Design of a Deep Post Gripping Perception Framework for Industrial Robots

Firas Zoghlami
Applied Sciences and Mechatronics
Munich University of Applied Sciences
Munich, Germany
Email: firas.zoghlami@hm.edu

Philip Kurrek
Applied Sciences and Mechatronics
Munich University of Applied Sciences
Munich, Germany
Email: philip.kurrek@hm.edu

Mark Jocas
Applied Sciences and Mechatronics
Munich University of Applied Sciences
Munich, Germany
Email: mark.jocas@hm.edu

Giovanni Masala
Computing and Mathematics
Manchester Metropolitan University
Manchester, United Kingdom
Email: g.masala@mmu.ac.uk

Vahid Salehi
Applied Sciences and Mechatronics
Munich University of Applied Sciences
Munich, Germany
Email: salehi-d@hm.edu

The use of flexible and autonomous robotic systems is a possible solution for automation in dynamic and unstructured industrial environments. Pick and place robotic applications are becoming common for the automation of manipulation tasks in an industrial context. This context requires the robot to be aware of its surroundings throughout the whole manipulation task, even after accomplishing the gripping action. This work introduces the deep post gripping perception framework, which includes post gripping perception abilities realized with the help of deep learning techniques, mainly unsupervised learning methods. These abilities help robots to execute a stable and precise placing of the gripped items while respecting the process quality requirements. The framework development is described based on the results of a literature review on post gripping perception functions and frameworks. This results in a modular design using three building components to realize planning, monitoring and verifying modules. Experimental evaluation of the framework shows its advantages in terms of process quality and stability in pick and place applications.

1 Introduction

The integration of robotic solutions in industrial environments has been one of the most important pillars of automation. According to the International Federation of Robotics [1], the robotic adoption in the manufacturing industry increased by an annual global average of 12% between 2011 and 2016. The structuring of industrial environments needed for the integration of these robots comes with the cost of the flexibility in these environments [2]. The trend in the production environments is a continuously decreasing number of predictable assembly lines [3]. This means that developing new automation solutions has to follow agile methods that reduce the time to produce [4–8]. Also, the flexibility of industrial robots has to be increased, so that they can be deployed in different stations. This can be realized by the use of computers in tasks that require knowledge, perception, reasoning, learning, understanding and similar cognitive abilities, also called artificial intelligence [9]. Machine learning, which is the use of learning agents that improve their performance on future tasks after making observations about the world [10], and mainly deep learning, meaning to learn using artificial neural networks with hidden layers, has known a wide application spectrum [11] and an important success recently [12]. The use of this artificial intelligence ability increases the autonomy of robots and their ability to adapt to new conditions, which allow their usage in novel applications [13–16].

Need for Post Gripping Perception In environments that can be dynamic and unstructured, industrial robots need adaptive capabilities like machine learning and robotic perception. Industrial environments are dynamic when unplanned events

can occur and change the state of running processes. On the other hand, they are unstructured when objects are not fixed in known positions. The challenges in such environments lie in the gripping and placing steps of robotic manipulation [17].

The focus in the research field has been on the perception for the gripping step [18–20] to the detriment of the placing action [21, 22]. Analogically, the focus of the industrial research and development lies in the perception based gripping of items in structured or unstructured environments [23, 24]. This is reflected in the points distribution for the score calculation in the Amazon Picking Challenge, where failing to place an object causes a penalty of only 15% of the picking points [25].

However, object placing is an important topic, that has more optimization potential than the pre-gripping phase [26], where research about motion planning, machine learning and perception can meet [27]. Furthermore, the robustness and quality of manipulation tasks can be increased by reducing the placing error [28]. In this work, it is believed that robotic automation of dynamic and unstructured environments leads to the necessity to go beyond current *"perceive to grip"* research.

Contribution   The focus of this work is based on the results of the literature review on designing a holistic framework for post gripping perception in dynamic and unstructured robotic environments. It introduces a holistic framework that includes all post gripping perception functions found in the literature, namely planning, monitoring and verification. It is based on the use of deep convolution neural networks for object detection and deep convolution auto-encoders for anomaly detection because of their ability to generalize to unknown objects or events, which can ensure high scalability and reduce the setup time of new pick and place applications, in comparison to deterministic frameworks [29]. The design of the introduced framework prioritizes the use of unsupervised learning for training deep neural networks, by collecting training data from running processes that are already supervised. This presents an advantage compared to collecting failure data as in [30], since they are less probable to occur in a running system. Consequently, the introduced deep post gripping perception framework allows the optimization of perception development costs and enables the termination of pick and place manipulation tasks with a confirmed stable placing while respecting the quality requirements in industrial applications.

Research Methodology   This work is an applied descriptive research [31] based on secondary literature review data. It represents the first descriptive study and the prescriptive study of the overall research project about *deep post gripping perception for industrial logistics robots*, that follows the *Design Research Methodology* [32]. The research work of
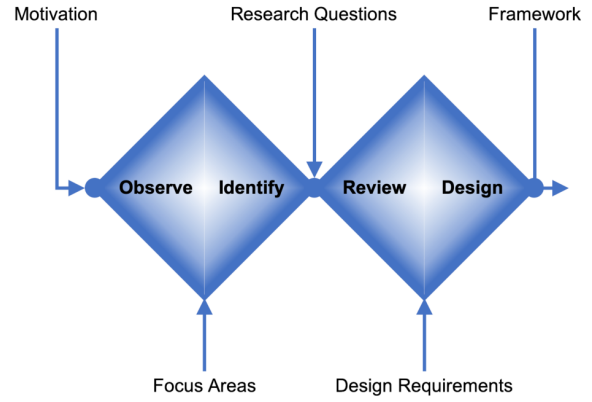

Fig. 1.   The adopted double diamond research methodology.

this paper follows, as described in Fig. 1, a double diamond approach [33] with four phases:

- a first diverging exploration phase, where different areas are considered for the use of artificial intelligence in new robotic applications,
- a first converging phase, where the research gap and research questions are identified in the intersection of different areas,
- a second diverging phase, where a literature review is carried out in order to define the current state of the art and the requirements for the work's contribution and
- a second converging phase, where the design of a novel framework is developed as a result of the research effort.

This work starts with a related work section for the literature review. The defined contribution from this review is then realized in the framework design section. Furthermore, an implementation of the designed framework in a real robotic use case is presented and evaluated. Finally, the contribution of the presented framework is discussed and the intended future work is described.

2   Related Work

To the best of the authors' knowledge, it has been only in 2006 that the first work, that focuses on robotic object placing in a pick and place context, has been published in [34]. In this work, the authors solved an object placing problem on a flat shelf based on passive compliance and force control. While their method is based on force measurements, other methods are mostly based on image data for post gripping perception.

