

## Article

# Vision-Based Guiding System for Autonomous Robotic Corner Cleaning of Window Frames

Tzu-Jan Tung <sup>1</sup>, Mohamed Al-Hussein <sup>1</sup> and Pablo Martinez <sup>2,\*</sup> 

<sup>1</sup> Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 2N1, Canada; tzujan@ualberta.ca (T.-J.T.); malhussein@ualberta.ca (M.A.-H.)

<sup>2</sup> Department of Architecture and Built Environment, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

\* Correspondence: pablo.rodriguez@northumbria.ac.uk

**Abstract:** Corner cleaning is the most important manufacturing step of window framing to ensure aesthetic quality. After the welding process, the current methods to clean the welding seams lack quality control and adaptability. This increases rework, cost, and the waste produced in manufacturing and is largely due to the use of CNC cutting machines, as well as the reliance on manual inspection and weld seam cleaning. Dealing with manufacturing imperfections becomes a challenging task, as CNC machines rely on predetermined cleaning paths and frame information. To tackle such challenges using Industry 4.0 approaches and automation technology, such as robots and sensors, in this paper, a novel intelligent system is proposed to increase the process capacity to adapt to variability in weld cleaning conditions while ensuring quality through a combined approach of robot arms and machine vision that replaces the existing manual-based methods. Using edge detection to identify the window position and its orientation, artificial intelligence image processing techniques (Mask R-CNN model) are used to detect the window weld seam and to guide the robot manipulator in its cleaning process. The framework is divided into several modules, beginning with the estimation of a rough position for the purpose of guiding the robot toward the window target, followed by an image processing and detection module used in conjunction with instance segmentation techniques to segment the target area of the weld seam, and, finally, the generation of cleaning paths for further robot manipulation. The proposed robotic system is validated two-fold: first, in a simulated environment and then, in a real-world scenario, with the results obtained demonstrating the effectiveness and adaptability of the proposed system. The evaluation of the proposed framework shows that the trained Mask R-CNN can locate and quantify weld seams with 95% mean average precision (less than 1 cm).



**Citation:** Tung, T.-J.; Al-Hussein, M.; Martinez, P. Vision-Based Guiding System for Autonomous Robotic Corner Cleaning of Window Frames. *Buildings* **2023**, *13*, 2990. <https://doi.org/10.3390/buildings13122990>

Academic Editor: Antonio Caggiano

Received: 18 October 2023

Revised: 20 November 2023

Accepted: 28 November 2023

Published: 30 November 2023

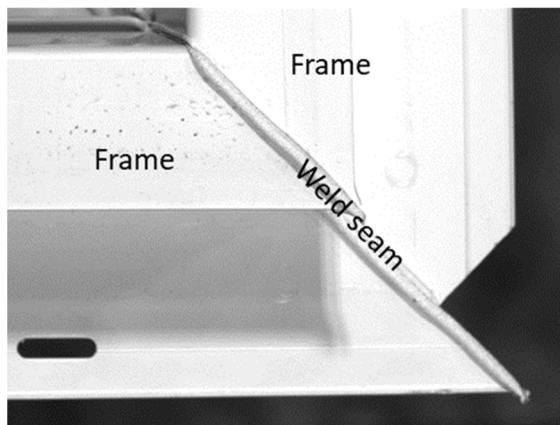


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the emergence of the concept of Industry 4.0 and the ongoing development of related technologies, automation in the construction industry has drawn increasing attention in recent years. Successful applications in the manufacturing industry in particular have inspired the pursuit of further automated applications in the industrialized construction field. With the integration of novel technologies such as robotic arms, computer vision, and various real-time sensors, repetitive tasks can be executed more flexibly and accurately [1]. In the window manufacturing industry, for instance, a typical plastic framing production line may consist of a combination of semi-automated or automated processes: an initial cutting station where frames are cut to length using saws, a hot plate welding station where individual frame elements are fused together to create the window frames, and a final corner-cleaning station where excess material on the corner of joined window frames is removed [2]. This excess material is generated as pressure is applied to different individual elements to secure the frame during the welding process. As such, the material to be

removed is always located on the exterior surface of the weld area, around the perimeter of the window frame. An example of the corner of a window frame with its main elements prior to cleaning can be seen in Figure 1. The current practice used by window manufacturers to remove the excess of material in the seams follows two sequential steps: (1) a series of knifing and milling tools, driven by linear actuators, are responsible for cleaning the window frame surface and leaving it as aesthetically pleasing as possible so that the weld seam is barely noticeable; and (2) a manual inspection ensures the quality of the cleaning process, and, if needed, a worker removes any residual seam material manually.



**Figure 1.** Overview of the weld area of a thermoplastic window frame.

However, this procedure has crucial flaws, including low adaptability to variations in window placement and unpredictable quality due to the volume of the weld seam material to be cleaned being unknown. Going into more detail, in the first part of the cleaning process, since the machine tasks are planned in advance based on idealistic CAD window models, the efficiency of the tool in cleaning the weld seam depends on the correct and accurate placement of the window, as well as the geometric precision and overall quality of the anterior processes (cutting and welding operations). Window placement, which is, in most cases, a manual operation, becomes more complicated with larger window frames, as they are more difficult to manually handle by workers and tend to cause ergonomic issues. Furthermore, the weld seam characteristics, namely, the volume and shape, change depending on the precision of the anterior hot plate welding process. Under the current practice, with its inflexible execution of the corner-cleaning process, even a small variance between the design and the manufacturing output, such as 1 mm in seam thickness or 1-degree difference in weld angle, can lead to an unsatisfactory outcome of the cleaning process. These limitations challenge the reliability of current corner-cleaning machines to achieve the desired outcomes and threaten product integrity. In fact, if product tolerances are considered, by introducing additional variability to the product status, the current approach to automated corner cleaning is rendered ineffective and incapable of ensuring product quality. By introducing visual sensors, for example, the cleaning process could be self-aware of the cleaning task ahead and perform the required motions (activities) without predetermined knowledge of the window frame. In addition, the capabilities of robotic systems to perform almost any motion accurately is a benefit that the cleaning procedure could benefit from to improve productivity and quality over the current practice.

In the second part of the cleaning process, the quality of manually removing excess material highly depends on the skill of the individual workers; therefore, the quality is uncertain and difficult to control due to individual differences. In some instances, the frame may become damaged to the extent that it fails to meet the quality standards of a marketable product. Although workers exhibit better adaptability to varying and unpredictable product specifications than machines do, this manual process has the disadvantages of uncertain quality and high labor cost. Due to these inherent disadvantages, workers only play an assistive role in the cleaning procedure. However, the manual inspection of the

window frames is still time consuming, tedious, and an inefficient use of labor resources. In fact, any additional work that operators have to do to surmount the inefficiencies of the automated cleaning process can be considered as waste (following lean thinking), and efforts must be made to reduce or completely remove the need for them. In summary, the current practice for cleaning weld seams in window manufacturing is suboptimal and needs to be improved. Automation in manufacturing has progressed significantly within the last decade, such that the use of robots and vision assistance on production lines can no longer be considered novel ideas. Combined with the rapid evolution of deep learning algorithms, which are now being widely applied in several research areas, manufacturing is gradually becoming more closely aligned with the vision of Industry 4.0, and several Industry 4.0 approaches have already achieved impressive outcomes in manufacturing [3]. Building on the success of these previous applications of Industry 4.0 principles, in the present research, a framework of vision-based robotic window weld cleaning system is proposed to overcome the flaws in the current corner-cleaning procedure. To enhance its adaptability to accommodate variability and ensure cleaning quality, vision techniques and robotics are incorporated in the framework to identify and locate window frames, detect the weld seams on window frames, and generate moving paths for the robotic arm. Specifically, the present study seeks to achieve the following two objectives: (1) to develop a working framework of a vision-based robotic system that guides the cleaning tool operation following window frame welding; and (2) to investigate the feasibility of Mask R-CNN image segmentation for window weld seam identification.

Achieving these objectives would represent a large improvement over the current practice in several aspects. First, it would enable automated welding of non-standard window frames, as current approaches are limited to standard 4- or 6-point boxes, by leveraging the higher order of degrees of freedom in robotic motion when compared to CNC linear actuators. Considering that these types of frames tend to be custom made and, hence, more profitable, having an adaptable automated system that can flexibly accommodate custom frames would reduce the reliance on manual labor. It is also expected to increase the accuracy of the cleaning process by actively detecting and adapting the cleaning path to each individual weld seam instead of predetermined paths based on ideal outputs from the welding process. This approach would allow manufacturers to rely further on the quality of the cleaning process and better manage uncertainty during the welding process.

## 2. Literature Review

### 2.1. Recent Advances in Window Frame Manufacturing

Windows are key components of a sustainable and well-functional building. The comfort level of the interior environment of a building is largely derived by physical variables that have close relationships with openings, such as lighting, acoustics, flow of air, humidity, and temperature [4]. To adapt to various climates around the world, strategies such as specialized window designs may be devised as passive design measures to better handle the outer environment [5]. For example, the feasibility of increased thermal resistance within windows can be discussed by analyzing its frame design and introducing novel combinations of materials [6]. In the current practice, depending on various contributing factors, the materials used for window frame production may include wood, aluminum, wood-cladding, fiberglass, polyvinyl chloride (PVC), and other composite materials.

Nonetheless, a current window performance analysis is always based on idealistic designs based on simple geometric features and does not incorporate potential manufacturing deviations and defects, which are always present in any manufacturing process [7]. This approach simplifies the design and testing processes but enforces strict quality control to ensure the thermal performance of manufactured windows. To support such efforts, research in operational performance, mostly to support more efficient management of framing processes, has been conducted for the last decade, and is especially supported by Industry 4.0 technologies and the advancement of data analytics.

On the one hand, the available research regarding applying engineering management methodologies for window frame production lines is numerous. In a recent publication, an application model has been proposed for operational planning and scheduling in window frame manufacturing [8]. The study showcases the capability of proper management strategies to minimize quality risks due to improper logistics and planning, while minimizing waste along the production line. Indeed, current research focuses on the application of well-known methodologies to reduce waste. A recent research study on the cutting processes of aluminum profiles for window frames was conducted to reduce material waste by optimizing the cutting patterns [9]. A lean 4.0 approach was presented to help decision-making regarding waste generated in the processes based on simulation models and decision support systems [10].

On the other hand, the required operations in window framing have not changed much in the past several decades. Window framing is based on permanent union manufacturing processes, e.g., hot plate welding in the case of thermoplastic profiles. Hot plate welding is a common welding technique to fuse components and is widely used in PVC window production due to its simplicity and relatively low cost. To achieve a marketable product, not only does the whole welding process require precise execution, but the weld seam cleaning task should also be applied with fine work. In fact, the weld quality between different window frame elements has a significant impact on the final product quality [2], while proper weld seam cleaning ensures a correct aesthetic finishing. Considering the current operations, two areas for future research seem necessary to improve window frame manufacturing operations: (1) modify or use an alternative manufacturing method to fuse window frames together that reduces or eliminates the need for posterior weld seam cleaning; or (2) modify the current window frame cleaning operations to increase accuracy and quality. This study explores the second research area and proposes the introduction of robotics, guided by vision systems, to replace current computer numerical controlled systems driven by linear actuators.

