

Article

Vision AI System Development for Improved Productivity in Challenging Industrial Environments: A Sustainable and Efficient Approach

Changmo Yang ^{1,2} , JinSeok Kim ², DongWeon Kang ² and Doo-Seop Eom ^{1,*} 

¹ Department of Electrical and Computer Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea; cmyang83@gmail.com

² Hyundai Motor Company, 37 Cheoldobangmulgwan-ro, Uiwang 16088, Republic of Korea; cmyang@hyundai.com (C.Y.)

* Correspondence: eomds@korea.ac.kr

Abstract: This study presents a development plan for a vision AI system to enhance productivity in industrial environments, where environmental control is challenging, by using AI technology. An image pre-processing algorithm was developed using a mobile robot that can operate in complex environments alongside workers to obtain high-quality learning and inspection images. Additionally, the proposed architecture for sustainable AI system development included cropping the inspection part images to minimize the technology development time, investment costs, and the reuse of images. The algorithm was retrained using mixed learning data to maintain and improve its performance in industrial fields. This AI system development architecture effectively addresses the challenges faced in applying AI technology at industrial sites and was demonstrated through experimentation and application.

Keywords: vision AI; industrial sites; AI technology; image pre-processing; mobile robot; sustainable AI system



Citation: Yang, C.; Kim, J.; Kang, D.; Eom, D.-S. Vision AI System Development for Improved Productivity in Challenging Industrial Environments: A Sustainable and Efficient Approach. *Appl. Sci.* **2024**, *14*, 2750. <https://doi.org/10.3390/app14072750>

Academic Editors: Patrick Wong and Yifan Zhao

Received: 26 February 2024

Revised: 18 March 2024

Accepted: 22 March 2024

Published: 25 March 2024



Copyright: © 2024 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

In recent years, extensive research has been conducted on the development of machine vision technology for product quality and work-ability improvement in industrial fields [1,2]. The rapid development of deep learning technology has enabled us to perform quality inspections on products with a diversity and complexity that was difficult to achieve in the past. However, the implementation of deep learning technology enables quality inspection primarily on small unit parts that can obtain relatively high-quality inspection images in an extremely limited environment. However, it is challenging to implement an inspection system using deep learning technology in an industrial environment where it is difficult to obtain high-quality images due to the structural diversity and frequent environmental changes, such as the assembly process of automobile manufacturing plants. Over 70% of automobile manufacturing plants use automated systems for assembly. However, parts such as wiring and connectors, which are difficult to automate due to their flexibility and versatility, still depend on manual assembly. Due to the structural characteristics of automobiles, during the assembly of parts, the next assembly part covers the previous assembly part, making it impossible to detect assembly defects through visual inspection using the human eye in the next process. If defects and omissions are detected through electrical inspection after the vehicle is assembled, the parts must then be disassembled in the reverse order to rectify these assembly defects, which involves considerable time and cost. Additionally, defective assembly of fixing clips, bolts, nuts, and so on cannot be detected electronically due to field claims such as vibration and noise produced while driving a vehicle. Recently, there has been a shift in the assembly process from the conveyor method to a cellular production method that is efficient for the small-volume production of

multiple vehicles and options. The cellular method enables the assembly of up to five times more parts than conventional conveyors within a single process. This has increased the probability of parts assembly omission as well as erroneous assembly by workers, and has also increased the cost of poor assembly quality. Therefore, it is crucial to develop a new visual inspection system technology for real-time image acquisition and assembly defect detection after manual assembly by workers in an environment where it is difficult to obtain high-quality images in real-time. Two factors must be considered for real-time visual inspection in the manual assembly processes. Firstly, images must be obtained in real-time by collaborating with workers. Secondly, a visual inspection algorithm must be developed to detect assembly defects using the obtained images.