2.1   Post Gripping Perception Functions

The research work about robotic perception for object placing can be categorized in two major contexts: industrial robotic use cases and service robots for human support. The motivations for considering post gripping perception in these two domains are different. In fact, while industrial assembly

robotic systems target high precision in insertion tasks such as mounting two parts together, service robots need to fulfill stability and quality requirements, such as placing an object in the appropriate compartment in a refrigerator and leaving it undamaged in a stable pose.

2.1.1 Industrial Robots

One major context for the post gripping perception research in the industry is the peg-in-hole task. This task can occur in almost one third of the assembly tasks that can be automated [35]. The high precision and reliability requirements together with the dynamic brought by the environment, process or manipulated objects of this task motivate the use of perception systems in different stages of the post gripping process, meaning from the moment the robot grips the item until it inserts it completely in the target slot.

Several research works deal with equipping the robot with some sensors for post gripping perception. Bo *et al.* [36] worked on the problem of automatic plugging of electronic components in printed circuit boards by identifying and localizing the insertion holes. The authors used a template matching classifier based on edge features of insertion holes. This planning information has then to be sent to the actuator controller in the automated production system. Wang *et al.* [37] integrated guidance and alignment modules in a robotic solution for the assembly of large-scale parts. The guidance module uses one vision sensor to calculate the 3D position and orientation of the installation location, in order to plan the target pose of the manipulator. In a second step, an alignment module, composed of one vision sensor and three 1D laser sensors, computes with the help of template matching the positioning error of the object relative to the target installation and define a fine-positioning plan for precise insertion. Another approach to plan the placing step in a pick and place task, is by simulation as introduced by Karako *et al.* [38]. The authors focused on predicting potential slipping of the object from the suction cups while transporting it. They targeted higher grasping quality for a safe placing with shorter cycle time and higher process efficiency.

Force sensors equipped with camera devices are also used for the precise manipulation in normal and micro-scale robots [39]. Nozu *et al.* [40] introduced a tactile sensor equipped with a camera in order to perform the in-hand object localization in a bolt insertion task. The proposed sensor is a combination of a reflective membrane and marker displacement based on the camera input. The computed in-hand position and orientation of the object are used to control the robot manipulator motion for the insertion task. Hayakawa *et al.* [41] worked on shaft insertion in a moving object with the help of multiple light sensors called position sensitive detectors. The realized system is a complex configuration of sensors on the end-effector that monitors the target hole

during the manipulation. De Gregorio *et al.* [42] consider the peg-in-hole problem in the context of wire insertion in commercial electro-mechanical components with the help of cameras and tactile sensors. This is a challenging task, since the manipulated object is not rigid, which means that its in-hand pose changes along the manipulation task.

Other than planning and monitoring the placing process, some research work has focused on verifying the outcome of the post gripping process, as in the work of Erkent *et al.* [43]. Here, the authors used a two-step deep neural network to predict the status and outcome of a robotic manipulation task based on color image data containing red, green and blue channels (RGB) from an external camera facing the manipulation task scene. Their "deep task status estimator" is composed of a Faster R-CNN object detection network [44] followed by a modified classification AlexNet network [45] applied on resulting bounding boxes of the first step. The result of the estimator is the task outcome (success, fail, in progress) together with the pose and in-plane orientation of the manipulation item and its target container.

The post gripping perception for object placing is also considered in other application fields. Zhang *et al.* [46] propose, for instance, an automated robotic vitrification system capable of precisely and reliably picking and placing multiple embryos in a medical environment. Besides the medical application, the post gripping perception is considered in microscopy assembly. For instance, Komati *et al.* [47] used computer-aided design (CAD) based visual tracking as part of a visual servoing strategy for inserting micro parts into the grooves of micro-assembly substrate.

2.1.2 Service Robots

The second major area where the problem of post gripping perception is addressed in the literature is domestic service robots. Manipulation tasks in such a context are usually a pick and place action of domestic items. The challenges that arise in this case are that manipulated objects are usually unknown, since humans can purchase new items that are be added to the robot's environment on daily basis. Also, the structure of the robot's surrounding continuously changes by the human influence, since no standard processes are defined in a domestic environment. These two aspects affect the object placing task since the robot has to place unknown items in an unknown scene.

In order to tackle this problem, research efforts have been canalized in solving the problem of placing objects on surfaces, usually a table, that can be free or cluttered. Harada *et al.* [48] introduced an approach to plan the location and orientation of gripped objects on the target surface. Based on point-cloud data, two geometrical surfaces from the manipulated object and the target surface are defined, while fulfilling stability and inclusion requirements. These two planes rep-

resent the contact surface between the placed object and the table. Another placement planning method for a given task was introduced by Toris *et al.* [49]. The authors defined the optimal placement location of a gripped object on a planar surface based on a simulation fed by a database of associations between the defined task and possible end states.

Jiang *et al.* [22] claim to be the first to work on placing objects in a complex scene. A complex scene in this context means a cluttered planar or non planar surface with no free space to place an object. The authors considered two requirements for a successful placing, namely stability and semantic preference. Their proposed approach is based on a support vector machine [50] classifier of placing data with features such as support stability, preferred orientation, supporting contacts and caging. The scenario with the highest score is considered to plan placing actions. Planning object placement on a cluttered surface was also considered in the work of Schuster *et al.* [21], where the authors introduced a supervised learning approach for the segmentation of placing scenes in clutter and surface segments. The surface segments are then planned as a potential placing target while considering some semantic conditions such as the inconvenience of the table edges for item placing. Besides planning the placing of an object in a cluttered scene, some approaches focus on planning a placing with a preparing action. For instance, Cosgun *et al.* [27] proposed to use a pushing action on items situated on cluttered tables in order to clear enough space for the footprint of the object of interest. In his research work, Baumgartl [51] proposed a novel approach for placement planning based on reconstructed surface models of the manipulated objects and environment. A stable placement on a complex and non-planar surface was achieved by planning a fine-positioning motion for the placing action.

2.2 Post Gripping Perception Frameworks

The literature review in this work has shown that there exist efforts to study the post gripping perception in a robotic context. They focus on three main topics, namely planning the placing action based on the target detection, monitoring the transport of the item after gripping it and verifying the manipulation outcome. However, there is a need for holistic frameworks to include all post gripping perception functions. The American National Institute of Standards and Technology (NIST) is working on the modeling of robotic kitting applications with the help of the Web Ontology Language (OWL) for the assessment and increase of robotic agility [29]. This work evaluates for each action in a kitting application its precondition and its effect for further planning of next actions [52]. This includes the actions in the pre-gripping and post gripping phases. With the help of the NIST framework, post gripping functions can be integrated in a kitting application. However the failure detection in the precondition or effect of an action

is based on logical operations with region connection calculus (RCC8) and a listing of known failures [53]. This can lead to a cost intensive setup of new applications, since all possible process specific known failures have to be modeled. Learning techniques can be used to generalize these failures, which can also detect unknown failures that can occur in running processes.

