International Conference on Advanced Intelligent Mechatronics (AIM) Boston, MA (USA) 15-19 July 2024) [10.1109/aim55361.2024.10637081].

Availability:
This version is available at: 11583/2992010 since: 2024-08-29T09:50:13Z

Publisher:
IEEE

Published
DOI:10.1109/aim55361.2024.10637081

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

07 May 2026

Monocular Estimation of Connector Orientation: Combining Deformable Linear Object Priors and Smooth Angle Classification

Alessio Caporali, Kevin Galassi, Giovanni Berselli, Gianluca Palli

***Abstract*—In this paper, a novel method for monocular estimation of connector orientation in wire and cable harnesses is introduced. The proposed approach combines Deformable Linear Objects (DLOs) priors with a learning-based technique for smooth angle classification, enabling precise prediction of connector orientation using only a single RGB image. The integration of DLO perception is crucial in recovering an initial coarse understanding of the connector pose. This is accomplished by linearly projecting the identified DLO endpoint using the predicted spline-based DLO representation model. To estimate the axial orientation of the connector, the proposed approach incorporates a smooth labeling technique in the angle classification process. This ensures effective handling of the circular nature inherent in angular data. Additionally, a self-supervised acquisition and annotation of the dataset samples is employed. To assess the effectiveness of the proposed method, we conducted experiments with a collection of real-world connectors sourced from the automotive sector. The outcomes underscore the potential applications of the proposed method in tasks related to the robotic manufacturing and assembly of complex deformable linear objects, such as wire harnesses.**

***Index Terms*—Deformable Linear Objects, Pose Estimation, Data Augmentation, Industrial Manufacturing**

I. Introduction

Wires, cables, and wire harnesses are pivotal in industries like automotive and aerospace [1]. An example of an automotive wire harness is depicted in Fig. 1. Currently, the manufacturing and assembly processes for these components rely heavily on manual execution, giving rise to significant challenges such as operator fatigue, ergonomic concerns, and issues related to product quality [2]–[4].

In the literature, cables and wires are commonly classified as Deformable Linear Objects (DLOs), emphasizing their characteristic of having a single main dimension. In contrast, wire harnesses are formed by interconnecting multiple DLOs, resulting in the creation of several branches. In the literature, they are frequently denoted as Deformable Multi-Linear Objects (DMLOs), highlighting their close association with the DLOs category.

Alessio Caporali, Kevin Galassi and Gianluca Palli are with DEI - Department of Electrical, Electronic and Information Engineering, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy.

Giovanni Berselli is with the Dept. of Mechanical, Energy, Management and Transportation Engineering, University of Genova, Via alla Opera Pia 15, 16145 Genova, Italy.

This work was supported by the Horizon Europe project *IntelliMan - AI-Powered Manipulation System for Advanced Robotic Service, Manufacturing and Prosthetics* [grant number 101070136].

Corresponding author: *alessio.caporali2@unibo.it*

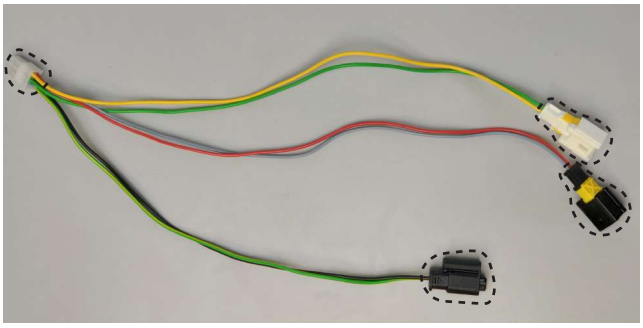

Fig. 1: Deformable Multi-Linear Object (DMLO) from the automotive sector with connectors attached at the extremities, highlighted with dashed black lines.

Connectors and other rigid elements, such as clamps and fixtures, are typically combined with DLOs and DMLOs. These rigid objects are essential to ensure proper functionality, signal transmission, and adherence to mounting and assembly constraints [1].

Integrating robotics and automation in industrial scenarios involving DLOs and DMLOs with connectors faces challenges, particularly in dealing with deformability, the high number of degrees of freedom, complex perception, and small sizes of DLOs [5]–[8]. Additionally, detecting connectors and estimating their poses is hindered by the scarce availability of 3D/CAD models and the susceptibility of depth data to noise and artifacts caused by the small sizes involved.