## 2.2. Vision-Guided Robots in Manufacturing

A wide range of vision-based methods and systems are applied in inspection and quality control processes, while also being combined with robot manipulation to enhance positioning accuracy. These methods include photogrammetry, stereo vision, structured light, time of flight, and laser triangulation [11]. Machine vision is gradually taking the lead in the blueprint of the future of robot guidance, and its applications have drawn increasing attention in the manufacturing field. The most intuitive and simple way to utilize vision is to mount cameras monitoring the working area or within the robotic system itself. Shah et al. proposed this type of approach for the automatic recognition of butt-welding joints. The authors mounted a fixed camera upon the working area to locate the position of the weld seam and create a path for the robot in two-dimensional coordinates, which simplifies the coordinates' transformation as long as the vertical distance between the weld element and the camera is fixed [12]. Abdelaal et al. conducted a similar experiment to control an industrial robot arm to pick and place moving objects with a real-time computer vision system to detect the object's exact location on a moving conveyor belt. The system conquers the challenges of dynamic picking point estimation with only two fixed cameras mounted surrounding the work area as a stereo vision system, which proved to be an efficient use of fixed cameras monitoring the working area to position objects precisely in dynamic situations [13].

However, there is no consensus on the most appropriate setup for vision-guided robotic operations. Several viable options are available based on the literature, and each study seems to provide decisions based on hardware limitations or cost effectiveness. For example, Kleppe et al. proposed an approach using a three-dimensional camera and a two-dimensional camera on a robotic arm to assist automated assembly tasks. The system first found a rough estimate of the target object posed with the 3D camera, and then the two-dimensional camera on the robotic arm was utilized to define a fine estimate

of the target. This approach avoids the viewing range limitation of robotic cameras by estimating the rough location first to guide the robotic arm to the working area and demonstrates a cooperative structure of fixed cameras and robotic cameras to position the target accurately [14]. In addition, with the popularization of deep learning approaches based on computer vision, a learning-based approach to hand–eye coordination for robotic grasping has been developed to enhance the awareness of collaborative pick and place robotic operations. Since the target objects are scattered randomly in this case, optimal real-time control is achieved using a deep learning model trained from the robotic arm grasping sensor and process data. The capability of handling variability from deep learning and vision systems was explored and validated [15]. Overall, vision-guided robots provide a more suitable platform to handle variability and clear capability to adapt to novel scenarios.

The advantages of welding robots have emerged for decades, including quality improvement and worker safety enhancement. Not only has vision-aided robotic welding been applied, but several applications in welding related scenarios have also been explored [16]. As a result, the utilization of supporting sensor technologies, such as vision sensing and self-learning neural control, have been deployed and investigated for robotic welding processes, as well [17]. While the developed vision techniques are involved in the control process for robotic welding, they facilitate the advancement of seam tracking and identification, which is an important task in which to automate robots on the correct welding path [18]. Since the age of Industry 4.0 began, robotics and deep learning have become crucial roles in the blueprint of future manufacturing systems, and the technology's evolution brings updated applications dealing with automated weld seam identification that enables intelligent welding operations. The combination of computer vision and welding robots carrying a camera provides considerably accurate weld seam identification [19]. As deep learning models are developed, the image processing techniques and the accuracy of welding seam recognition models have been improving as well, which also shows the potential of computer vision to handle dynamic environments by realizing the identification of multi-type weld seams [20,21]. Overall, the accuracy that can be obtained using such models can possibly increase as newer soft computing techniques appear in the future [22,23]. In order to provide an adaptive solution for welding seam cleaning, this paper proposes a vision-based robotic system integrating the use of robotic arms and computer vision, especially related to image segmentation techniques.

### 2.3. Weld Seam Detection

Detecting weld seams has become a much easier task thanks to recent developments in computer vision and deep learning. Image segmentation is a general designation of techniques utilized in digital image processing to define objects and boundaries by dividing the image into several meaningful segments. The ability of segmentation models to discretize weld shapes contained in images provides a clear step ahead from traditional image processing approaches, where shapes had to be defined by geometrical (edges) or numerical features (splines). The recent success of image segmentation relies heavily on the active development of accurate deep learning models, with significant improvement on feature extraction and real-time performance. Deep learning is generally proficient in tackling the issue of features learning so that it can overcome the limitation of domain knowledge requirements easily [24]. Among the sizable algorithms for image segmentation, region-based convolutional networks models, such as Mask R-CNN, are typically deployed for instance segmentation and have shown accurate capability in object detection and image segmentation applications. Mask R-CNN architecture is a modification of the Faster R-CNN by adding a third output of the model that generates an object mask on top of the original outputs, a bounding box, and a class label. This outcome is possible by replacing the region of interest pooling module with an align module; hence, most of the convolutional structure remains identical.

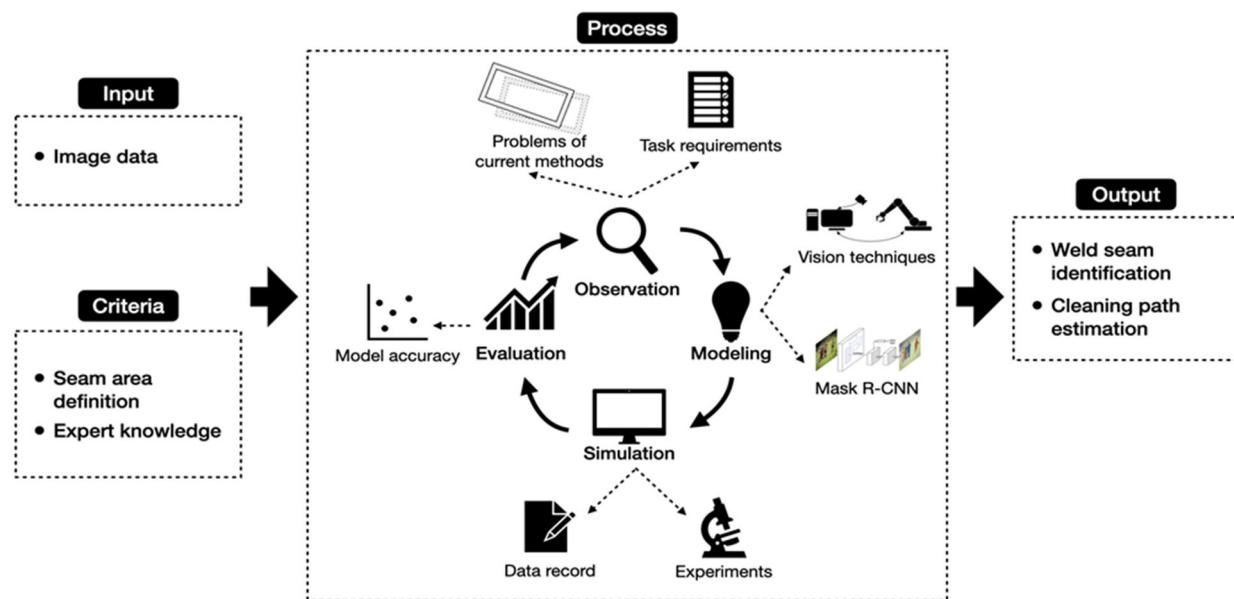
Applications of the Mask R-CNN model in manufacturing are numerous and can be found mainly in inspection systems. Due to the crucial role of assembly processes in the automation manufacturing field, the advantages of Mask R-CNN, including the capability of multi-object classification and identification, can simplify the inspection process by finding multiple defects at once [25]. A similar application is used in vehicle appearance component inspection, showcasing that Mask R-CNN-based methods show powerful capabilities to provide accurate detection of specific shapes and objects [26]. When it comes to applications in weld seam identification, the shape of weld seams is usually not as specific as the examples discussed above. However, a Mask R-CNN model was trained to detect the melt pool area within a visual sensing system for a robotic wire arc additive manufacturing system. Such a method shows the potential of detecting objects without a specific shape [27]. Even for micro-gap welds, a similar model could overcome the difficulty of recognizing welds with a width less than 40  $\mu\text{m}$  providing enough feature information regarding the weld seam [28]. In addition, a method was developed to recognize small batches. The results show the potential for more complex welding seam location, as pixel-level accuracy can be achieved [29]. The identification and classification were utilized in a trajectory planning method to achieve the pixel-wise coordinate data of small-target objects to be scanned, in which the data were extracted by a cloud point method based on Mask R-CNN [30]. Overall, the deep learning algorithms have shown great performance on vision-based weld identification. This study follows the current trend of using Mask R-CNN for the purpose of weld seam identification.

#### 2.4. Summary and Research Gaps

Industry 4.0 expands the scope of manufacturing innovation, particularly in the form of automated manufacturing solutions. Moreover, recent technological advancements provide the opportunity to integrate vision techniques and robot manipulation for tackling the dynamic operations and working conditions characteristic of automated manufacturing. The existing literature in this area touches on the current practice of window manufacturing, vision-guided robotics applications in manufacturing (i.e., welding-related cases), and applications of the Mask R-CNN deep learning model for image segmentation. The literature review above reveals a research gap, which is investigated in this paper. Although several welding-related applications of vision-guided robots have been explored, to the authors' knowledge, the welding seam cleaning application has yet to be investigated. As mentioned, the current practice has notable flaws in terms of low adaptability and unpredictable quality, so innovation of the cleaning process is needed in order to ensure a high-quality final product. In this context, to provide an adaptive solution for weld seam cleaning, a vision-based robotic system integrating the use of robotic arms and computer vision for image segmentation is a promising solution to fill the gap.

### 3. Methodology

For this research, a design science research (DSR) methodology was adopted to develop a vision-based robotic system that is capable of locating, identifying, and cleaning the weld seams at window corners. The DSR's goal is to develop an artifact: something that has a meaningful use and improves the understanding of a problem at the identified research gaps. The process of developing an artifact includes a rigorous procedure of literature review and developing the artifact following the evaluation methods in a structured and replicable manner while clearly communicating its outputs [31]. The artifact developed in this paper consists of a vision-based framework to enable robotic corner cleaning of window frames. In possession of this artifact, an alternative solution to the current industrial machinery is presented that increases flexibility and reliability in the operations. The methods applied in this research are demonstrated in Figure 2 and are divided into four stages:



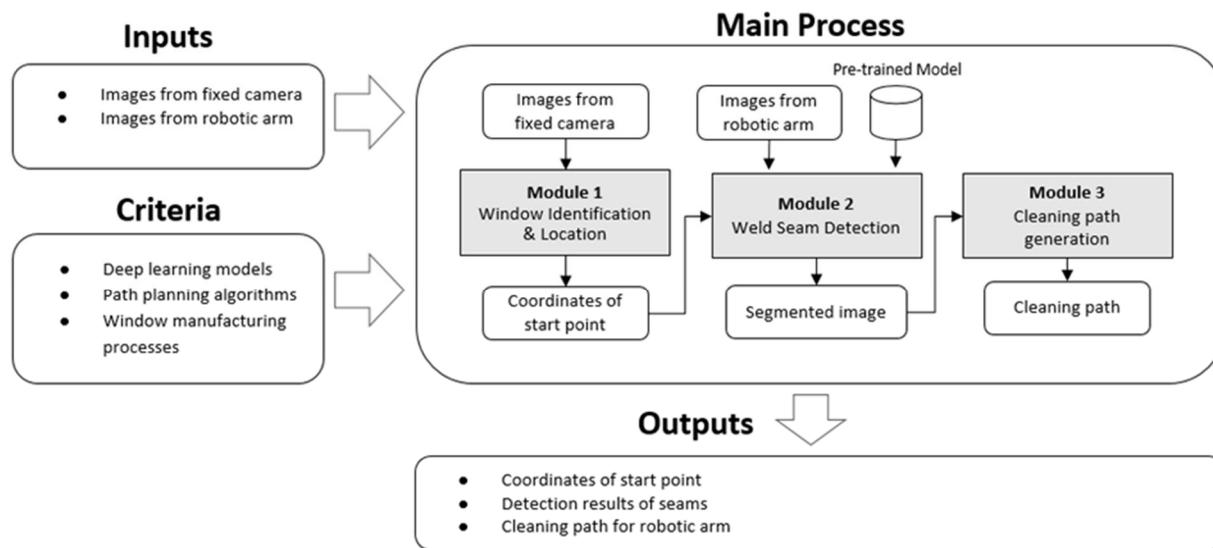
**Figure 2.** Overview of the proposed research methods.