Using machine learning for detecting process failures in such pick and place applications has been studied in the work of Altan *et al.* [30]. In their work, the authors have used Long Short-Term Memory (LSTM) [54] based deep neural networks for the identification of robotic process anomalies, based on supervised training and focusing on determining their causes. This is used for planning new pick and place actions in further processes. However, the authors have not studied the direct use of failure detection to improve the stability of the running process. Using machine learning in the form of deep neural networks to model running processes in an unsupervised manner and using these models to detect process failures as anomalies will be studied in the introduced deep post gripping perception framework. Furthermore, different sensors have been used for the post gripping perception functions, as described in the literature review. A common solution using one sensor hardware that can be integrated in all robots is needed. The introduced deep post gripping framework uses only image data collected by a camera that can be integrated in common robot manipulators and their applications, namely in their bases or end-effectors.

3 Deep Post Gripping Perception Framework

In order to define the needed functions for the post gripping perception in an unstructured and dynamic environment, an analogy to the more intuitive pre-gripping perception is conducted based on the process-environment map, as depicted in Fig. 2. The main question that has to be answered through this analysis is: which information about the robot, the task item and/or the task target should the robot perceive in the manipulation steps?

In an unstructured environment, the robot has to detect and localize the task items and their targets. The target class has also to be defined based on the item class or its quality. In fact, an item with quality issues has to be placed in dedicated targets. It also needs to define the occupancy map of the obstacles on the way to the target that can influence its manipulation plan. For this purpose, a camera has to be installed externally on the top of the robot cell for example. This limits the flexibility needed for intelligent robots that can be deployed in different applications. Hence, the focus in the deep PGP framework has been shifted from the obstacle poses to the consequences of their poses and dynamics on the manipulation process. This can be realized by focusing on the dynamic of the task environment by detecting deviations
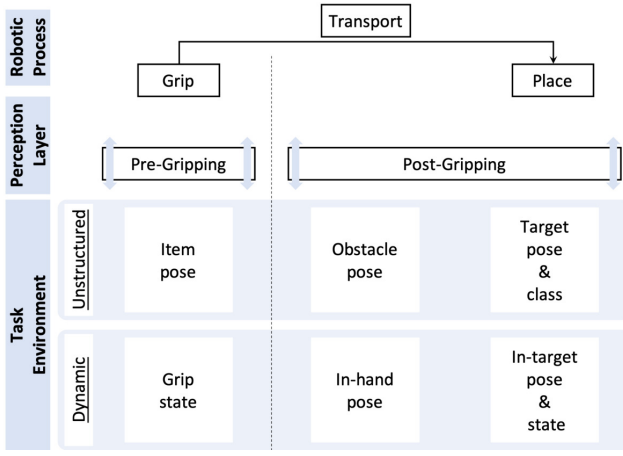
Fig. 2. Analysis of the pre-gripping and the post gripping phases in a robotic manipulation process for the definition of the perception information needed in unstructured and dynamic environments. All poses have to be computed in the robot base coordinate system, so that they can be reached or avoided by the robot. The focus in the design of the deep PGP framework in this work lies in the transport step in dynamic environments and the placing step in unstructured and dynamic environment.

to the normal optimal process. In fact, the changes to the optimal grip position through hardware calibration inaccuracies or collision with obstacles, together with damaged items through the transport or the placing steps have to be detected and adjustments have to be planned and executed.

This means that in an unstructured and dynamic environment the post gripping perception has to detect and localize not only objects in the environment, but also needs to detect and localize the events that have an impact on the process stability or quality with the help of anomaly detection. Anomaly detection is the task of classifying test data representing process deviations that differ from available training data representing the normal process as anomalies [55, 56]. For designing the proposed framework, functional components are defined based on common inputs and outputs and are used to build the framework modules. These modules are defined based on the pick and place process.

3.1 Building Components

In order to define the functions of the deep PGP framework, the perception tasks in the post gripping phase are analyzed, so that their common characteristics can be identified. This enables the definition of building components that can be used in different contexts in the framework. The input needed for one function and its output can be used as the common characteristics to define functional components. The input of each perception component is defined as the environment element that has to be perceived: the robot, task item or task target. The output of a perception building component in a post gripping context is the world pose of the environ-

ment element consisting of its position and orientation, the evaluation of its quality or the evaluation of its pose. The quality of an element is evaluated by describing the presence of unaccepted defects in it. For the quality evaluation, item properties are classified to define whether it should run the normal process or should be handled differently because of process quality violations such as damages or dirt. The pose evaluation means deciding whether the pose of the elements individually or in a group context (such as their in-hand, in-stack or in-target pose) lead to a success in the current and following manipulations.

Based on the description of the needed building blocks for post gripping perception, three building components are defined: object detectors, quality anomaly detectors and pose anomaly detectors:

- *Object Detector (OD)*: The function of this building component is the localization of a known object, which can be a task item or a target, in the robot base coordinate system based on image data. The pose of the robot is not considered, since it is intrinsically known. The input of this component is the gripping scene and the class of the object that has to be detected and its output is the object pose.
- *Quality Anomaly Detector (QAD)*: The function of this building component is the detection of a deviation from the normal quality of an object, which excludes its pose in the world. The input of this component is a scene with the object where the anomaly should be looked for and its output is whether an anomaly is detected.
- *Pose Anomaly Detector (PAD)*: The function of this building block is the detection of a deviation from the normal pose of an object or the relative pose between multiple objects. The input of this component is a scene with the objects to be considered in the pose evaluation and its output is whether an anomaly is detected.

3.2 Modules

The post gripping phase in a dynamic and unstructured environment leads to the need of a planner that computes the placing pose based on the robot-item relative pose and the target pose. Furthermore, a monitor that detects any deviation in the robot-item relative pose should be integrated. In case of a deviation to the normal process, the planner should be triggered. Finally, a verifier should be considered in order to detect any deviation in the placing outcome, which can affect further processes. The verifier also triggers the planner, if an outcome correction can be performed by the robot. Otherwise, it interrupts the process, in order to demand a human intervention. The mentioned planning, monitoring and verifying perception processes constitute the modules of the deep PGP framework as described in Fig. 3. For the design of deep PGP framework modules, their requirements are derived in
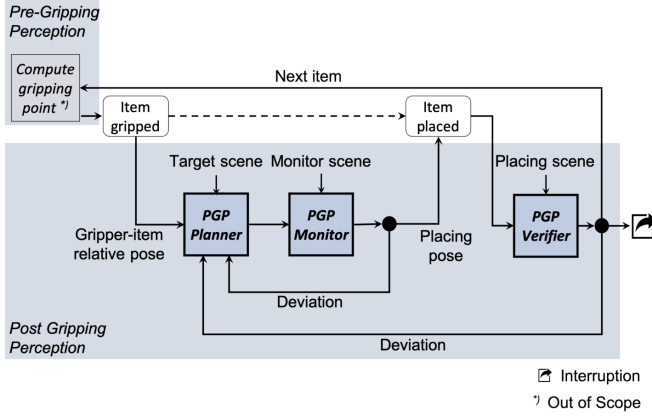
Fig. 3. The post gripping perception framework integration in the pick and place process.