This paper tackles the relevant problem of connector detection and pose estimation, with a specific focus on the estimation of the axial orientation of a connector attached to a DLO or DMLO using a single RGB image, i.e. within monocular settings. Vision-based deep-learning solutions, like those in [6], [9], can reliably perceive DLOs by outputting sequences of keypoints or curves. Therefore, we can leverage the usual high stiffness of cables and wires to approximate the connector pose based on the obtained DLO curve-based representation. However, determining the axial orientation of the connector based solely on DLO knowledge remains ambiguous. Hence, this paper introduces a data-driven approach to estimate the axial orientation (and additionally the class type) of the connector, employing a self-supervised data-collection procedure.

In summary, the paper contributes to:

- Leveraging prior knowledge about DLOs to simplify the

problem of connector pose estimation;
- Self-supervised acquisition and annotation of dataset samples;
- Developing a data-driven approach for monocular angle axial classification of connector orientation, demonstrating generalization across various backgrounds and lighting conditions.

In the remainder of the paper, the related works are discussed in Sec. II while the proposed method is presented in Sec. III. The experiments are reported in Sec. IV. Finally, a discussion about the limitations of the approach is proposed in Sec. V, and the conclusions are drawn in Sec. VI.

II. Related Works

A. Object Pose Estimation in Robotics

Identifying an object in an image and determining its position and orientation relative to a coordinate system is a task known as object pose estimation. Learning-based solutions addressing this problem have gained significant attention, as evidenced by works such as [10]–[12]. Object pose estimation is crucial for robotic manipulation tasks, where simpler RGB devices are preferred due to their low noise, variability in resolutions, frame rates, and working modalities, as well as their affordability, compared to 3D devices [10], [12].

The 2022 BOP Challenge [13], a public benchmark dedicated to object pose estimation, has observed remarkable progress in tested methods, both in terms of accuracy and efficiency. Deep neural network-based methods have now surpassed traditional approaches relying on point pair features. The leading method in 2022 exclusively employs RGB image channels for both the training and testing phases. Notably, the top-performing method of 2022 outperforms RGB-D approaches from 2020 [13].

B. Vision-based Connectors Detection

Various studies have explored the perception of connectors [14]–[20]. In [15] and [14], the focus is on connector detection. [15] proposes a straightforward image-processing technique suitable only for structured environments. Conversely, [14] examines a range of learning-based object detection methods and introduces a dataset featuring diverse connectors.

In [19], connector detection is realized by exploiting torque readings traced along a DLO contour. Meanwhile, [20] implements a vision-based approach to estimate the pose of multiple connectors, emphasizing a connector-agnostic approach that, however, omits axial orientation estimation.

For small connectors, [17] employs a high-precision 3D camera and a data-driven strategy on point-cloud data, incorporating an ICP-based registration method.

Investigating the connector-mating process in wiring harness assembly, [16] adopts a visual servoing solution. Similarly, [18] delves into the connection between a wiring harness connector and its socket, utilizing a data-driven approach and integrating a shape-based 3D matching method for pose estimation, with the caveat of necessitating the 3D CAD model of the considered connector.

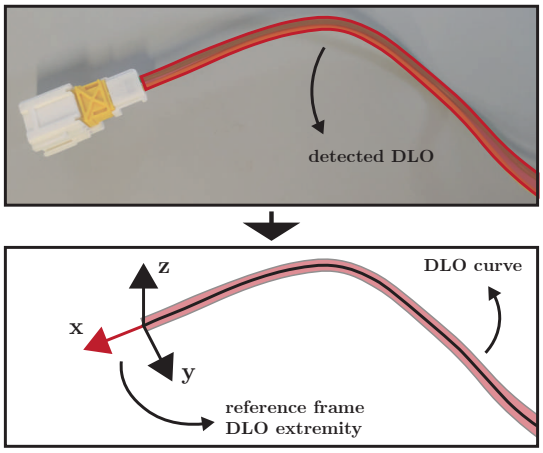

Fig. 2: DLO perception system providing both a mask-based and curve-based description of the object. In this illustration, a top-down view is assumed, such that the $x-y$ plane is parallel to the surface underneath.