In the observation stage, a concise understanding of the PVC window manufacturing was achieved by investigating the welding process outcomes and identifying problems with the current window cleaning methods and task requirements. The descriptions of the major flaws resulted in relevant data, which were used to identify suitable technologies that can minimize or eliminate the current process limitations. The following modeling stage involved the development of the necessary models based on the selected technologies: in this case, model building following the proposed approach with Mask R-CNN and vision techniques. Based on the information provided by the observation stage, the structure of a vision-based robotic system was developed utilizing computer vision algorithms for image segmentation and vision techniques to monitor and control the working environment. The simulation stage contained several experiments for validating the proposed approach and hypothesis in the previous stage. The data generated through the validation process in this stage were recorded and analyzed. In the evaluation stage, the proposed approach was evaluated and tested in a real scenario. The result performance compared with the hypothesis evaluated the efficiency of the proposed approach and revealed the limitations and potential. The definition of seam area and expert knowledge were required to describe the model training and operation involved, and the final outputs were provided as the performance of weld seam identification and cleaning path estimation.

#### 4. Vision-Based Robotic Corner Cleaning of Window Frames

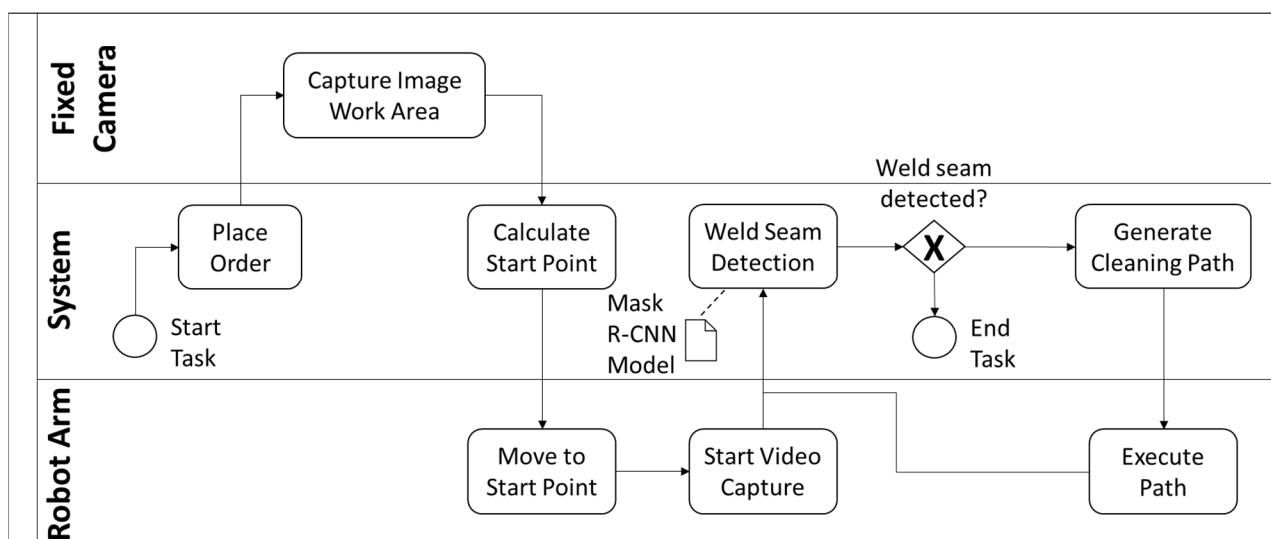
##### 4.1. Proposed Framework

The purpose of the framework presented in this study is to guide a robotic arm in executing weld cleaning on window frames with the assistance of computer vision. As shown in Figure 3 and as mentioned above, the proposed framework comprises three consecutive modules: (1) window and location identification; (2) weld seam detection; and (3) cleaning path generation. The first module uses the Hough transform to perform accurate edge detection from the image captured from a fixed camera to define the location of the target. Throughout the second module, the camera attached to the robotic arm captures a close-range image of the target. The framework processes the image to detect the seams with a pre-trained Mask R-CNN model and generates a mask to quantify the amount of weld seam to remove. In the last module, based on the information extracted from the previous stage, the system calculates a cleaning path for the robotic arm and transforms the coordinates from the two-dimensional image to three-dimensional real world. Each module is explained in detail in the following subsections.



**Figure 3.** Proposed vision-based window corner cleaning framework.

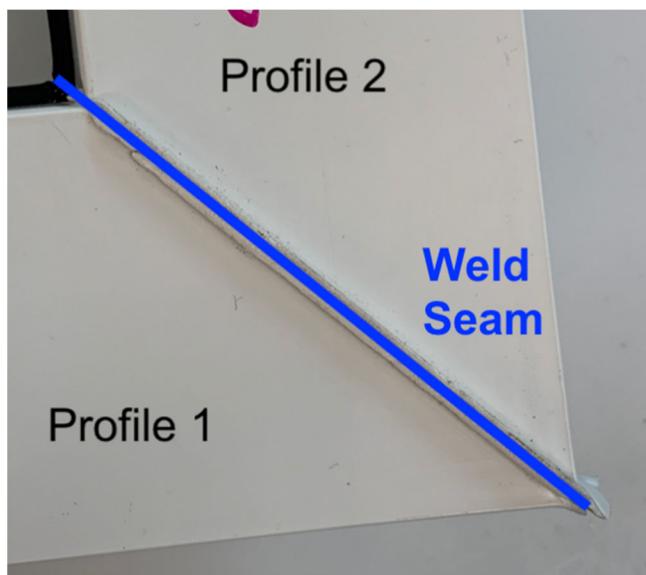
The framework's process model illustrates the process of one working cycle (where a cycle represents an individual detected weld seam) and the interactions and logic among the three main elements: the fixed camera, the robotic arm, and the control system. The model can be found in Figure 4 and represents the use case application of the proposed framework for a single robot cell, a single corner cleaning process. First, the working cycle starts once a task is assigned and an order is placed to the system. This order can be generated via user input or through proximity sensors that indicate the presence of a new window. The fixed camera is activated and captures an image of the work area in order to locate the window frame. Next, the system detects the location of the window's corner area, calculates the coordinates of the edge points, and sends it to the robotic arm as a target point. After the robotic arm moves to the location above the start point, the camera attached to the robot starts to capture images and the system detects if there is any seam in the image using the pre-trained image seam detection model. If no seam is detected, the order will be completed, and if any seam exists, the cleaning path for the robotic arm is generated and sent to the robotic arm. Then, the robotic arm executes the cleaning task just assigned and returns to the start point to restart the cycle again. The loop keeps working until there is no seam captured by the moving camera.



**Figure 4.** Proposed vision-based window corner cleaning process.

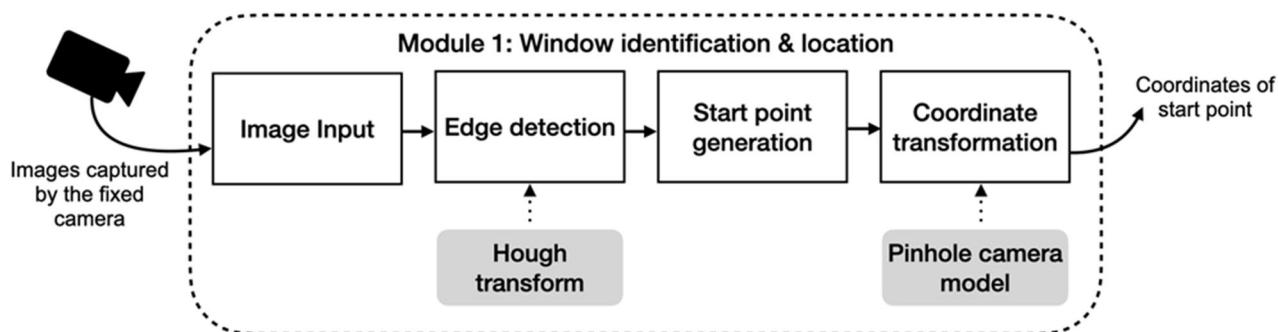
#### 4.2. Module 1: Window Identification and Location

In this module, images from the fixed camera are processed to obtain the current location and orientation of the window. This information will support an initial approximation of the robot towards its target. Considering that the vision range of the camera attached to the robotic arm is highly dependent on the current position of the robot, a fixed camera looking over the corner cleaning workstation is utilized to initially guide the robot towards the target and perform a controlled approach towards the weld seam. Since weld seams are produced due to the pressure applied on the window profiles, the only possible location for the generated weld seams on window frames is along the intersection of both profiles. An illustration of a frame corner area is shown in Figure 5.



**Figure 5.** Window profiles and frame corner with the weld seam.

To be more specific, the location is only possible on the intersecting line between the edge points on the corner of the window and with an identical inclined angle between both edges of the window (assuming orthogonal window boxes, which is the most common design). A summary of the operations performed in this module is shown in Figure 6.



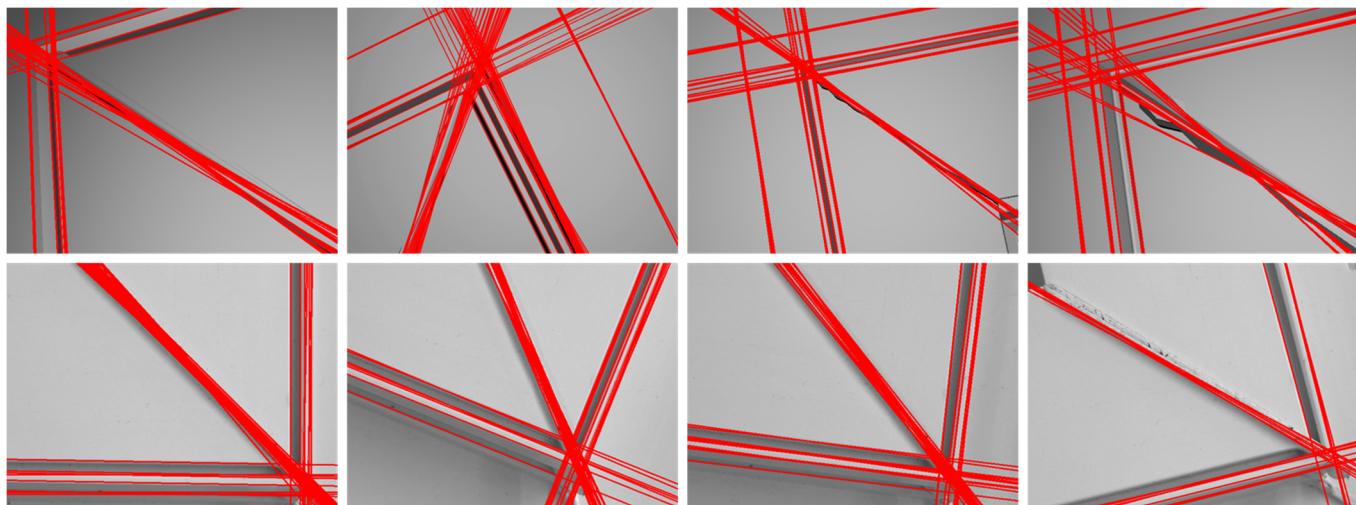
**Figure 6.** Proposed window identification and location system.