Tab. 1. Furthermore, the realization of these modules with the introduced functional building blocks is described.

3.2.1 Planning Module

The post gripping perception planner (PGP planner) is the first module in the framework, whose output is needed after the robot grips task items. In order to plan the transport motion and placing actions, the target pose has to be defined. The main function of the PGP planner is to compute the placing pose based on three types of information: 1) the relative pose between the robot gripper and the task item after gripping it 2) the target class, where the task item should be placed, and 3) its pose. This is due to the fact that the targets in unstructured environments such as the intra-logistics are movable and selecting the target depends on the item class and its quality. For instance, damaged items have to be placed in a container for defective parts.

Based on the derived requirements from Tab. 1, the functional architecture of the PGP planner can be set. The main component of this module is a post gripping object detector (OD) introduced in Sec. 3.1. Since it has to deal with a high variance of objects with different backgrounds, different lighting conditions and different quality states such as soiled or damaged objects, a convolution neural network is integrated as the first element in this component. This can be justified by the ability of hidden layers in deep neural networks to extract hidden features that are invariant to the mentioned variance. The localized targets have to be validated. Only validated targets are selected for further processing based on prior spatial or process quality rules, in order to reduce false positive predictions. Finally, the computed placing poses in the image are transformed to real world poses based on the camera model and robot kinematics. The selected targets are candidates for the final target, that is selected based on the class of the item and its quality. The item quality can be predicted by a quality anomaly detector (QAD) introduced in

Sec. 3.1, which detects the items that present process quality violations and selects the defective container as their target. The QAD is composed of a deep auto-encoder followed by a classifier. This auto-encoder is composed of convolution layers in the encoder and decoder and reconstructs the input image based on extracted features in the hidden layers. The classifier determines if the auto-encoder reconstruction error corresponds to a normal or an anomalous class. The output of the PGP planner is a list of placing poses in real world coordinates for specific item classes. This information is then sent to the robot decision or execution, in order to derive the next actions based on the perception result.

3.2.2 Monitoring Module

In a dynamic environment, the post gripping process, especially the transporting step, can be affected by some changes that influence placing of the task item in the target. The main function of the post gripping perception monitor (PGP monitor) is to detect these deviations in the relative pose between the robot and the item. In a dynamic environment, the gripped item can slip in any direction or even get lost. Monitoring areas should be defined, where image data of the item gripped by the robot can be collected and deviations to the planned process can be detected. The output of this module should be the detection of a pose anomalies. In the positive case, placing pose adjustments has to be conducted by the PGP planner.

Based on the introduced requirements in Tab. 1, the functional architecture of the PGP monitor can be designed. Since this module deals with detecting a deviation from the normal relative pose between the robot and the item, its task can be formulated as a pose anomaly detection. This means detecting an anomaly in the scene based on a spatial context with the help of a PAD introduced in Sec. 3.1. Since this module has to deal with high variance in possible anomalies that can occur in the process, deep unsupervised learning techniques for anomaly detectors should be used. The PAD gets image data of the monitoring scene, which gets injected in a deep auto-encoder followed by a classifier. This auto-encoder is composed of convolution layers in the encoder and decoder and fully connected layers in the bottleneck. In fully connected layers, global features that can describe spatial relations can be extracted. The classifier determines whether the auto-encoder reconstruction corresponds to a normal or to an anomalous scene. If an anomaly is detected the image data of the monitoring scene gets forwarded to an object detector, where an updated placing pose should be computed, based on the old target pose and the new robot-item relative pose. The relative pose deviation can also be computed in an unsupervised way based on the difference between the input and reconstructed scene. The adjusted placing pose is then sent to the robot decision or execution, in order to derive next

Table 1. Requirements definition for the deep post gripping perception framework modules

Requirement	Planner	Monitor	Verifier
Input	color and depth image data	color and depth image data	color and depth image data
Output	target pose and class based on the item class and quality	relative pose state between the robot and the item	relative pose state between target and the items quality
Variance	high variance of the target classes, item quality and scene background	high variance in the item classes, its quality and the scene background	high variance of the item and target classes, their quality and the scene background
Modularity	integrable in different stages of the process, where placing pose adjustments are necessary	integrable in different stages after the gripping and before the placing, for deviation detection	integrable after the placing action, for the confirmation of the final task result
Predictions	False positives have to be more penalized for target pose prediction and false negatives have to be more penalized for item quality prediction, since they can lead to items being placed in wrong targets	False negatives have to be more penalized, since not detected anomalies can lead to a risk of damages to the items during the placing	False negatives have to be more penalized, since not detected pose anomalies can lead to unstable or damaged items in the target and not detected quality anomalies can lead to quality issues in further processes

actions based on the perception result.

3.2.3 Verifying Module

In a dynamic environment, process changes can affect the quality or pose of the task items and placing targets, not only during the transporting step but also during the placing step. This can lead to a misplacement that can only be detected after placing the item. For this purpose, the post gripping perception verifier (PGP verifier) is introduced. Its main function is to detect any deviations in the item quality or its pose in the target, such as an incorrect stack of items that can collapse. As input, the verifier is fed with image data of the target scene after the placing action is performed. It should detect two types of anomalies, namely a pose anomaly in the relative pose between the target and the item and a quality anomaly in the placed item. If a pose or quality anomaly is detected, the PGP planner can be triggered, in order to derive autonomous correction actions. If the placing result has an anomaly that cannot be corrected autonomously, the process should be interrupted and an external intervention is needed.

The PAD detector described in the monitoring section can also be used here for detecting the pose anomaly in a placing scene. The same scene can be used as input for a QAD to detect quality anomalies in the placed item, such as damages caused by the robot.

3.3 Overall Design

The deep PGP framework optimizes the use of the defined building components, in order to share the same archi-

tecture or data in different modules. Fig. 4 describes the overall design of the deep PGP framework, which is based on deep supervised learning for object detection, and mainly unsupervised learning techniques for quality and pose anomaly detectors.

The PGP planner is composed of an OD, that is fed by the target scene together with the item class so that the placing pose for each item can be computed. The list of placing poses can be generated at the start of each task, where multiple items have to be placed in one or multiple reachable targets. The second component in the PGP planner is a QAD that evaluates the quality of the considered item, in order to decide, whether it should be placed in the normal target or in the target for defective items.

Once the item is gripped, a monitoring step can be included before the placing action, in order to detect any anomalies in the item in-hand pose by the PAD of the PGP monitor. A detected pose anomaly triggers an adjustment of the placing pose by the OD in the PGP planner. For this purpose, the OD is used to localize the gripper and the item and compute their relative pose, called in-hand pose. The computed error in the in-hand pose is used to adjust the placing pose.