C. Perception of Deformable Linear Objects

Among all possible perception systems, the vision-based approach is commonly preferred for DLOs due to the seamless integration of diverse cameras into robotic systems [8]. One key task in DLO perception is semantic and instance segmentation, as addressed in [7], [21]. The process of estimating the shapes of DLOs is discussed in [6], [22], [23], proposing recent real-time 2D approaches.

The 3D domain is less explored due to difficulties in sensing thin objects like DLOs [8]. Practitioners typically tackle the 2D shape estimation problem first and subsequently use depth data to transform the estimated shape into Cartesian space, as discussed in [23].

Conversely, [9] explores reliable 3D detection of DLOs through a multi-view stereo approach. This involves combining a 2D camera in an eye-in-hand configuration, a robotic arm, and shape estimation algorithms designed for 2D data to acquire the 3D shape of DLOs. However, a notable limitation of this method is its exclusive suitability for static scenes.

III. Method

The proposed approach exploits only a monocular RGB image to estimate the orientation of a small connector attached to a DLO.

The orientation of a rigid object can be described by 3 rotation components. With reference to Fig. 2, a reference frame is fixed to the DLO endpoint. Therefore, the orientation components can be organized into two main groups: those not oriented along the DLO neutral axis, i.e. rotation around the y and z axis of Fig. 2; the one oriented along the DLO neutral axis, i.e. rotation around the x axis. The former orientation components are estimated based on the DLO perception, as detailed in Sec. III-A. For the latter orientation component, a specifically developed data-driven method is proposed, as discussed in Secs. III-B, III-C and III-D.

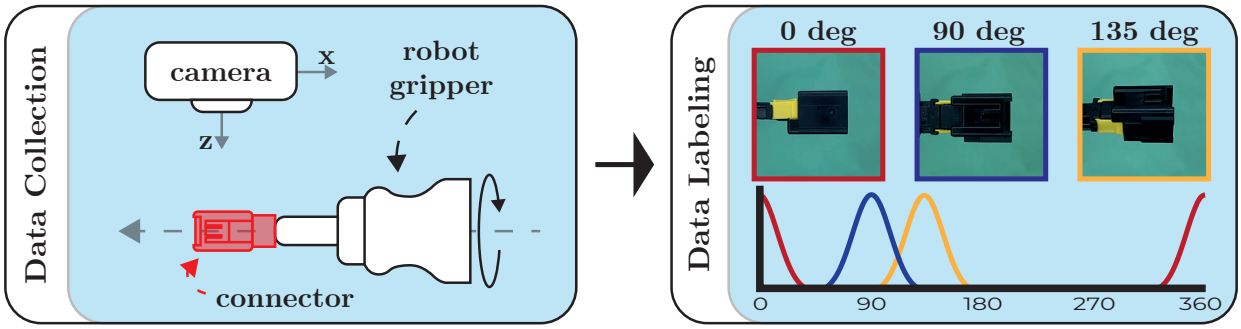

Fig. 3: Schematic view of the self-supervised data collection and labeling approach.

A. Preliminaries: DLO Perception for Orientation Estimation

In this section, our emphasis is on the 2D domain, specifically to estimate the z-axis rotation component. It is important to note that the rotation component around the y-axis can only be addressed in the 3D domain. Nevertheless, it is worth mentioning that the methodology developed for the 2D domain is readily extendable to the 3D context.

The process of deriving the z-axis rotation component starts with the analysis of an RGB image capturing the scene, as illustrated in Fig. 2. From the image, DLO instances are obtained using the real-time DLO perception algorithm *RT-DLO* [6]. This algorithm integrates a deep convolutional neural network for background segmentation [7] and utilizes a graph-based representation of the foreground scene to extract individual instances of DLOs. Each DLO instance is then defined as a sequence of 2D keypoints within the image plane. If required, one can obtain a 3D representation of a detected DLO shape by applying established methods, like [9], resulting in 3D keypoints in Cartesian space.

In both 2D or 3D scenarios, a cubic B-spline is fitted to the keypoints, providing a continuous representation of the considered DLO.