Based on the vision-based inspection method proposed by Martinez et al. [2], the edge detection method utilized here follows the same principle to detect the window edges. The Hough transform is a feature extraction technique widely used in image analysis to detect lines, among other geometrical shapes, which was first introduced to detect complex patterns of points in binary image data [32]. It is applied on the image captured by the fixed

camera and obtains the parametric lines that are recognized as edge lines, which follows Equation (1) below:

$$L(\rho, \theta) := \begin{cases} \rho = \sqrt{x^2 + y^2} \\ \theta = \tan^{-1}(\frac{y}{x}) \end{cases} \quad (1)$$

where  $L(\rho, \theta)$  is the set of lines defined by the Hough transform and  $(x, y)$  represent the coordinates of the points of each edge from the image. Sample image results where edge detection was achieved using the Hough transform can be found in Figure 7. As observed, three main groups of lines are consistently detected, namely, the edges of both individual frame elements and the weld seam. Thus, the detected lines can provide rough start points that can be generated by calculating the intersection points in the next step. As observed, the Hough transform can identify lines in any orientation, providing robust initial detection that is independent of the window placement operations. This is important, as it allows the algorithm to respond to variability in the angle that the window may be manually placed in the workstation by the operators.

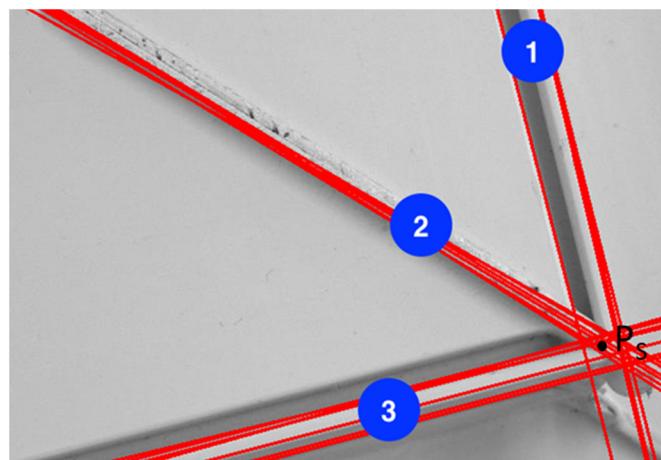


**Figure 7.** Examples of edge detection results using the Hough transform over a window corner. (**Top**) Image samples from a virtual environment. (**Bottom**) Image samples from the real environment.

Based on the slope of the lines, each line detected can be categorized into three clusters: (1) edges from the vertical frame element; (2) the weld seam; and (3) edges from the horizontal frame element. The clusters are obtained using k-means clustering around the value of  $(\theta_{cj})$  of each line. An illustration of the identified line clusters can be found in Figure 8. To simplify the calculation of the intersection point, each cluster is integrated into an average line,  $C(\rho_c, \theta_c)$ , to represent the cluster. The equation of  $C(\rho_c, \theta_c)$ , is defined in Equation (2) below:

$$C_j(\rho_{cj}, \theta_{cj}) := \begin{cases} \rho_{cj} = \frac{1}{n} \sum_{i=0}^n \rho_i \\ \theta_{cj} = \frac{1}{n} \sum_{i=0}^n \theta_i \end{cases} \quad (2)$$

where  $(C_1)$ ,  $(C_2)$ , and  $(C_3)$  are the average lines from each cluster, satisfying  $\theta_{c1} < \theta_{c2} < \theta_{c3}$ . Additionally, it is important to mention that the Hough transform may detect lines that belong to background elements that can be considered noise (lines detected in other orientations that do not belong to the window profile). This averaging step reduces the numerical impact that background detections may have on the approximation step as long as the correct detections are in much larger quantity than the background detections. As the field of view contains, purposefully, the window profile area and weld seam, it is assumed that background noise is not an issue for this application. This is confirmed later by the numerical results obtained by the experimental setup tested.



**Figure 8.** Numbered clusters of detected lines (in red) based on the Hough transform angle and final intersection point ( $P_S$ ).

Next, considering the possibility that the three lines cannot intersect at one point, three intersection points ( $p_i$ ) of any two lines from them are calculated, as defined in Equation (3). The average point of the three intersection points calculated is the final start point, ( $p_S$ ).

$$p_S = \frac{1}{n} \sum_{i=1}^n p_i, \text{ where } p_i = \frac{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 - a_2 & b_1 - b_2 \end{vmatrix}}, \forall (a_1, a_2, b_1, b_2) \in \mathbb{R} \quad (3)$$

where any two segments representing different clusters are defined by two sets of points, ( $a_1, a_2$ ) and ( $b_1, b_2$ ).

After the start point on the image is defined, the camera pinhole model is utilized to transform the coordinates from a two-dimensional image into three-dimensional space that serves as a global reference frame for the fixed camera and the robotic arm. Additionally, this transformation is necessary to provide the robot with a feasible motion target point. The relationship between a point  $M(X, Y, Z)$  in the global reference frame and its image projection  $m(x, y)$  is stated in Equation (4).

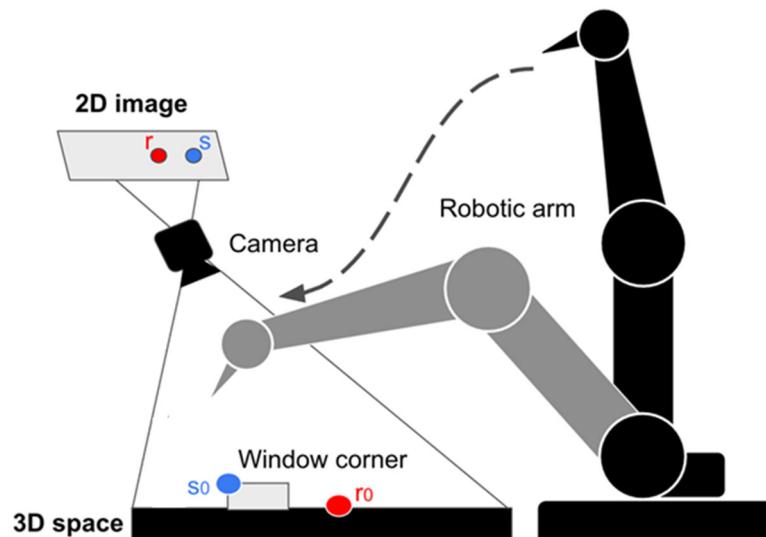
$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = A[\mathbf{R} \ \mathbf{t}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \text{ with } A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

where ( $s$ ) is an arbitrary scale factor;  $[\mathbf{R} \ \mathbf{t}]$ , known as the extrinsic parameters, are the rotation and translation matrices, respectively, which relate the global coordinate system to the camera coordinate system; and ( $A$ ) is called the camera intrinsic matrix with  $(u_0, v_0)$  being the coordinates of the principal point, ( $\alpha$ ) and ( $\beta$ ) the scale factors in the image on the  $u$  and  $v$  axes, respectively, and ( $\gamma$ ) the parameter describing the skew of both the image axes. The parameters above can be obtained using a well-known camera calibration method [33]. The application of the camera pinhole model is shown in Figure 9, where ( $r$ ) and ( $s$ ) are the image counterparts of the reference ( $r_0$ ) and sample ( $s_0$ ) three-dimensional points. Therefore, the target for the robot is computed following Equation (5).

$$P_S = A^{-1}[\mathbf{R} \ \mathbf{t}]^{-1} p_S + h_S \quad (5)$$

where ( $h_S$ ) is a safety coefficient that ensures that any potential camera calibration error does not cause a collision between the robot and the window during the approximation motion. This coefficient is in the order of magnitude of several centimeters in each axis for the robot arm chosen in this study, as it must account for the camera calibration error, the

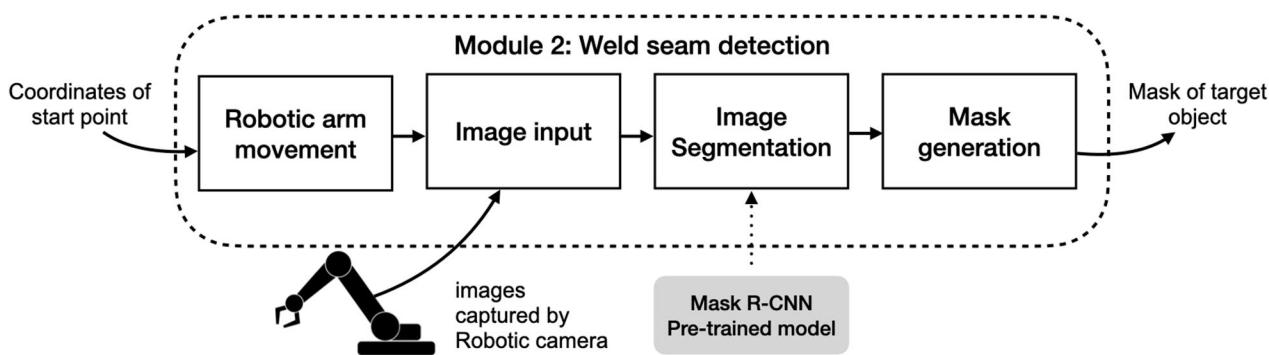
robot motion accuracy, and the end effector (tool) size. In the end, the expected position of the end effector is a few millimeters away from one of the edges of the weld seam that requires clearing.



**Figure 9.** Application of the pinhole model in the studied workstation.

#### 4.3. Module 2: Weld Seam Detection

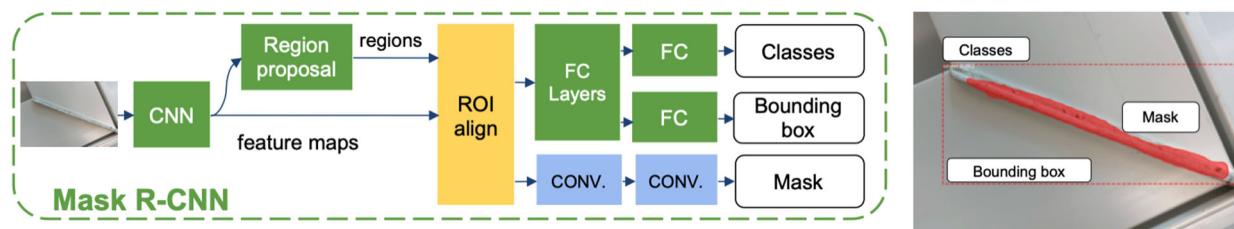
In this module, the goal is to determine the boundaries of the area of the weld seam about to be cleaned in order to guide the corner cleaning tool through the material. The process starts with the robot approximation following the coordinates recently obtained. Meanwhile, the camera in the robot turns on and starts streaming images to the main computer system. Then, the trained Mask R-CNN model looks for the weld seam, and the obtained mask is stored for future use. A summary of the operations performed in this module can be found in Figure 10.



**Figure 10.** Proposed vision-based weld seam detection system.

First, as the final output of the first module is a set of three-dimensional coordinates, these coordinates need to be translated into a language that the robot arm can understand. In this study, considering that a Universal Robots UR5e robot arm is being used, the coordinates are translated using RoboDK motion planning following the URScript programming language. For this particular motion, a “go-to-point” approach is used, which is represented by the “movej” function. A typical format for the use of that function in URScript is `movej(q, a, v, t, r)`, where `(q)` is a vector with all the robot joint’s positions (translated from the coordinates obtained previously), `(a)` is the robot joint’s acceleration along the leading axis, `(v)` stands for the joint’s speed along the leading axis, `(t)` is the time delay (usually 0), and `(r)` is the blend radius (usually zero).