Once the item is placed, two anomaly detection steps in the placing scene regarding the item in-target pose and its quality are integrated in the process. This can be performed by the PAD for the in-target pose evaluation and by the QAD for the item quality evaluation. This can be performed from the same QAD used in the planning module. If no quality and pose anomalies are detected, the next pick and place item can be handled. A detected anomaly by the PGP verifier
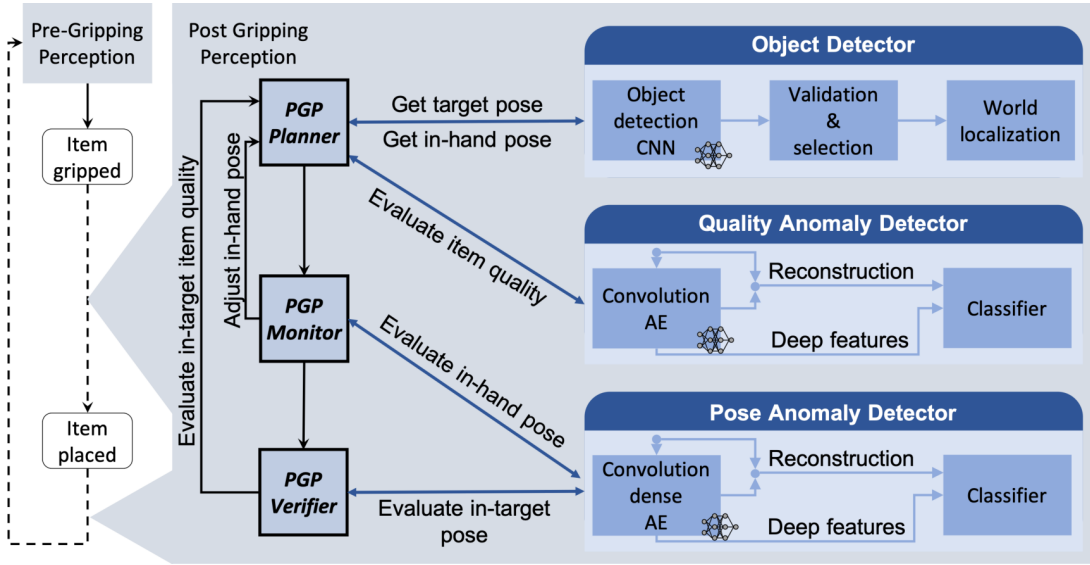
Fig. 4. The deep post gripping perception framework architecture: The components OD, QAD and PAD introduced in Sec. 3.1 are shared between the planning, monitoring and verifying modules (from top to bottom).

can trigger an autonomous correction of the outcome or an interruption of the process by the human supervisor, so that the pose or quality anomaly can be corrected.

Unsupervised learning saves the data labeling step, since no ground truth is needed. For this reason, the use of unsupervised anomaly detection techniques are prioritized in the PAD and QAD designs. These two components can save the use of the computation extensive object detection step, if no deviation is detected.

The deep PGP framework has a modular structure. The planning, monitoring and verifying modules can be combined depending on the task, process and environment of the robot. For instance, if the environment is only unstructured, where no quality control is required, the robot only needs a PGP planner, that determines the placing pose based on the target position in the world. In a dynamic environment, where the targets poses are fixed and do not have to be very precise, such as in use cases where objects can be dropped in bins, and only quality deviations can occur (such as damages to the items), the post gripping perception consists only of a PGP verifier, that detects quality anomalies. Once they are detected, information about the anomalies have to be indicated to the process supervisor, so that he can derive the needed actions.

4 Experimental Work

The target of the introduced deep PGP framework is to organize different perception functions that can be needed for the post gripping phase in a robotic process. In order to evaluate the advantages of this framework and prove its applicability, an implementation of the deep PGP framework in an experimental pick and place setup described in the first subsection is presented. The implementation in this work targets

the optimization of the third phase in the described robotic process, namely the placing step. For this purpose, two post gripping functions are integrated in the pick and place robot following the deep PGP framework. These functions are integrated as PGP planning and verifying modules.

4.1 Setup Description

For the evaluation of the proposed deep PGP framework, a packaging robot is studied in an experimental setup. The task is a logistics task consisting of picking drugstore items and stacking them in a packaging container, in order to be transported to another location. The presented items have the same dimensions and textures. However, they can have different quality, a normal state or an anomalous state. Having an anomalous state in a quality context means that the item is damaged or dirty, which makes it not admissible because of quality standards. The experimental setup is composed, as described in Fig. 5, of a 6-axis robot arm Universal Robot UR5e, a cobot vacuum pump with a gripper set from Schmalz and a stereo-vision color and depth image camera Intel Realsense D415. The target of the pick and place task is the container box, where the items are stacked. The baseline implementation of the robot conducts a gripping detection step based on a deep convolution neural network for object detection [57] with two object classes representing the items and the gripping points on top of the items. Once the gripping point is computed, the item can be gripped and transported to the fixed target container, where it can be stacked.

4.2 Implementation

In the described setup, two potential optimizations can be addressed: an increase of the process 1) quality and 2) stabil-

1 Robot **2** Camera **3** Items **4** Target
5 Quality anomaly **6** Pose anomaly

Fig. 5. The experimental setup for evaluating the introduced deep PGP framework.

ity. With the baseline implementation, the anomalous items can be gripped with the help of the object detection algorithm output. Then, they can be placed in the target container based on their classes. Placing quality violating items in a container for defective items increases the quality of the pick and place process. On the other hand, the non-deterministic object detection and camera calibration inaccuracies can lead to an unstable or shifted stacking. Correcting these misplacements increases the process stability, by allowing a stable and safe stacking in further steps.

For these purposes, planning and verifying modules are integrated in the existing baseline. For each module, a state in the state machine is added. The state machine is implemented with the help of the *smach* Robot Operating System (ROS) package. In each added module state, a ROS service is called in order to perform the PGP function. With this baseline extension no hardware modification is required. The data used for these functional components is collected with the existing end-effector camera shown in Fig. 5. The input data represents planning and verifying scenes from two poses executed by the robot in the planning and verifying states. For each of the integrated functions, a trained deep neural network is loaded from the library of the trained models. For the planning and verifying modules integrated in this experiment, anomaly detector based on deep convolution auto-encoders have been selected for the evaluation of the framework. Fig. 6 describes the architecture of the implementations tested in this experiment, namely the baseline implementation, the extension with the PGP planning module and the further extension with the verifying module.