Therefore, an estimate of the connector orientation along the considered axis can be easily obtained by leveraging the knowledge of the DLO shape. Specifically, assuming a predominantly linear behavior due to the stiffness of the DLO, a linear curve can be aligned to match the derivative of the spline at the designated extremity. In a simpler 2D case, the result is a direction coincident with the one depicted by the x-axis in Fig. 2. However, considering a more general 3D scenario, the presence of a rotation around the y-axis will result in the definition of a new axis direction in the $x-z$ plane. Thus, the connector pose is defined at the extrapolated terminal region of the DLO instance. The main assumption exploited is the knowledge that a connector is attached to that end of the DLO. This assumption is reasonable in industrial applications where the structure of the DLOs and DMLOs being manipulated is usually known.

The only orientation component that cannot be estimated with the outlined procedure is the axial one, as a rotation around the x-axis is not detectable in the DLO shape from the image data. Therefore, a data-driven approach, detailed in the following, is developed to address this specific goal.

B. Data Collection and Labeling

Collecting and labeling data for learning-based solutions is a tedious and expensive process [10]. A common strategy is to rely on synthetic data [11]. In general, object pose estimation has witnessed a convergence in task accuracy between synthetic and real domain approaches [13]. However, exploiting synthetic data requires 3D models of the objects of interest, which may not always be available, as in the setup considered.

Therefore, we deploy a robotic system for the automatic collection and labeling of data. This decision removes the burden of both the collection and annotation processes from human users. The robotic setup, schematized in Fig. 3, employs a manipulator with a parallel fingers gripper and a statically mounted 2D RGB camera with a top-down view. The camera is intrinsically and extrinsically calibrated with respect to the robot.

The object of interest, in our case a connector, is firmly grasped with the gripper at a known initial axial orientation and moved inside the camera field of view. The robot's configuration is selected such that the end-effector approach axis and the camera planes are parallel. Therefore, an axial angle range within ± 180 degrees is considered, thus covering the entire spectrum of possible orientations. This range is discretized with a user-defined angular step, and an image of the connector is collected for each discrete available value. Within each image, the angular displacement with respect to the initial configuration is recorded. This value is used to automatically assign an angular label value to each image. In order to focus the learning process on the specific connector of interest, square crops of the objects are obtained from the saved images. The areas for these crops are determined based on the projected location of the gripper fingertips in the image plane, a value easily obtained through the robot's forward kinematics and the extrinsic matrix of the calibrated camera. The result of the data collection and labeling approach is shown on the right side of Fig. 3.

C. Learning-based Axial Angle Estimation

The input RGB image crop is processed by a CNN-based model, which predicts the angular orientation of the object along the axial direction. To enhance information output during network deployment, a classification branch is also

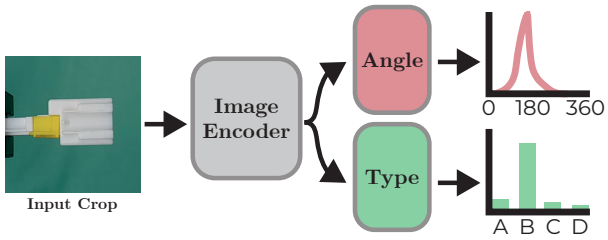

Fig. 4: Network structure composed of an image encoder and two classification heads (angle and object type).

incorporated into the network structure to predict the object type, for example, connector A or B.

1) Smooth Angle Encoding: Predicting angular values using a learning-based method can pose challenges due to the periodic nature of angular data, leading to inaccuracies in distance representations when calculating the loss function. For instance, an angle of $0°$ is intuitively "close" to both $5°$ and $355°$, even though applying common losses like L1-loss or MSE-loss within the $[0°, 360°]$ domain results in significantly different loss values. To tackle the periodicity and ambiguity in loss computation, an approach introduced in [24] is adopted.

In this approach, a given angular value θ within the range $[0°, 360°]$ is encoded as a k-dimensional vector, where k is a function of the angular discretization step selected. For instance, with a step of 5 degrees, $k = 72$. This vector is created by applying a Gaussian function centered at θ with variance σ. This encoding method ensures a smooth propagation of the angle θ in its proximity, providing advantages during the loss computation process. During training, binary cross entropy serves as the loss function, framing the learning task as a classification problem for the angular value among the k available classes. As a result, extracting the actual predicted angle is straightforward, as it corresponds to the index of the vector associated with the maximum probability. The encoding of the angular label via a Gaussian function for classification purposes is graphically shown via color-codes in Fig. 3 (bottom right side).