Once the robot is in motion, the wrist camera attached to it is powered on and starts transmitting its video feed to the main computing system. Then, image processing techniques are used on the video feed to detect the weld seam and locate its area boundaries. Weld seams are not generated with a particular geometrical shape, mostly due to the nature of the hot plate welding process, where thermoplastic material is squeezed from both sides unevenly. This raises the level of difficulty to locate the exact area of any weld seam with simple feature recognition image processing techniques; however, image segmentation techniques can be utilized to define objects and boundaries by separating the image into several segments of a non-predetermined shape. For the purpose of segmenting the weld seam and accurately determining the boundaries of the area to be cleaned, a Mask R-CNN network is trained from scratch. An example of the desired detection result and the Mask R-CNN architecture is shown in Figure 11.

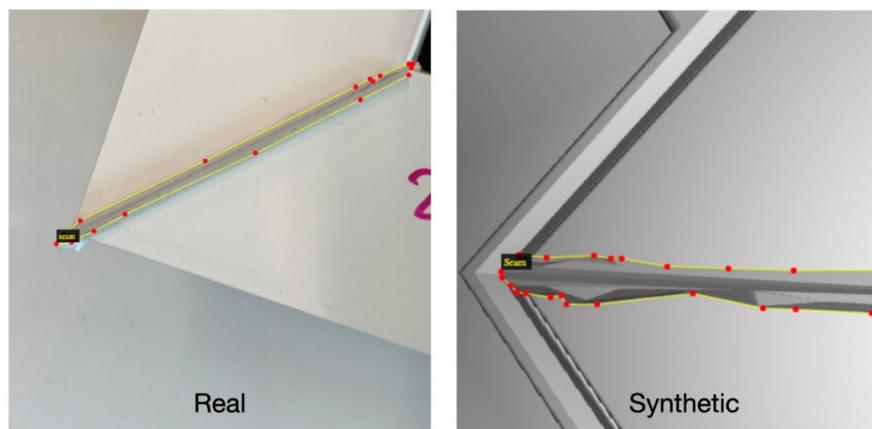


**Figure 11.** Diagram of the Mask R-CNN architecture with an example of a desired detection result over a window frame corner.

To train a weld seam detection model that can identify the boundaries of the area to be cleaned correctly, several steps are required:

Step (1): images of uncleared corners of window frames are collected as training material. The image dataset is constructed using images taken from the Robotiq wrist camera integrated within the robot arm. The dataset contains 396 images of weld seams of real windows complemented with 421 images of synthetic weld seams made in CAD software, for a total of 817 images.

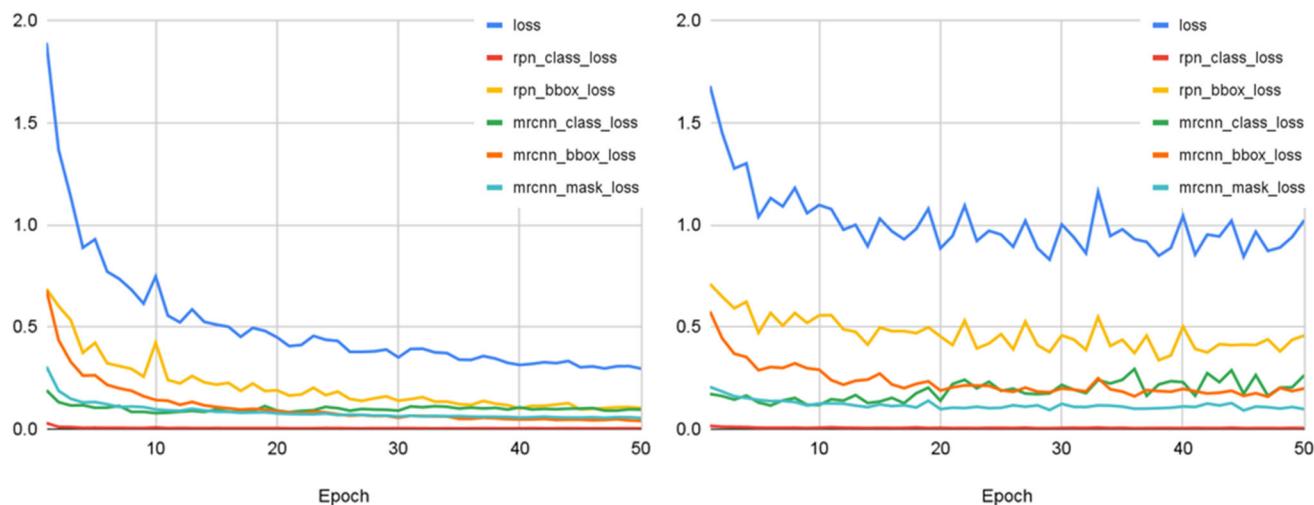
Step (2): to teach the model to recognize the weld seam, each image is manually labeled and divided into two datasets, the training dataset and the validation dataset, with a proportion of 3 to 2. The software utilized is VGG Image Annotator (VIA), which is an online manual annotation software developed by the Visual Geometry Group at the University of Oxford [34]. Two examples of images from the training database can be found in Figure 12.



**Figure 12.** Example of two labeled images from the training dataset.

Step (3): the model is trained with the training dataset, and then the resulting model is tested with the validation dataset. The model is trained within Google Collaboratory with GPU boosting and 12 GB of RAM, with Tensorflow (version 1.15.0) and Keras (version 2.1.5) installed. The COCO dataset, which is a large-scale object detection, segmentation, and captioning dataset, is used as the initial training weights on the pre-trained Mask R-CNN model. With this computational environment, the training process took 3.5 h to complete 50 epochs of model training.

The model had a final training loss of 0.2959 and a final validation loss of 1.0228, as shown in Figure 13. As expected, the model's loss (error) gradually decreased as the number of training epochs (runs) increased. In the Figure 13, “loss” indicates the overall model loss during the training and validation steps and the other indexes such as “rpn\_class\_loss” and “rpn\_bbox\_loss”, which stand for the region proposal losses for the classification and localization (bounding box); and “mrcnn\_class\_loss”, “mrcnn\_bbox\_loss”, and “mrcnn\_mask\_loss” stand for the classification, localization (bounding box), and segmentation (mask) losses during the with Mask R-CNN training and validation steps. It is important to note that for the desired application of this neural network, it is paramount that the final value for “mrcnn\_mask\_loss” is as small as possible. With the used dataset, the final segmentation loss achieved is 0.0542 for the training and 0.0976 for the validation.



**Figure 13.** Mask R-CNN loss results using the training (**left**) and validation (**right**) datasets.

#### 4.4. Module 3: Cleaning Path Generation

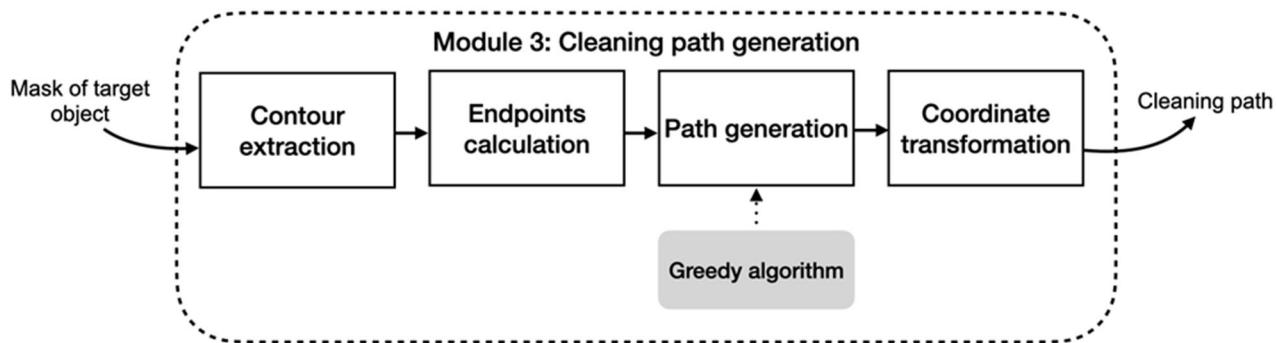
In this module, the cleaning path is finally generated based on the object mask obtained from the Mask R-CNN model detection. The overall module procedure can be found in Figure 14. First, the contour of the mask is extracted, as it represents the boundaries of the area to clean. The extracted contour is not of a specific geometrical shape; however, it usually has a string-like shape, being that it is much longer than it is wide. Usually, weld seams are around 2 to 5 mm wide with a longitude from 5 to 15 cm. The most common and most efficient approach given the tool used to clean weld seams is to go through the centerline of the seam with a wide enough tool in a single pass; therefore, the objective of this module is to estimate the centerline of the weld seam.

As the product of the last module is a strip-like mask defining the seam area, the first step in this module is to extract a contour from the mask. This contour defines, with as many points as necessary, the mask obtained by the Mask R-CNN model. Each contour point is stored in a list and assigned an index following its location on the contour. Due to the strip-like shape of the area, by comparing the distances between any two points on the

contour, the two points with the longest distance are chosen as the start and end points for the robot's path. Both endpoints satisfy Equation (6) below.

$$(E_1, E_2) = \{(p, q) \mid \max(d(p, q))\} \text{ where } d(p, q) = \sqrt{(q_x - p_x)^2 + (q_y - p_y)^2} \quad (6)$$

where  $(E_1)$  and  $(E_2)$  represent the start and end points for the robot's path and  $d(p, q)$  is the Euclidean distance between two cartesian points  $p(p_x, p_y)$  and  $q(q_x, q_y)$  on the contour. These end points have the property that they belong to the frame and the weld seam. Thus, assuming that the desired outcome of the cleaning procedure is a flat surface, all the points in the cleaning path will have the same depth coordinate, obtained from either  $(E_1)$  or  $(E_2)$ . This simplification allows the path to be generated without estimating the weld volumes and using only the detected areas.



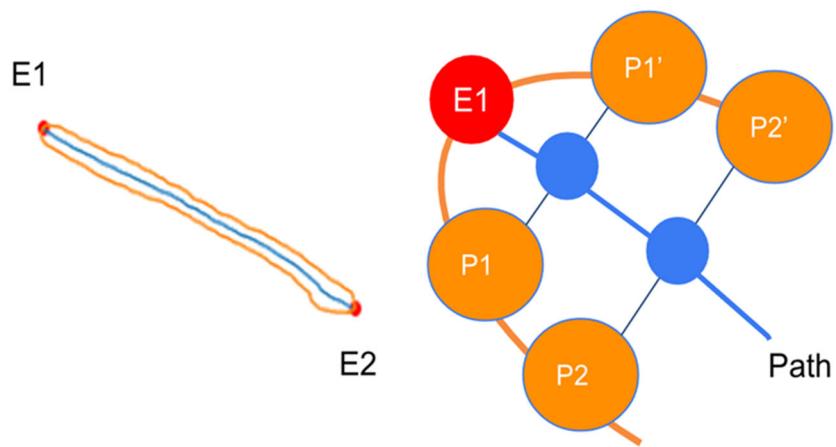
**Figure 14.** Proposed cleaning path generation system.

From the remaining contour points, the robot's path is determined using the greedy algorithm. This algorithm is usually applied when obtaining an optimal result is not feasible in a single solution or iteration, and it divides the optimization problem, in this case, a path planning problem, into several minimization problems that aim to yield locally optimal solutions at every iteration and eventually reach an optimal solution. Thus, the cleaning path generation is calculated by dividing the whole area into several sections and gathering the local center points from each section as a final path. To clean the weld efficiently, the path should go through the center of the seam, so the best local solution for the division is determined as the midpoint of each two counterpart points. All  $(N)$  points are defined while extracting the contour; thus, the following points of the robot path are the midpoints,  $(CP_i)$  of the previous and next points, represented by  $(P_i)$  and  $(P'_i)$  in Figure 15. Consequently, the third point in the path, named  $(CP_2)$ , is the midpoint of  $(P_2)$  and  $(P'_2)$ , and so on.