1.1. Image Acquisition Device

It is difficult to control lighting and environmental changes during manual assembly processes in the automobile industry, unlike small parts inspections. Furthermore, the mixed production of multiple vehicle models presents difficulties in image acquisition due to frequent changes in the inspection items. Therefore, highly mobile robots and manipulators with excellent flexibility are necessary for image acquisition in automobile manual assembly processes. Additionally, a device with high safety standards, obstacle avoidance, and excellent mobility that can be operated by workers must be developed. Various commercial devices have been implemented for image acquisition, but it is difficult to find a suitable device for manual assembly processes in the automobile industry. Moreover, it is challenging to acquire images inside the vehicle with a fixed camera. Although a 360-degree camera can be installed inside the vehicle before assembly to capture images, it is not suitable for inspection purposes due to low image resolution. Wearable glasses produce low-quality images due to shaking, and drones are not applicable due to safety and noise issues. A device must be developed that can collaborate flexibly with workers in automobile manual assembly processes and acquire high-quality image data through stability, obstacle avoidance, and excellent mobility.

1.2. Vision Inspection Algorithm Development

When a new car is introduced into a car production plant, the car body, color, and parts are changed. For approximately 100 days after the production of the vehicle, there are several defective parts assemblies due to the inexperience of the workers. After 100 days, the skill level of the operator improves, and the number of assembly defects is drastically reduced. Therefore, an assembly defect inspection system must be established at the initial stage of the production of new cars. However, the effect of implementing the existing rule-based visual inspection system is insufficient since an engineer requires at least six months to develop the visual inspection algorithm corresponding to the parts to be assembled in a new car. Several attempts are being made to apply deep learning technology to reduce the development period of the rule-based algorithm [3]. However, deep learning vision technology also requires high-quality learning data to develop new car assembly inspection algorithms. It takes more than 100 days to acquire the normal and defective data required for deep learning algorithm learning, due to which the visual inspection of assembly parts in the early stages of new car production is impossible, similar to the existing rule-based vision system. In this study, we propose a solution for the manual process, such as car assembly in the production plant, where it is difficult to implement the existing visual inspection system due to environmental changes and inspection item modifications. Our proposed solution includes a mobile robot suitable for image acquisition and a method for developing AI algorithms that can reduce the development period. We also demonstrate the performance of our proposed solution.

2. Related Works

Deep learning technology has been reported to outperform conventional rule-based visual inspection systems in detecting various types of assembly defects and conducting

quality inspections in complex industrial settings [4–6]. However, high-quality training images are required to improve the performance of deep learning-based inspection systems. Most studies conducted on deep learning technology have been implemented on small-scale inspection targets with limited variation under controlled lighting and environmental conditions. In several industrial settings, it is difficult to obtain high-quality training images due to uncontrolled lighting and environmental conditions, along with frequent changes in the inspection targets, making it difficult to acquire sufficient training data. Previous studies have attempted to implement image acquisition and deep learning inspection under conditions similar to those of actual industrial sites. In Wang's study, assembly workers directly used wearable lenses and headsets to acquire images. The time required to capture the part images was set by using the worker's position and gaze information [7]. The evaluation results demonstrated that the system accuracy was low, at 85%. This was because even in a laboratory environment where the lighting conditions were consistent, there were image variations based on the distance and angle at which the worker captured the image. In Mazzetto's study, deep learning technology was implemented to inspect the surface treatment quality of automobile assembly parts. After sufficient training data were obtained, deep learning technology showed superior inspection performance when compared to the existing rule-based visual inspection method [3]. However, the inspection algorithm was only limited to the surface inspection of brake pedal parts that did not change when a new vehicle was released, and it was developed only after obtaining sufficient training data. Research is being conducted to address the challenge of obtaining sufficient data in industrial settings. The data acquired in actual industrial settings are small in quantity, but there is a significant difference in the OK and NG ratios. In the case of manual assembly processes for automobiles, data acquisition can be performed with an OK rate of 99% and an NG rate of 1%. Therefore, NG items must be created arbitrarily to acquire sufficient training data within a short period. This involves substantial time and cost. The one-class neural-network method, a type of semi-supervised learning, has been proposed to address this issue. This method learns using only a small number of normal images and detects samples that differ from the normal samples as outliers [8]. However, uncontrollable environments such as lighting can cause severe image variations due to image exposure. In the case of automobile parts comprising flexible cables and connectors, the position of the cable varies even in normal images, along with the position of the surrounding parts. Therefore, it is difficult to determine the boundary between bad and normal images. Even if an appropriate boundary is set, it frequently changes due to the rapidly changing environmental conditions, resulting in decreased inspection accuracy and increased maintenance costs. Therefore, a system must be developed that can acquire and inspect images under conditions where it is difficult to secure learning data of sufficient quality due to frequent changes in the inspection parts at industrial sites. Additionally, the high time and cost requirements for repeatedly maintaining detection algorithms due to frequent changes in the inspection items make it difficult to implement deep learning technology. In actual industrial sites, developing and implementing a deep learning inspection system can be challenging owing to the high cost required for the development and maintenance of the inspection system. In this study, we present a development methodology for a deep learning visual inspection system that can be implemented in an actual industrial field. First, image acquisition and pre-processing techniques using mobile robots acquire high-quality image data required to train the deep learning algorithms. Second, we propose a method to improve the performance of deep learning algorithms by using a small amount of data that can be acquired within a short period of time in actual industrial sites. Third, we propose a deep learning algorithm for the re-learning method to respond to frequent changes in the inspection parts and environmental conditions in industrial sites and to maintain the detection performance. Finally, the proposed technology is demonstrated through empirical evaluation.