2) Object Type Classification: Beyond predicting the angle, an additional output for the object type has been introduced. The label for the object type is readily acquired during the data collection process and is represented as an ID number (e.g., 0 for connector A, 1 for connector B, etc.). This label is encoded as a one-hot vector, and the optimization of the network weights is carried out using binary cross-entropy as the loss function.

D. Network Architecture

The input image undergoes encoding by a pre-trained backbone network, such as ResNet [25], resulting in a linear layer with dimensions that vary depending on the chosen backbone size. To ensure consistent dimensions, an extra linear layer is introduced to transform the feature dimensions of the backbone to a fixed size. Subsequently, the output from the backbone is fed into two classification heads — namely, the angle head and the type head. Each classification head is formed by a multi-layer perception (MLP) composed of two linear layers. The output dimension of the angle head is equal to $360/k$. Instead, the output dimension of the type head is equal to the number of classes present. For both outputs, an argmax operation is performed to retrieve the predicted values. The overall network structure is illustrated in Fig. 4.

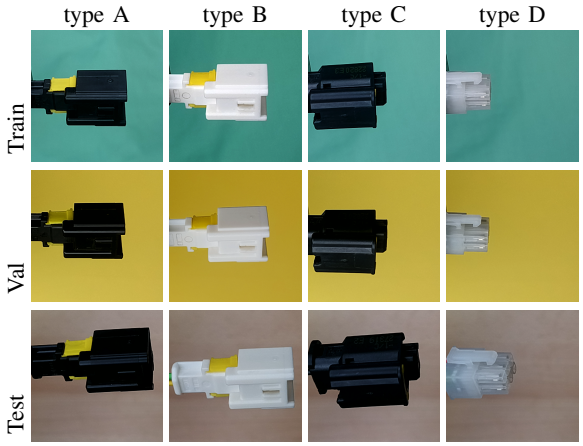

Fig. 5: $45°$ configuration of the different connectors across the dataset splits.

IV. Experiments

The approach outlined in Sec. III is assessed using a robotic configuration consisting of a UR5 robot equipped with a Robotiq Hand-E parallel fingers gripper. In the robotic work cell, an OAK-1 RGB camera from Luxonis is securely mounted with a top-down perspective, mirroring the configuration schematized in Fig. 3. The camera's resolution is constrained to 1920×1080 pixels. An Ubuntu 20 PC with an Intel i9-9900k CPU, 64GB of RAM, and an Nvidia RTX 2080 Ti is used as the computing platform.

A set of four distinct connectors is utilized to validate the approach. The connectors come from the wire harness of Fig. 1 and are depicted in more detail in Fig. 5. These connectors, commonly found in the automotive field, are characterized by small dimensions on the order of a few centimeters, making them challenging to detect with common depth sensors like Intel RealSense devices.

Data and source code are openly released at the following repository: https://github.com/lar-unibo/connector_orientation.

A. Optimization Details

To generate the training dataset, each connector is made free of cables/wires to enable secure and accurate characterization of its axial orientation, eliminating possible ambiguity caused by the deformability of DLOs during the learning process. Then, the data collection and labeling methodology explained in Sec. III-B is employed to generate the training dataset. An angular step of 5 degrees is utilized, so a total of $k = 72$ images is collected. For each image, a crop of 512×512 pixels around the connector is performed. Since the generation of the dataset is quite efficient, requiring only a couple of minutes, for the same connector a set of 5 different backgrounds is

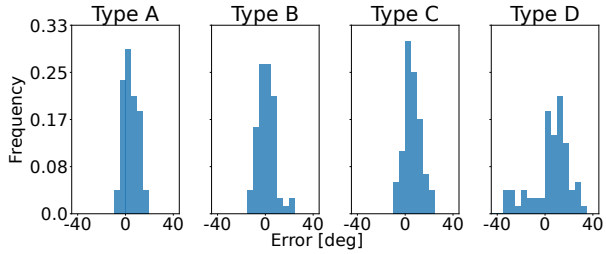

Fig. 6: Histogram of angular error (in degrees) computed between predicted and ground truth values across the different connectors.