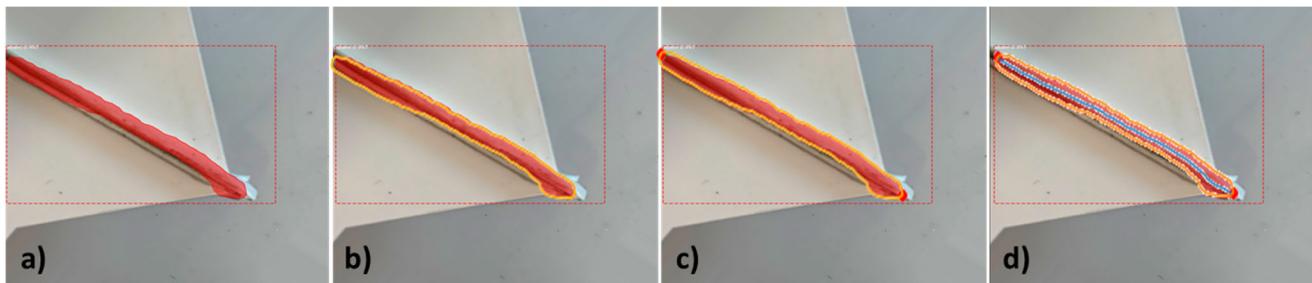
Therefore, generalizing for all the contour points  $(N)$ , the set of points that comprise the cleaning path,  $(CP_i)$ , can be determined. Let  $(k)$  be the index of the starting point  $(E_1 = N_k)$ . Then, all the contour points can be described referring to the starting point, following the nomenclature in Figure 15,  $(P_i = N_{k-i})$  and  $(P'_i = N_{k+i})$ . Note that both  $(P_i)$  and  $(P'_i)$  can be reversed and the path generated would remain identical. Thus, the coordinates of all the points in the cleaning path are defined by Equation (7) below.

$$CP_i(x_{ci}, y_{ci}) := \begin{cases} x_{ci} = \frac{x_i + x'_i}{2} \\ y_{ci} = \frac{y_i + y'_i}{2} \end{cases} \quad (7)$$

where  $P_i (x_i, y_i)$  and  $P'_i (x'_i, y'_i)$  are the corresponding points on the contour. A step-by-step overview of the path generation process can be found in Figure 16. To obtain the depth coordinate of the cleaning path and guide the robot motions, a similar effort to the one mentioned previously (see Equations (4) and (5)) is necessary.



**Figure 15.** Diagram of the application of the greedy algorithm over the weld seam contour to determine the robot's path.



**Figure 16.** Illustration of the steps required to generate the corner cleaning path alongside the detected weld seam. (a) Mask R-CNN detection mask (red area); (b) contour extracted from the mask (yellow line); (c) estimated start and end points for the corner cleaning operation (red points); (d) cleaning path obtained using the greedy algorithm (blue line).

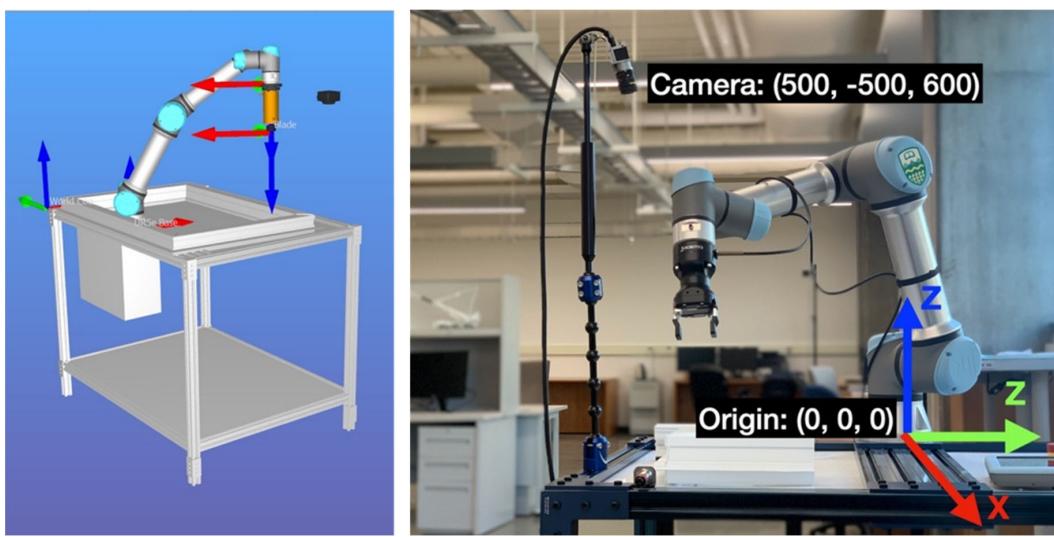
## 5. Results

This section aims to validate the proposed vision-based robotics system in two different environments: first, a simulation environment and then, a real scenario. In the following subsections, the experimental setup is initially defined, then the results are presented and analyzed following the structure of the three modules in the system, and, lastly, the limitations are discussed.

### 5.1. Experimental Setup

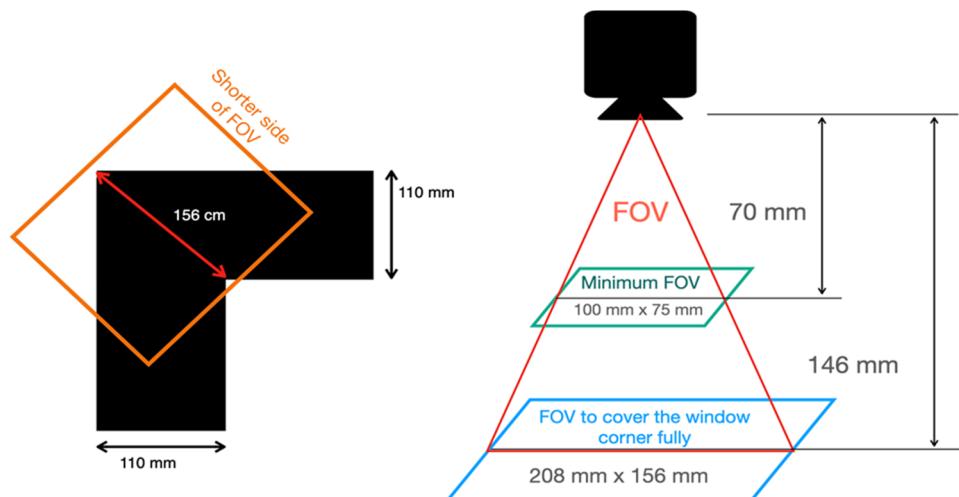
As an experimental setup to validate the proposed approach to corner cleaning in window frames, two cameras and a robot arm were utilized. These included a Basler ACE camera with an Optron 35 mm lens (fixed camera), which monitored the working area, and a Robotiq wrist camera attached to a Universal Robots UR5e. The virtual environment was built within the RoboDK software, providing a true representation of the real environment. In the robotic reference frame, the origin was defined on the center base of the robot, which made the coordinates of the fixed camera position as  $(500, -500, 600)$  in millimeters. Both experimental setups in their environments can be found in Figure 17.

The windows used during the reported experiments were single box frames made of polyvinyl chloride (PVC) that are mainly used for residential purposes. In more detail, the height of these frames was  $3.26 \pm 0.15$  inches ( $82.80 \pm 3.81$  mm) and their width was  $4.31 \pm 0.20$  inches ( $109.47 \pm 5.08$  mm). Their geometry was modeled in CAD software before being transferred to the virtual environment in STL format.



**Figure 17.** Validation environments for the proposed vision-guided robot system. **(Left)** Virtual environment in RoboDK software 5.6. **(Right)** Real-world setup.

For the wrist camera attached to the robot arm UR5e, its field of view was between 100 mm by 75 mm and 640 mm by 480 mm, and the focus range started from 70 mm to infinity. Based on the specifications and the height of the window researched in this study, a required minimum field of view (FOV) to fully cover the corner area is around 110 mm by 110 mm; however, considering that the placement angle of the window frame might vary, the narrowest FOV should not be less than 156 mm, which is the maximum length of the seam, as shown in Figure 18. If the shorter width should reach at least around 156 mm, then the working distance from the camera should be around 146 mm above the object. In addition, the height of the window frame (83 mm) should be taken into consideration, as well. Therefore, the proper vertical distance from the camera to the working table should not be less than 229 mm, which equals the height of the window plus the working distance. In order to reserve enough space for the end effector (cleaning tool), a fixed distance was eventually set to 300 mm in the z axis. Since the camera distance to the workstation was fixed, as well as the height of the robotic arm at the initial position, the height measurements (z axis) were kept as a constant true value and not estimated to simplify the system's operations.



**Figure 18.** Camera placement considerations. **(Left)** Illustration of the decision for the minimum width of field of view. **(Right)** Illustration of the calculation for the minimum working distance needed to completely cover the window corner area.

## 5.2. Framework Validation

### 5.2.1. Module 1

The first module, window identification and location, aimed to locate the rough position of the start point to enable the robot to perform an initial approach towards the corner of the window frame. To validate the coordinate estimation process, several window corners were placed one by one at a random position within the working area. Then, the system was run to obtain the location estimation. The actual location was measured as well using a manual tape measure with 1 mm of accuracy to provide the ground truth of the experiment. The results of locating the window corner area are presented in Table 1.

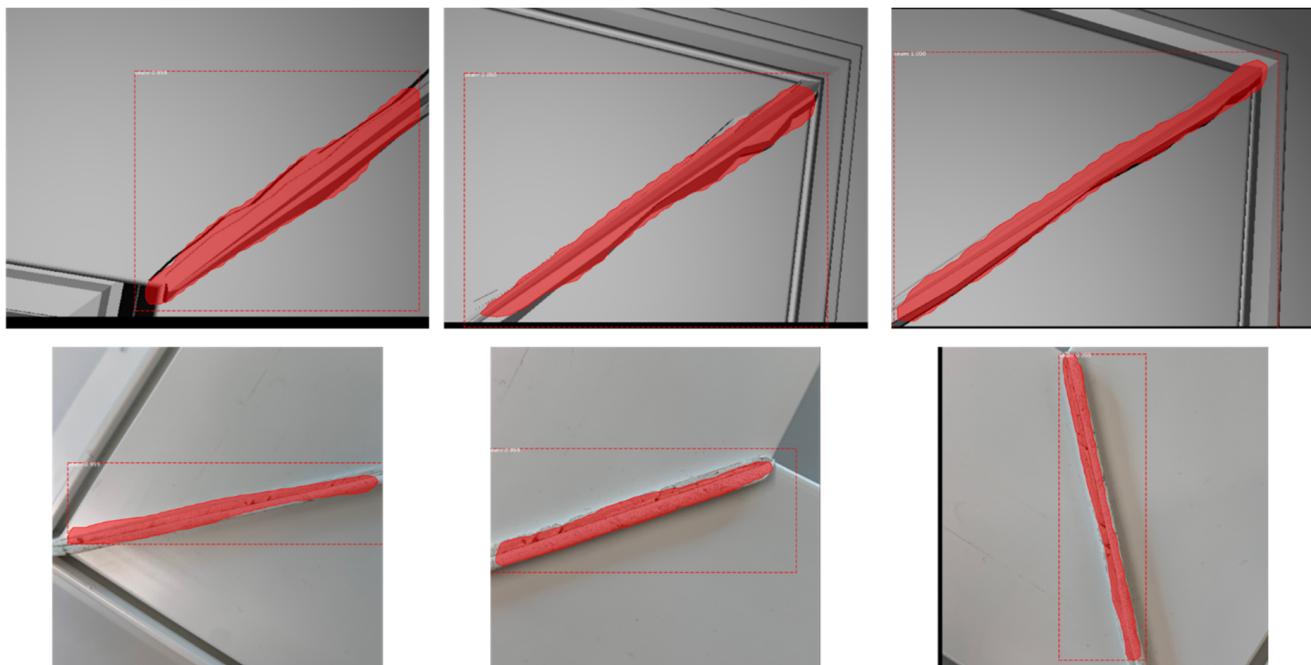
**Table 1.** List of the starting point estimations using the proposed system in module 1.