4.2.1 Planning Module

The planning module in the presented deep PGP framework is a platform to host all functions that contribute to the computation of the placing pose P_p in a pick and place process. This computation is based on the pose of the target containers P_T and the classification vector of the gripped item c_I. P_T is a matrix containing the poses of the available target containers in the use case, either set by the user in a fix manner or computed by an object detector. c_I is the unity vector, with a one corresponding to the class of the gripped item. In the presented implementation, the function added to the PGP planning module detects whether the gripped item is anomalous following the quality standards. This anomaly detection defines the target container, whether it is the normal stacking container T_1 or a container for defective items T_2. The placing pose P_p is computed as

$$P_p = P_T \cdot c_I = \begin{bmatrix} p_{T_1} & p_{T_2} \end{bmatrix} \cdot \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{for normal items} \\ \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{for anomalous items} \end{cases} \quad (1)$$

The classification of the gripped item is realized by an unsupervised QAD, as described in the previous section. The anomaly detector is fed with a color image representing the detected item, which it uses to predict the item quality (normal or anomalous). By using an unsupervised learning technique, only normal unlabeled data is needed for training the anomaly detector. This data should be available in the running packaging process, since most of the items have normal quality and are ready to be stacked in the container. For the experiment in this work, three unsupervised anomaly detectors are evaluated, namely One-Class Support Vector Machines (OCSVM) [58], L2-norm deep convolution auto-encoder reconstruction based anomaly detection (L2-AE) [14, 59] and a Generative Adversial Networks based anomaly detection (AnoGAN) [60]. The evaluation of the PGP planning module is conducted based on the process quality rate (QR):

$$QR = \frac{\# \text{ items with normal quality in T1}}{\# \text{ items in T1}} \quad (2)$$

4.2.2 Verifying Module

The verifying module in the proposed PGP framework is a platform to host all functions that contribute to the decision making on the manipulation result. This decision defines further actions after placing the item:
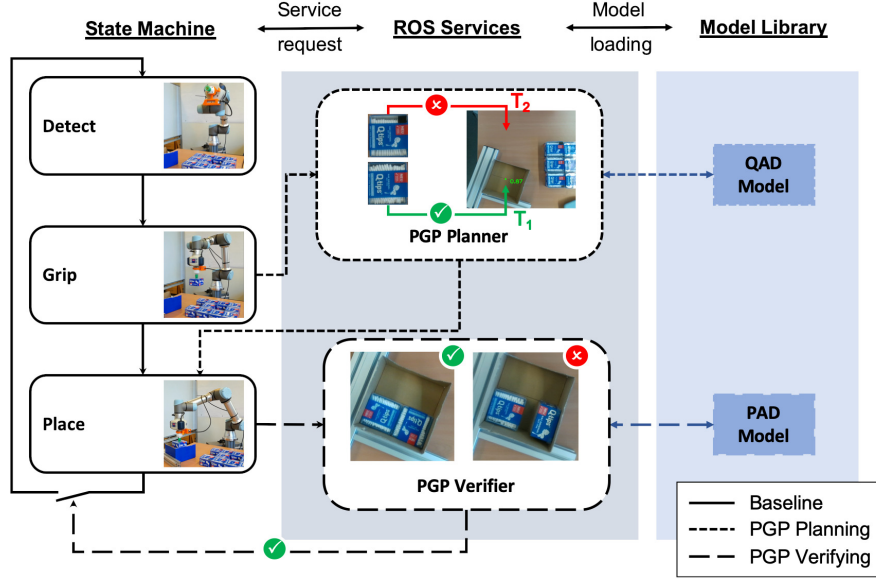
Fig. 6. The baseline implementation and the integration of the deep post gripping perception framework with two modules: planning and verifying.

- The placing is successful and further handling of new items is possible.
- The placing is not successful and an autonomous or manual correction is needed.

The classification of the manipulation result can be formulated as an anomaly detection problem, defining the result of a successful placing as normal and the result of an unsuccessful placing as anomalous. This classification function has been added to the verifying module in the presented implementation. The pose of placed items has to be evaluated in order to compute the result of this function. For this purpose, a pose anomaly detector, as described in the previous section, is integrated in the verifying module. The pose anomaly detector can also be unsupervised, since data from successful placings is available in the baseline implementation. For the pose anomaly detector of the verifying module, the same three anomaly detection methods from the quality anomaly detector in the planning module will be evaluated, based on the process stability (PS). The process stability is calculated as the ratio of the time needed for the correction of unsuccessful placings (correction down time) and the productive time needed for the pick and place process:

$$PS = \frac{\text{productive time}}{\text{correction down time} + \text{productive time}} \quad (3)$$

4.3 Results and Discussion

The anomaly detectors integrated in the planning and verifying modules are selected after comparing their performance

based on the harmonic mean of precision and recall [14], also called F1-score and computed as

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

Furthermore, the accuracy and prediction time of each method are summarized in Tab. 2. The presented results show that the deep learning based reconstruction approach L2-AE outperforms OCSVM in terms of F1 and accuracy. This can be explained by the ability of hidden layers in the reconstruction network to find the latent features of the data and describe its distribution. Since the prediction time of both approaches have the same scale of magnitude, the L2-AE is selected for the use in the QAD and PAD in this experiment. The iterative extension of the baseline implementation with a planning module followed by a verifying module is evaluated based on the defined indicators for the process quality and

Table 2. Comparison of unsupervised anomaly detection approaches for the QAD of the planning module: The dataset is composed of 447 normal images for training, and 63 normal and 85 anomalous images for testing. The prediction time is computed with an Intel i5 CPU and an NVIDIA GeForce GTX 1050 GPU.

Algorithm	OCSVM	L2-AE	AnoGAN
F1	0.74	**0.78**	0.41
Accuracy	0.63	**0.75**	0.42
Prediction time [s]	**0.01**	0.09	3.88

Table 3. Experimental results for evaluating the PGP modules compared to the baseline implementation. The mean and standard error of the mean are calculated for both implementations.

Implementation	Baseline	+ PGP
QR [%]	71.2 ± 1.9	$\mathbf{75 \pm 1.9}$
PS [%]	94 ± 0.3	$\mathbf{97.5 \pm 1.3}$

stability. In order to compute these parameters, ten packaging cycles with eight items each are conducted with each implementation. The results of the experiment are summarized in Tab. 3.

The results shown in Tab. 3 describe an optimization in process quality and stability brought by the baseline extension with two PGP modules. In fact, the PGP extension has enabled a relative increase of 5.3% and 3.7% in QR and PS respectively compared to the baseline implementation. The integration of both modules in the existing implementation in form of ROS services proves the modularity of the framework. The design of these modules minimizes the interfaces with baseline implementations, since the input data for both modules, which are color images from two scenes, can be captured by the existing camera in the setup. Only new poses have to be added to the process, where these scenes can be captured. The scalability of the proposed deep PGP framework can be supported by the unsupervised learning techniques that save development time for new functions, since no labeling effort is needed. The use of deep neural networks in the framework has shown the need for features learned by hidden layers for describing the data distribution.

The benefit brought by the PGP modules is higher than the error deviation of the mean for the quality (QR) and stability (PS). This means that the augmentation in these indicators with the help of the PGP modules is not random and can be extended. In fact, using other anomaly detection techniques that are able to achieve higher accuracy than the algorithms from Tab. 2, can lead to higher performance of the PGP framework.