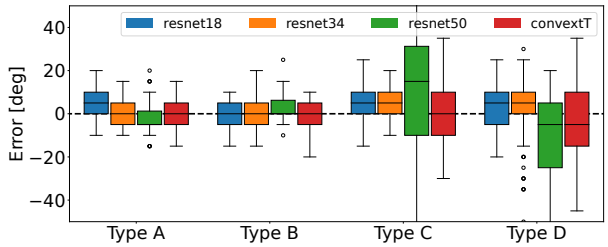

Fig. 7: Error comparison among different backbones architectures, specifically ResNet [25] (18, 34, and 50 layers) and ConvNext [26] (Tiny) concerning the different test connectors.

employed to increase the dataset variance. The process is repeated for each connector resulting in a total dataset of 1440 samples. During the network optimization, 4 backgrounds are employed for the training split, while one background is for the validation split. The model described in Sec. III-D is optimized using the mentioned training and validation datasets. ResNet34 [25] is selected as the backbone network. The training process involves 50 epochs with a batch size of 4 and a learning rate of 1×10^{-4}. To enhance the dataset, augmentation techniques such as random brightness and contrast, random rotation (limited to 15 deg), and random crop are employed. The final weights are selected as the ones having the minimum validation loss.

B. Test Dataset and Error Metric

Following a similar procedure to the one mentioned in Sec. IV-A, a test dataset is collected to quantitatively evaluate the approach. In this case, the connectors are kept with the wires attached to replicate a test scenario more closely resembling actual settings. The presence of the wires makes the connector move slightly, resulting in unwanted rotated shapes as opposed to the perfectly horizontal orientation of the training/validation dataset (see Fig. 5). The labeling of the test dataset follows the same approach detailed in Sec. III-B.

The error between the predicted and the ground truth value is computed as the evaluation metric.

C. Results

The results for the test set, depicted in Fig. 5, are further illustrated in Fig. 6. This plot showcases the behavior of angular errors across various test connectors. The mean and standard deviation errors for each connector type are as follows: type A 1.8 ± 6.2 deg; type B -1.4 ± 7.2 deg; type C 3.7 ± 7.1 deg; type D 1.8 ± 16 deg. These results highlight the capability of the proposed method in estimating angular axial values close to the actual ones. Notably, the type D connector poses the most significant challenge. We argue that this difficulty arises from its smaller size and the lack of relevant features, making it more challenging for the model to learn the key differences between different viewpoints.

Regarding the classification of connector types within the 4 classes, all test connectors were successfully classified. This outcome was expected, given the tasks' simplicity in comparison to angular axial estimation.

To assess the impact of backbone selection on angular axial estimation, various backbone architectures were tested, specifically those from the ResNet [25] family, including the 18, 34, and 50-layer versions, and from the ConvNext group, employing the ConvNext-Tiny model [26]. The choice of backbones aligns with the objective of evaluating different sizes in terms of the number of parameters. The selected set of backbones encompasses models with parameter counts ranging from 11.4M, 21.5M, and 24.9M for the ResNet versions, and 28.5M for ConvNext-Tiny. The results are depicted in Fig. 7. The plot illustrates that larger backbones result in more pronounced outlines in error values, emphasizing the preference for smaller models. Hence, we opted for ResNet34 in the analysis presented in Fig. 6. The small network size enables fast inference speeds even on non-high-end hardware. For example, on a laptop with an Nvidia GTX1050Ti, inference takes 12 ms for 512×512 image crops. On the outlined workstation setup, the same image is processed in 7 ms. Both scenarios achieve over 60 FPS.

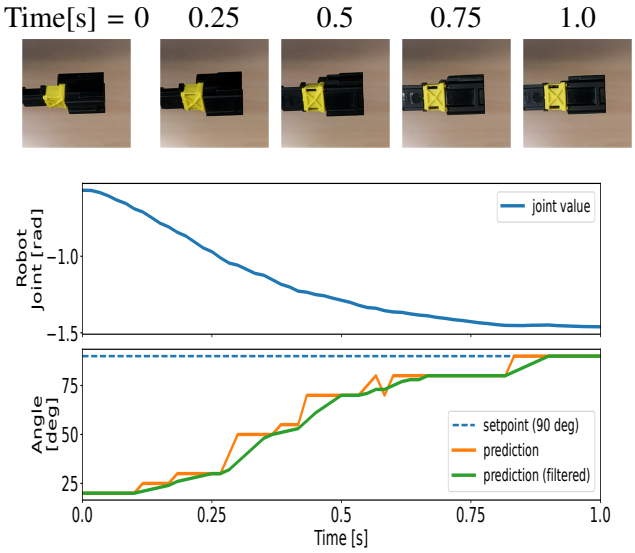

Fig. 8: Visual serving experiment on connector type A. The robot aligns the connector with the target angle of 90 deg.