#	Coordinates of the Starting Point (x, y)									
	Virtual Environment					Real Environment				
	Estimated		Actual		Error	Estimated		Actual		Error
#	x [mm]	y [mm]	x [mm]	y [mm]	ε [mm]	x [mm]	y [mm]	x [mm]	y [mm]	ε [mm]
1	494	-497	496	-499	2.828	486	-510	489	-506	5.000
2	499	-490	501	-503	13.153	470	-488	467	-479	9.487
3	488	-504	492	-510	7.211	478	-502	481	-496	6.708
4	512	-508	510	-516	8.246	478.0	-497	475	-499	3.606
5	502	-509	507	-501	9.434	472	-500	471	-502	2.236
6	469	-405	476	-401	8.062	496	-408	500	-401	8.062
7	490	-388	497	-391	7.616	512	-388	500	-390	12.166
8	482	-393	480	-405	12.166	488	-420	494	-415	7.810
9	485	-402	480	-398	6.403	496	-420	500	-412	8.944
10	483	-401	481	-405	4.472	497	-413	498	-400	13.038
11	589	-499	593	-501	4.472	602	-498	600	-500	2.828
12	586	-502	592	-508	8.486	596	-480	597	-483	3.162
13	605	-509	608	-505	5.000	604	-518	601	-515	4.242
14	601	-505	602	-503	2.236	598	-483	600	-481	2.828
15	600	-487	601	-480	7.071	601	-495	610	-496	9.055
Average [mm]					7.124	Average [mm]				
Standard Deviation [mm]					2.975	Standard Deviation [mm]				

Based on the results in Table 1, the performance of the identification algorithm can be statistically analyzed. The average measurement error (mean  $\pm$  standard deviation) of the window location is  $7.124 \pm 2.975$  mm and  $6.612 \pm 3.390$  mm for the virtual environment and the real environment, respectively. Given the dimensions of the window corner and required field of view (FOV) shown in Figure 18, the area of the whole window corner was 110 mm by 110 mm, and the FOV of the set height was around 232 mm by 309 mm. The FOV at the set height was at least 86 mm wider than the required FOV to fully cover the corner for both sides. Therefore, a 6 to 7 mm measurement error represents around a 5% error relative to the window corner area and is thus considered acceptable, considering the purpose of the algorithm (approximation step).

### 5.2.2. Module 2

The main function of the second module was to detect the weld seam with the pre-trained image segmentation model. The Mask R-CNN applied dataset, which consists of 78 new images from the virtual and real environments, was used for validation of the model. Examples of the weld seam classification, location, and segmentation results are presented in Figure 19.



**Figure 19.** Results of the weld seam segmentation using the Mask R-CNN model on images of the test dataset. (**Top**) Images of the simulated environment. (**Bottom**) Images of the real environment.

To validate the Mask R-CNN detection results, image segmentation performance metrics were used. These metrics quantify the capacity of the model to predict the boundaries of the weld seam when compared to manually identified ones. This is commonly achieved using the intersection over union (IoU) metric. (IoU) is a coefficient that compares the area of overlap between the segmented result and the manually obtained boundaries over the area of union of both, as shown in Equation (8).

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (8)$$

where ( $A$ ) is the detected weld seam area and ( $B$ ) is the manually defined weld seam area. (IoU) is often utilized as a given threshold to classify the prediction as a true positive or as a false positive. (IoU) defines the threshold in a range from 0 (no union between the ground truth and detection) to 1 (detection and ground truth are identical). As such, different (IoU) thresholds are applied to describe the accuracy of the model in correctly identifying the boundaries of the weld seam. As the threshold value increases, the determination of what a true positive detection is gets stricter. Additionally, the mean average precision (mAP) is commonly used to evaluate object detection models. As there is only one type of object or class to predict, (mAP) represents the average precision, which is the area under the precision–recall curve (see Equation (9)).

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i} \times \frac{TP_i}{TP_i + FN_i} \quad (9)$$

where  $(TP)$ ,  $(FP)$ , and  $(FN)$  represent the number of true positive, false positive, and false negative detections as determined by the (IoU) threshold for all  $(n)$  images. Table 2 shows the proposed seam-detection model's performance with variable (IoU) values ranging from 0.5 to 0.9 for the applied aforementioned dataset.

**Table 2.** Seam-detection performance results using the Mask R-CNN model to detect window weld seams with an (IoU) between 0.5 and 0.95.

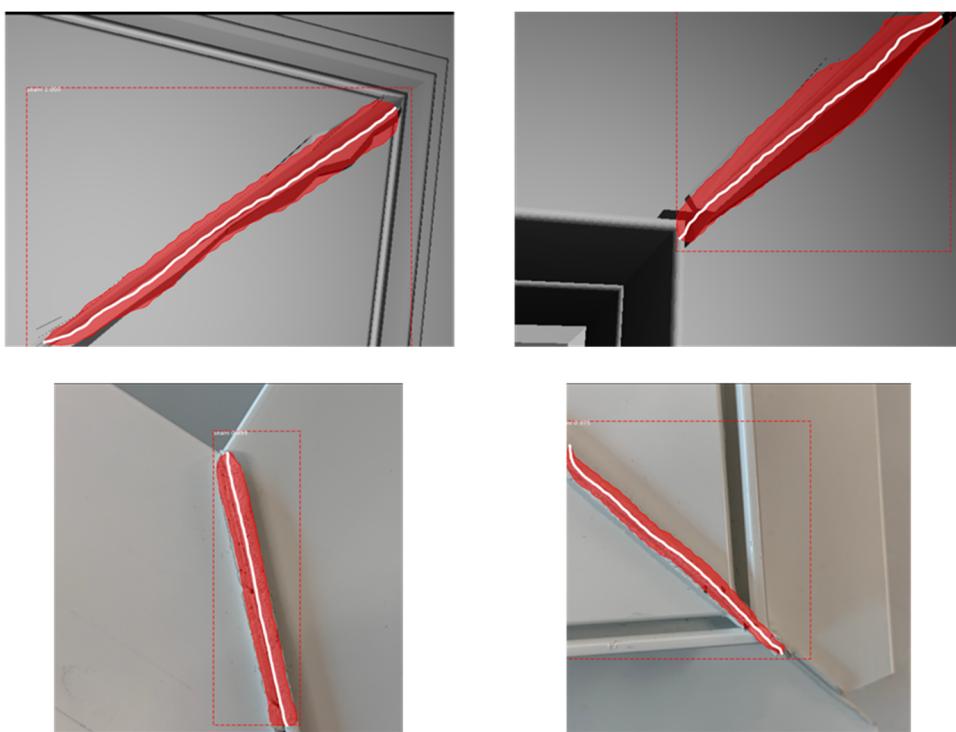
IoU	0.5	0.6	0.7	0.75	0.8	0.85	0.9	0.95
mAP	0.953	0.851	0.784	0.755	0.707	0.664	0.611	0.512

For this study, the (IoU) was tested starting at 0.5 (commonly used as the minimum acceptable value) and continuously increasing it until it reached 0.95. Based on the results, as expected, the model performance decreased gradually as the (IoU) value increased, with a mAP of 0.953 when the (IoU) was set to its lower value (0.5) and reaching a minimum of 0.512 when the (IoU) was set to 0.95. The model performance obtained from the weld seam model in these experiments is comparable with similar applications found in the literature [35]; thus, the proposed model is considered good enough to provide an accurate enough mask that enables the path generation process in the next module.

### 5.2.3. Module 3

The outcome of module 3 is a list of coordinates or target points in the three-dimensional space that represent the cleaning path for the robot arm. As the calculation of the cleaning path from the weld seam's detected contour is deterministic, any error in the path is directly related to the accuracy obtained from the image segmentation process discussed above. Aside from the accuracy of path location, the evaluation of this module also relates to the force control and the accuracy capacity of the robotic arm executing the task. The force control process is similar to the position control: each joint must be adjusted to match the required output. However, the process is more complicated than the position control due to both requirements to be met simultaneously: having a specific trajectory in a certain direction for the end effector and the application of the exact force needed.

In this study, several assumptions are taken that support the validation of this module. For the position control, it is assumed that the generated path is within the reach of the robot and that the motions are feasible. This means that it is considered that all potential paths for cleaning that can be obtained using the proposed system are feasible from the perspective of the robot controller. For the force control, it is assumed that the tooling used for cleaning the window frames is unchanged. In other words, the current practice allows for the removal of the excess material by applying a certain force; therefore, if the robot applies the same force, the same material removal should be observed. In summary, the outcome of this module does not require validation based on the assumptions taken as long as the appropriate robot arm is selected (enough reach and enough force). Both requirements depend on the industrial application and are discussed in the following subsection. Thus, the results here contain only visualization results, which can be found in Figure 20.



**Figure 20.** Illustration of the generated cleaning paths (white lines) on a detected weld seam. **(Top)** Images of the simulated environment. **(Bottom)** Images of the real environment.

### 5.3. Discussion and Limitations

The proposed vision-based robotic corner cleaning system is capable of identifying and locating window frames, detecting weld seams, and eventually generating the cleaning path to guide the robotic arm. However, there are two main limitations of this research related to some results presented in the validation results section. The related topics worth exploring further in the future will be discussed, as well.

First, the choice of end effector (cleaning tool) can lead to huge differences for the path planning. The types of industrial blades used in window corner cleaning include circular saws, straight knives, and deburring blades. The end-effector of choice in this study was a deburring blade, which is the most utilized for edge deburring, chamfering, countersinking, and scraping of material in window manufacturing. The current industrial practice is to use blades, which can clean weld seams in a single unidirectional motion; in other words, the current blades are wider than the usual seam width. However, with various designs and tools available, the path planning brought about in this research cannot be considered a universal solution for all kinds of end effectors. As a result, a solution that links the end effector in use to automatically adapt the cleaning path planning could be worth discussing in future research. In addition, the window profiles used are the most common and basic profiles available, which do not contain difficult geometrical shapes alongside its transversal section; however, more ornamental (and more expensive) window profiles exist that contain different features that are not seen in the studied profiles, e.g., gaps, cavities, or chambers. With the current approach, which does not estimate the weld seam volume, such features would be impossible to clean. A different approach would need to be taken to support such cleaning operations, which must be researched in the future.

Secondly, although the results of the neural network model show adequate performance, they also bring concerns of overfitting and a lack of generalization in the dataset. Due to the limited variability in the window profiles tested, the image segmentation model is trained on a dataset that consists of image data with a certain homogeneity, as they come from a unique production line and the same industrial environment. To obtain a more general model with different window types and materials, additional efforts need

to be made to gather data, retrain the model, and determine the viability of the proposed approach. On top of that, the cleaning procedure for different materials varies; hence, the proposed system needs to be flexible enough to handle an increased number of outputs while maintaining the achieved accuracy. Further validation of the selected model when including diverse sample sources will yield a more general application for the window manufacturing industry.