Furthermore, the experiments conducted for evaluating the planning and verifying modules show that further optimization of the placing success rate can be achieved by placing pose adjustment in case the item deviates from the optimal gripping pose because of detection inaccuracies for instance. This adjustment can be triggered by a pose anomaly detector in the PGP monitoring module. The data collection for the anomaly detector can be automated, since the PGP verifying can decide on the class of the gripping result based on the result of the placing step. The monitoring module has been studied in further work about pose anomaly detection in deep post gripping perception for logistics robots [59].

5 Conclusion

This work introduces the post gripping perception as a research field that frames different computer vision efforts for intelligent robots to achieve robust placing. A framework for the integration of post gripping perception functions in industrial robots deployed in unstructured and dynamic environments is presented. The framework has a modular structure based on an optimized use of functional components realized with deep supervised and mainly unsupervised learning methods. The increase of the process quality and stability caused by the integration of modules from the introduced framework is shown in a real robotic experimental setup. While the design of the framework has been optimized by sharing the building components by different modules, the generalization and integration of such unsupervised learning based functions in an industrial context will be studied in future work.

References
[1] IFR, 2018. Robots and the workplace of the future. Tech. rep., International Federation of Robotics Frankfurt, Germany.
[2] Lammer, U. A., 2017. "Funktionsvereinigung in der lagertechnik". PhD thesis, Technische Universitaet Muenchen.
[3] Winkelhake, U., 2017. *Die digitale Transformation der Automobilindustrie: Treiber - Roadmap - Praxis*. Springer Vieweg.
[4] Salehi, V., and Wang, S., 2017. "Using point cloud technology for process simulation in the context of digital factory based on a systems engineering integrated approach". In Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 3: Product, Services and Systems Design.
[5] Salehi, V., 2019. "Development of an agile concept for MBSE for future digital products through the entire life cycle management called munich agile MBSE concept (MAGIC)". *Computer-Aided Design and Applications,* **17**, 05, pp. 147–166.
[6] Salehi, V., and Wang, S., 2019. "Munich agile MBSE concept (MAGIC)". *Proceedings of the Design Society: International Conference on Engineering Design,* **1**, 07, pp. 3701–3710.
[7] Kurrek, P., Zoghlami, F., Jocas, M., Stoelen, M., and Salehi, V., 2020. "Q-model: An artificial intelligence based methodology for the development of autonomous robots". *ASME Journal of Computing and Information Science in Engineering.*
[8] Kurrek, P., Zoghlami, F., Jocas, M., Stoelen, M., and Salehi, V., 2020. "Reinforcement learning lifecycle for the design of advanced robotic systems". In 3rd IEEE International Conference on Industrial Cyber-Physical Systems.
[9] Saloky, T., and Seminsky, J., 2005. Artificial intelligence and machine learning. KEGA research project, Virtual Program Modules of AI Systems.
[10] Russell, S., and Norvig, P., 2009. *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall Press, Upper Saddle River, NJ, USA.
[11] Ghosh, D., Olewnik, A., and Lewis, K., 2017. "Application of feature-learning methods toward product usage context identification and comfort prediction". *Journal of Computing and Information Science in Engineering,* **18**(1), 11, pp. 011004–011004–10.
[12] Shaw, S. B., and Singh, A. K., 2014. "A survey on cloud computing". In 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), pp. 1–6.
[13] Steder, B., 2013. "Feature-based 3D perception for mobile robots". PhD thesis, Uni Freiburg.
[14] Zoghlami, F., Kurrek, P., Jocas, M., Masala, G., and Salehi, V., 2019. "Usage identification of anomaly detection in an industrial context". *Proceedings of the Design Society: International Conference on Engineering Design,* **1**(1), pp. 3761–3770.
[15] Kurrek, P., Jocas, M., Zoghlami, F., Stoelen, M., and Salehi, V., 2019. "AI motion control a generic approach to develop control policies for robotic manipulation tasks". *Proceedings of the Design Society: International Conference on Engineering Design,* **1**(1), pp. 3561–3570.
[16] Jocas, M., Kurrek, P., Zoghlami, F., Gianni, M., and Salehi, V., 2019. "AI-based learning approach with consideration of safety criteria on

example of a depalletization robot". *Proceedings of the Design Society: International Conference on Engineering Design,* **1**(1), pp. 2041–2050.

[17] Zhang, H., Long, P., Zhou, D., Qian, Z., Wang, Z., Wan, W., Manocha, D., Park, C., Hu, T., Cao, C., Chen, Y., Chow, M., and Pan, J., 2016. "Dorapicker: An autonomous picking system for general objects". In 2016 IEEE International Conference on Automation Science and Engineering (CASE), pp. 721–726.

[18] Poss, C., Irrenhauser, T., Prueglmeier, M., Goehring, D., Zoghlami, F., Salehi, V., and Ibragimov, O., 2019. "Enabling robot selective trained deep neural networks for object detection through intelligent infrastructure". In Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering, CACRE2019, Association for Computing Machinery.

[19] Poss, C., Irrenhauser, T., Prueglmeier, M., Goehring, D., Zoghlami, F., and Salehi, V., 2019. "Perception based intelligent materialhandling in industrial logistics environments". In Proceedings of the 2019 11th International Conference on Computer and Automation Engineering, ICCAE 2019, ACM, pp. 146–151.

[20] Poss, C., Mlouka, O. B., Irrenhauser, T., Prueglmeier, M., Goehring, D., Zoghlami, F., and Salehi, V., 2019. "Robust framework for intelligent gripping point detection". In IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Vol. 1, pp. 717–723.

[21] Schuster, M. J., Okerman, J., Nguyen, H., Rehg, J. M., and Kemp, C. C., 2010. "Perceiving clutter and surfaces for object placement in indoor environments". In 2010 10th IEEE-RAS International Conference on Humanoid Robots, pp. 152–159.

[22] Jiang, Y., Lim, M., Zheng, C., and Saxena, A., 2012. "Learning to place new objects in a scene". *CoRR,* **abs/1202.1694**.

[23] Teodorescu, C. S., Vandenplas, S., Depraetere, B., Anthonis, J., Steinhauser, A., and Swevers, J., 2016. "A fast pick-and-place prototype robot: design and control". In 2016 IEEE Conference on Control Applications (CCA), pp. 1414–1420.

[24] Rengervé, A. D., Hirel, J., Andry, P., Quoy, M., and Gaussier, P., 2011. "On-line learning and planning in a pick-and-place task demonstrated through body manipulation". In 2011 IEEE International Conference on Development and Learning (ICDL), Vol. 2, pp. 1–6.

[25] Correll, N., Bekris, K. E., Berenson, D., Brock, O., Causo, A., Hauser, K., Okada, K., Rodriguez, A., Romano, J. M., and Wurman, P. R., 2016. "Lessons from the amazon picking challenge.". *CoRR,* **abs/1601.05484**.