D. Real Use Case: Visual Servoing Control

Visual servoing is tested as a real-world application of the proposed method. Given the timings of Sec. IV-C, the servoing

task is fixed at 60 Hz. Fig. 8 demonstrates the practical implementation, showcasing the successful integration of the approach into a control loop to guide the robotic system. The video of the experiment is available as supplementary material.

V. Discussion and Limitations

DLO Prior versus Object Detection: The DLO-based prior could potentially be replaced entirely by developing an object detection system to localize connectors, as explored in various studies [14]. However, this alternative approach requires the use of oriented bounding boxes to accurately capture the connector orientation along the two axes detailed in Sec. III-A. Nevertheless, a notable advantage of the DLO-based approach lies in its ability to automatically identify one side of the connector. Even with oriented bounding boxes, the top-down ambiguity persists. Therefore, a potentially more efficient strategy might involve combining these two methods to leverage their respective strengths, enhancing the overall system's robustness.

Generalization to Other Objects: The presented approach extends beyond the scope of connectors and DLOs. The primary assumptions involve knowledge of the "bottom" side of the object of interest (e.g., where the connector and the DLO are connected) and the asymmetry of the object along its axial axis. Consequently, the same methodology is applicable to other objects and will be investigated in future studies.

Angular Discretization Step: In this paper, we define the angular step as 5 degrees, yielding $k = 72$ angular classes. This choice balances accuracy with noticeable appearance variance in images. Experimentation with a 1-degree step proved challenging for the human eye to discern visual differences between nearby images and optimizing the network posed difficulties. Addressing this issue is a task for future research.

VI. Conclusions

The proposed approach for angular axial estimation tackles the challenging task of estimating the axial orientation of a connector attached to a Deformable Linear Object (DLO) using only a single RGB image. Priors on DLOs have proven valuable in simplifying the complexity of the problem, focusing on the estimation of the axial component of the orientation through a learning-based approach. A self-supervised collection and annotation of the dataset for the data-driven pipeline circumvent tedious and expensive human-based work. Additionally, the smooth labeling of the angular data addresses the circularity ambiguity associated with the angle domain. Experimental validation, conducted with real-world connectors from the automotive sector, demonstrates the effectiveness of the proposed method.

References

[1] H. G. Nguyen, R. Habiboglu, and J. Franke, "Enabling deep learning using synthetic data: A case study for the automotive wiring harness manufacturing," *Procedia CIRP*, 2022.

[2] O. Salunkhe, W. Quadrini, H. Wang, J. Stahre, D. Romero, L. Fumagalli, and D. Lämkull, "Review of current status and future directions for collaborative and semi-automated automotive wire harnesses assembly," *Procedia CIRP*, 2023.

[3] M. Pantano, Q. Yang, A. Blumberg, R. Reisch, T. Hauser, B. Lutz, D. Regulin, T. Kamps, K. Traganos, and D. Lee, "Influence of task decision autonomy on physical ergonomics and robot performances in an industrial human–robot collaboration scenario," *Frontiers in Robotics and AI*, 2022.

[4] M. Pantano, A. Curioni, D. Regulin, T. Kamps, and D. Lee, "Effects of robotic expertise and task knowledge on physical ergonomics and joint efficiency in a human-robot collaboration task," in *IEEE-RAS Int. Conf. HUMANOIDS*. IEEE, 2023.

[5] A. Caporali, P. Kicki, K. Galassi, R. Zanella, K. Walas, and G. Palli, "Deformable linear objects manipulation with online model parameters estimation," *IEEE Robotics and Automation Letters*, 2024.