The decision to choose robotic arms as the executor in this study was based on their ability of executing subtractive tasks precisely. However, aside from the two limitations mentioned above, force control is another topic that has not been explored in this study. In practice, the force control and position control, which were dealt with in this research, are calculated and planned as a hybrid force and position control to achieve precise execution. Such a topic is related to crucial factors when it comes to executing the cleaning task in practice, including friction and force/torque sensors. Since most of the weld seams lie horizontally, the force applied to remove the residual includes a horizontal force, as well. The friction between the window frame and the working table, as well as the horizontal force applied by the robot need to be taken into consideration when designing the desired force applied by the end-effector on a window frame. If a situation occurs in which the robot's forces cause a displacement of the window frame, the cleaning process will have an undesired execution. Although the workstation area in the study is assumed to be an ideal situation, some additional effort in the redesign of the work area may be needed to decrease the possibility of displacement, like clamps or magnet holders. Nowadays, most industrial robot arms are equipped with sensors to control force and torque because the force needs to be controlled differently to match target objects. Considering the diverse position and orientation of the window frames in this study, the optimal sensor controlling force and torque is another key problem to deal with. For example, calculation of the proper force that can clean the weld seam but not cause deformation or damage to the frames is required. With precise path planning, determining the force that performs perfectly and does not bring external variables to the working environment is a problem that needs to be addressed in the future to realize the automation of window corner cleaning.

## 6. Conclusions

Hot plate welding is a commonly used welding technique and plays an important role in thermoplastic window frame manufacturing nowadays. Weld seams are unwanted results that go along with the welding process that require an additional effort to obtain an aesthetically pleasant and marketable window frame. When considering the quality of weld seam cleaning, the current methods lack the ability to adapt to dynamic situations that risk product quality, leading to difficult predictability of the cleaning process and a complete reliability on the manual operator's capacity to perform any required adjustments or rework.

This study proposes a vision-based robotic corner cleaning system to enhance the adaptability of the current corner cleaning methods with vision systems and to execute a precise cleaning task execution with robotic arms. The proposed method relies on novel image processing tools to locate, identify, and quantify the area of the weld seam that requires cleaning before estimating an appropriate cleaning path for the robot. The location of the corner area and weld seam is determined by the edges of the frame and identified using the Hough transform, enabling the robot arm to perform a guided approximation that does not rely on predefined window positioning or orientation. Then, a Mask R-CNN model is trained to locate and segment the weld seam. The obtained mask is finally transformed using the greedy algorithm to automatically generate the required robot path to clean the weld seam.

Two scenarios were utilized to validate the proposed system: a virtual simulation and a real scenario using a UR5e robot with a built-in wrist camera and an overlooking Basler ACE camera on the workstation. The proposed approach resulted in less than 1 cm error for the window location and robot approximation and achieved a maximum of 0.95 mean

average precision on the weld seam detection and segmentation. For future research, with the limitations of different types of end effectors and generalization concerns, the proposed solution will be adapted by enlarging the sample database and adding flexibility to the path planning process. In addition, we will consider the introduction of more advanced computer vision algorithms to further enhance the system's perception of the process, as well as decision-making capabilities from other sensor sources (e.g., end-effector pressure sensors).

**Author Contributions:** Conceptualization, T.-J.T. and P.M.; methodology, T.-J.T. and P.M.; software, T.-J.T.; validation, T.-J.T. and P.M.; formal analysis, T.-J.T.; investigation, T.-J.T.; resources, M.A.-H.; data curation, T.-J.T.; writing—original draft preparation, T.-J.T.; writing—review and editing, P.M. and M.A.-H.; visualization, T.-J.T.; supervision, P.M. and M.A.-H.; project administration, M.A.-H.; funding acquisition, M.A.-H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (File No. IRCPJ 419145-15).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to manufacturing confidential information on the window frame designs by the industrial partner.

**Acknowledgments:** The authors acknowledge the support of the researchers at the Modular Construction Group at the University of Alberta.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Mittal, S.; Khan, M.A.; Romero, D.; Wuest, T. Smart manufacturing: Characteristics, technologies and enabling factors. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2019**, *233*, 1342–1361. [[CrossRef](#)]
- Martinez, P.; Al-Hussein, M.; Ahmad, R. Online vision-based inspection system for thermoplastic hot plate welding in window frame manufacturing. *Procedia CIRP* **2020**, *93*, 1316–1321. [[CrossRef](#)]
- Narula, S.; Prakash, S.; Dwivedy, M.; Talwar, V.; Tiwari, S.P. Industry 4.0 adoption key factors: An empirical study on manufacturing industry. *J. Adv. Manag. Res.* **2020**, *17*, 697–725. [[CrossRef](#)]
- Roslan, Q.; Ibrahim, S.H.; Affandi, R.; Nawi, M.N.M.; Baharun, A. A literature review on the improvement strategies of passive design for the roofing system of the modern house in a hot and humid climate region. *Front. Archit. Res.* **2016**, *5*, 126–133. [[CrossRef](#)]
- Aldawoud, A. Windows design for maximum cross-ventilation in buildings. *Adv. Build. Energy Res.* **2017**, *11*, 67–86. [[CrossRef](#)]
- Takada, K.; Hayama, H.; Mori, T.; Kikuta, K. Thermal insulated PVC windows for residential buildings: Feasibility of insulation performance improvement by various elemental technologies. *J. Asian Archit. Build. Eng.* **2021**, *20*, 340–355. [[CrossRef](#)]
- Martinez, P.; Ahmad, R. Quantifying the Impact of Inspection Processes on Production Lines through Stochastic Discrete-Event Simulation Modeling. *Modelling* **2021**, *2*, 406–424. [[CrossRef](#)]
- Kermani, A.P.; Abbaslu, L.; Alizadeh Shahi, S.R. A Model to Plan the Operations and Control the Orders (The Case Study: Profile Door and Window Manufacturing (UPVC)). *Biomed. Pharmacol. J.* **2015**, *8*, 219–229. [[CrossRef](#)]
- Machado, A.A.; Zayatz, J.C.; da Silva, M.M.; Neto, G.M.; Leal, G.C.L.; Lima, R.H.P. Aluminum bar cutting optimization for door and window manufacturing. *DYNA* **2020**, *87*, 155–162. [[CrossRef](#)]
- Itani, A.; Alghamdy, M.; Nazir, H.; Sharma, S.; Ahmad, R. A decision-making tool to integrate lean 4.0 in windows manufacturing using simulation and optimization models. In Proceedings of the 32nd European Modeling & Simulation Symposium (EMSS 2020), online, 16–18 September 2020; pp. 137–145. [[CrossRef](#)]
- Pérez, L.; Rodríguez, I.; Rodríguez, N.; Usamentiaga, R.; García, D. Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review. *Sensors* **2016**, *16*, 335. [[CrossRef](#)]
- Shah, H.N.M.; Sulaiman, M.; Shukor, A.Z.; Rashid, M.Z.A. Recognition of Butt Welding Joints using Background Subtraction Seam Path Approach for Welding Robot. *Int. J. Mech. Mechatron. Eng.* **2017**, *17*, 57–62.
- Abdelaal, M.; Farag, R.M.A.; Saad, M.S.; Bahgat, A.; Emara, H.M.; El-Dessouki, A. Uncalibrated stereo vision with deep learning for 6-DOF pose estimation for a robot arm system. *Robot. Auton. Syst.* **2021**, *145*, 103847. [[CrossRef](#)]
- Kleppe, A.; Bjørkedal, A.; Larsen, K.; Egeland, O. Automated Assembly Using 3D and 2D Cameras. *Robotics* **2017**, *6*, 14. [[CrossRef](#)]
- Levine, S.; Pastor, P.; Krizhevsky, A.; Ibarz, J.; Quillen, D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Int. J. Robot. Res.* **2018**, *37*, 421–436. [[CrossRef](#)]
- Lei, T.; Rong, Y.; Wang, H.; Huang, Y.; Li, M. A review of vision-aided robotic welding. *Comput. Ind.* **2020**, *123*, 103326. [[CrossRef](#)]

17. Chen, S.B.; Zhang, Y.; Qiu, T.; Lin, T. Robotic Welding Systems with Vision-Sensing and Self-learning Neuron Control of Arc Welding Dynamic Process. *J. Intell. Robot. Syst.* **2003**, *36*, 191–208. [[CrossRef](#)]
18. Micallef, K.; Fang, G.; Dinhamb, M. Automatic Seam Detection and Path Planning in Robotic Welding. In *Robotic Welding, Intelligence and Automation*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 23–32. [[CrossRef](#)]
19. Dinhamb, M.; Fang, G. Autonomous weld seam identification and localisation using eye-in-hand stereo vision for robotic arc welding. *Robot. Comput. Integrat. Manuf.* **2013**, *29*, 288–301. [[CrossRef](#)]
20. Duan, Y. Welding Seam Recognition Robots Based on Edge Computing. In Proceedings of the 2020 International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 1–2 August 2020; pp. 27–30. [[CrossRef](#)]
21. Tian, Y.; Liu, H.; Li, L.; Yuan, G.; Feng, J.; Chen, Y.; Wang, W. Automatic Identification of Multi-Type Weld Seam Based on Vision Sensor with Silhouette-Mapping. *IEEE Sens. J.* **2021**, *21*, 5402–5412. [[CrossRef](#)]
22. Rout, A.; Deepak, B.B.V.L.; Biswal, B.B. Advances in weld seam tracking techniques for robotic welding: A review. *Robot. Comput. Integrat. Manuf.* **2019**, *56*, 12–37. [[CrossRef](#)]
23. Xu, J.; Zhang, G.; Hou, Z.; Wang, J.; Liang, J.; Bao, X.; Yang, W.; Wang, W. Advances in Multi-robotic Welding Techniques: A Review. *Int. J. Mech. Eng. Robot. Res.* **2020**, *9*, 421–428. [[CrossRef](#)]
24. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [[CrossRef](#)] [[PubMed](#)]
25. Huang, H.; Wei, Z.; Yao, L. A Novel Approach to Component Assembly Inspection Based on Mask R-CNN and Support Vector Machines. *Information* **2019**, *10*, 282. [[CrossRef](#)]
26. Qianqian, Z.; Sen, L.; Weiming, G. Research on Vehicle Appearance Component Recognition Based on Mask R-CNN. *J. Phys. Conf. Ser.* **2019**, *1335*, 012026. [[CrossRef](#)]
27. Xia, C.; Pan, Z.; Zhang, S.; Polden, J.; Li, H.; Xu, Y.; Chen, S. Mask R-CNN-Based Welding Image Object Detection and Dynamic Modelling for WAAM. In *Transactions on Intelligent Welding Manufacturing*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 57–73. [[CrossRef](#)]
28. He, F.; Sun, X.; Wang, Y.; Rong, S.; Hu, Y. Research on Weld Recognition Method Based on Mask R-CNN. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2021; pp. 545–551. [[CrossRef](#)]
29. Jin, X.; Lv, L.; Chen, C.; Yang, F.; Chen, T. A New Welding Seam Recognition Methodology Based on Deep Learning Model MRCNN. In Proceedings of the 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), Guangzhou, China, 13–15 November 2020; pp. 767–771. [[CrossRef](#)]
30. Yang, Y.; Li, Z.; Yu, X.; Li, Z.; Gao, H. A trajectory planning method for robot scanning system using mask R-CNN for scanning objects with unknown model. *Neurocomputing* **2020**, *404*, 329–339. [[CrossRef](#)]
31. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q.* **2004**, *28*, 75. [[CrossRef](#)]
32. Illingworth, J.; Kittler, J. A survey of the hough transform. *Comput. Vis. Graph. Image Process.* **1998**, *44*, 87–116. [[CrossRef](#)]
33. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
34. Dutta, A.; Zisserman, A. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2276–2279. [[CrossRef](#)]
35. Wang, S.; Sun, G.; Zheng, B.; Du, Y. A Crop Image Segmentation and Extraction Algorithm Based on Mask RCNN. *Entropy* **2021**, *23*, 1160. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.