[26] Patil, G. G., 2013. "Vision guided pick and place robotic arm system based on SIFT". In International Journal of Scientific and Engineering Research.

[27] Cosgun, A., Hermans, T., Emeli, V., and Stilman, M., 2011. "Push planning for object placement on cluttered table surfaces". In 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4627–4632.

[28] Ongaro, F., Yoon, C., van den Brink, F., Abayazid, M., Oh, S. H., Gracias, D. H., and Misra, S., 2016. "Control of untethered soft grippers for pick-and-place tasks". In 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 299–304.

[29] Kootbally, Z., Kramer, T., Schlenoff, C., and Gupta, S., 2017. "Overview of an ontology-based approach for kit building applications". pp. 520–525.

[30] Altan, D., and Sariel, S., 2020. "What went wrong?: Identification of everyday object manipulation anomalies". *ArXiv,* **abs/2001.09084**.

[31] Kothari, C. R., 2004. *Research methodology : methods & techniques*. New Age International (P) Ltd.

[32] Blessing, L. T., and Chakrabarti, A., 2009. *DRM, a Design Research Methodology*. Springer-Verlag London.

[33] Council, D., 2007. Eleven lessons: managing design in eleven global brands. A study of the design process. Tech. rep., Design Council.

[34] Edsinger, A., and Kemp, C. C., 2006. "Manipulation in human environments". In 2006 6th IEEE-RAS International Conference on Humanoid Robots, pp. 102–109.

[35] Yamashita, T., Godler, I., Takahashi, Y., Wada, K., and Katoh, R., 1991. "Peg-and-hole task by robot with force sensor: Simulation and experiment". In Proceedings IECON '91: 1991 International Conference on Industrial Electronics, Control and Instrumentation, pp. 980–985 vol.2.

[36] Bo, C., Hu, H., Xu, J., Liu, Z., and Chai, P., 2017. "Classification learning method for PCB insertion holes based on shape context". In 2017 3rd IEEE International Conference on Computer and Communications (ICCC), pp. 1933–1937.

[37] Wang, P., Qin, Z., Xiong, Z., Lu, J., Xu, D., Yuan, X., and Liu, C., 2015. "Robotic assembly system guided by multiple vision and laser sensors for large scale components". In 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1735–1740.

[38] Karako, Y., Moriya, T., Abe, M., Shimakawa, H., Shirahori, S., and Saitoh, Y., 2017. "A practical simulation method for pick-and-place with vacuum gripper". In 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 1351–1356.

[39] Cappelleri, D. J., Krishnan, G., Kim, C., Kumar, V., and Kota, S., 2010. "Toward the design of a decoupled, two-dimensional, vision-based μN force sensor". *Journal of Mechanisms and Robotics,* **2**(2), 04, pp. 021010–021010–9.

[40] Nozu, K., and Shimonomura, K., 2018. "Robotic bolt insertion and tightening based on in-hand object localization and force sensing". In 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 310–315.

[41] Hayakawa, S., Yamada, Y., and Tsuchida, N., 2004. "Shaft insertion for moving object by using robot manipulator with one dimensional PSDs sensor". In Proceedings of the 2004 IEEE International Conference on Control Applications, 2004., Vol. 2, pp. 1118–1123 Vol.2.

[42] De Gregorio, D., Zanella, R., Palli, G., Pirozzi, S., and Melchiorri, C., 2019. "Integration of robotic vision and tactile sensing for wire-terminal insertion tasks". *IEEE Transactions on Automation Science and Engineering,* **16**(2), April, pp. 585–598.

[43] Erkent, O., Shukla, D., and Piater, J., 2017. "Visual task outcome verification using deep learning". In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4821–4827.

[44] Yuan, P., Zhong, Y., and Yuan, Y., 2017. "Faster R-CNN with region proposal refinement".

[45] Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. "Imagenet classification with deep convolutional neural networks". In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. Curran Associates, Inc., pp. 1097–1105.

[46] Zhang, Z., Liu, J., Wang, X., Zhao, Q., Zhou, C., Tan, M., Pu, H., Xie, S., and Sun, Y., 2017. "Robotic pick-and-place of multiple embryos for vitrification". *IEEE Robotics and Automation Letters,* **2**(2), April, pp. 570–576.

[47] Komati, B., Kudryavtsev, A., Clévy, C., Laurent, G., Tamadazte, B., Agnus, J., and Lutz, P., 2016. "Automated robotic microassembly of flexible optical components". In 2016 IEEE International Symposium on Assembly and Manufacturing (ISAM), pp. 93–98.

[48] Harada, K., Tsuji, T., Nagata, K., Yamanobe, N., Onda, H., Yoshimi, T., and Kawai, Y., 2012. "Object placement planner for robotic pick and place tasks". In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 980–985.

[49] Toris, R., Kent, D., and Chernova, S., 2015. "Unsupervised learning of multi-hypothesized pick-and-place task templates via crowdsourcing". In 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 4504–4510.

[50] Cortes, C., and Vapnik, V., 1995. "Support-vector networks". *Mach. Learn.,* **20**(3), Sept., pp. 273–297.

[51] Baumgartl, J., 2016. "Schnelle, konservative greif- und ablageplanung von unbekannten objekten". PhD thesis, Universität Bayreuth.

[52] Balakirsky, S., and Kootbally, Z., 2014. *An Ontology Based Approach to Action Verification for Agile Manufacturing*. Springer International Publishing, Cham, pp. 201–217.

[53] Kootbally, Z., Schlenoff, C., Weisman, T., Balakirsky, S., Kramer, T., and Pietromartire, A., 2014. *A Simulated Sensor-Based Approach for Kit Building Applications*. Springer International Publishing, Cham, pp. 241–257.

[54] Hochreiter, S., and Schmidhuber, J., 1997. "Long short-term memory". *Neural Computation,* **9**(8), pp. 1735–1780.

[55] Chandola, V., Banerjee, A., and Kumar, V., 2009. "Anomaly detection: A survey". *ACM Comput. Surv.,* **41**, 07.

[56] Pimentel, M. A. F., Clifton, D. A., Clifton, L. A., and Tarassenko, L., 2014. "A review of novelty detection". *Signal Processing,* **99**, pp. 215–249.

[57] Redmon, J., and Farhadi, A., 2018. "Yolov3: An incremental improvement". *CoRR,* **abs/1804.02767**.

[58] Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., and Platt, J. C., 2000. "Support vector method for novelty detection". In *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K. Müller, eds. MIT Press, pp. 582–588.

[59] Zoghlami, F., Kurrek, P., Jocas, M., Masala, G., and Salehi, V., 2020. "Unsupervised pose anomaly detection for dynamic robotic environments". In 3rd IEEE International Conference on Industrial Cyber-Physical Systems.

[60] Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G., 2017. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery". *CoRR,* **abs/1703.05921**.

List of Figures

List of Tables