[6] A. Caporali, K. Galassi, B. L. Žagar, R. Zanella, G. Palli, and A. C. Knoll, "RT-DLO: Real-time deformable linear objects instance segmentation," *IEEE Transactions on Industrial Informatics*, 2023.

[7] A. Caporali, M. Pantano, L. Janisch, D. Regulin, G. Palli, and D. Lee, "A weakly supervised semi-automatic image labeling approach for deformable linear objects," *IEEE Robotics and Automation Letters*, 2023.

[8] K. P. Cop, A. Peters, B. L. Žagar, D. Hettegger, and A. C. Knoll, "New metrics for industrial depth sensors evaluation for precise robotic applications," in *IEEE/RSJ Int. Conf. IROS*. IEEE, 2021.

[9] A. Caporali, K. Galassi, and G. Palli, "Deformable linear objects 3D shape estimation and tracking from multiple 2D views," *IEEE Robotics and Automation Letters*, 2023.

[10] X. Deng, Y. Xiang, A. Mousavian, C. Eppner, T. Bretl, and D. Fox, "Self-supervised 6d object pose estimation for robot manipulation," in *IEEE Int. Conf. ICRA*. IEEE, 2020.

[11] G. Wang, F. Manhardt, J. Shao, X. Ji, N. Navab, and F. Tombari, "Self6d: Self-supervised monocular 6d object pose estimation," in *Proc. of ECCV*. Springer, 2020.

[12] G. Billings and M. Johnson-Roberson, "Silhonet: An rgb method for 6d object pose estimation," *IEEE Robotics and Automation Letters*, 2019.

[13] M. Sundermeyer, T. Hodaň, Y. Labbe, G. Wang, E. Brachmann, B. Drost, C. Rother, and J. Matas, "Bop challenge 2022 on detection, segmentation and pose estimation of specific rigid objects," in *Proc. of the IEEE/CVF CVPR*, 2023.

[14] H. Wang and B. Johansson, "Deep learning-based connector detection for robotized assembly of automotive wire harnesses," in *IEEE Int. Conf. CASE*. IEEE, 2023.

[15] F. Yumbla, M. Abeyabas, T. Luong, J.-S. Yi, and H. Moon, "Preliminary connector recognition system based on image processing for wire harness assembly tasks," in *IEEE Int. Conf. ICCAS*. IEEE, 2020.

[16] H.-C. Song, Y.-L. Kim, D.-H. Lee, and J.-B. Song, "Electric connector assembly based on vision and impedance control using cable connector-feeding system," *Journal of Mechanical Science and Technology*, 2017.

[17] C. Ying, Y. Mo, Y. Matsuura, and K. Yamazaki, "Pose estimation of a small connector attached to the tip of a cable sticking out of a circuit board," *Int. Journal of Automation Technology*, 2022.

[18] H. Zhou, S. Li, Q. Lu, and J. Qian, "A practical solution to deformable linear object manipulation: A case study on cable harness connection," in *IEEE Int. Conf. ICARM*. IEEE, 2020.

[19] A. Monguzzi, A. M. Zanchettin, and P. Rocco, "Sensorless robotized cable contour following and connector detection," *Mechatronics*, 2024.

[20] A. Monguzzi, C. Cella, A. M. Zanchettin, and P. Rocco, "Vision-based state and pose estimation for robotic bin picking of cables," in *IEEE/RSJ Int. Conf. IROS*. IEEE, 2023.

[21] J. Dirr, D. Gebauer, J. Yao, and R. Daub, "Automatic image generation pipeline for instance segmentation of deformable linear objects," *Sensors*, 2023.

[22] A. Choi, D. Tong, B. Park, D. Terzopoulos, J. Joo, and M. K. Jawed, "mbest: Realtime deformable linear object detection through minimal bending energy skeleton pixel traversals," *IEEE Robotics and Automation Letters*, 2023.

[23] P. Kicki, A. Szymko, and K. Walas, "Dloftbs – fast tracking of deformable linear objects with b-splines," in *IEEE Int. Conf. ICRA*, 2023.

[24] X. Yang and J. Yan, "Arbitrary-oriented object detection with circular smooth label," in *ECCV*. Springer, 2020.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE/CVF CVPR*, 2016.

[26] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proc. of the IEEE/CVF CVPR*, 2022.