3. Proposal Method

In this study, we propose a methodology to develop visual AI technology that enables the effective inspection of assembly defects in automobile production from the early stages, using mobile robot technology for image acquisition. We focus on the development process and system architecture to create a visual AI inspection system capable of accurately detecting incorrect parts assembly by workers in the manual assembly process of an automobile manufacturing plant, as illustrated in Figure 1.

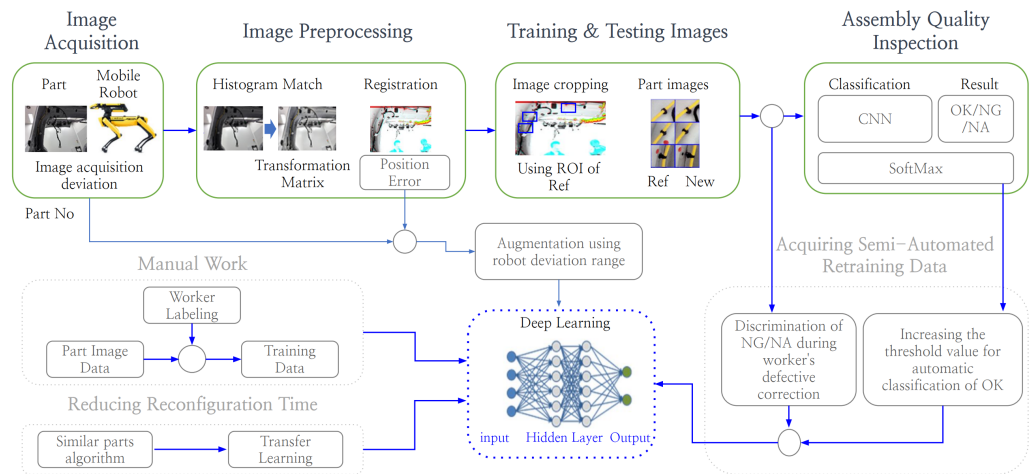


Figure 1. Flowchart of the proposed AI system for industrial site inspection.

We propose a new system to overcome the difficulties faced in implementing visual AI inspection in industrial settings, as depicted in Figure 2. Based on a standard visual AI inspection framework, this enhanced system additionally compensates for the variation in the images acquired by mobile robots due to changes in the industrial environment (①), and utilizes this information to crop specific component images (②) to generate optimized images that are crucial for assessing the assembly quality. Since automobile production facilities use more than 100 different components, each assembled in unique ways, the efficiency of visual AI inspections must be enhanced by individually applying tailored algorithms for each component (③), rather than employing a single generic algorithm. This methodology enables the straightforward reuse of algorithms for similar parts upon the introduction of new vehicle models or the application of simple transfer learning, thereby enhancing the operational management efficiency. From an operational perspective in industrial settings, we propose introducing an 'inspection error (NA)' category in deep learning image classification (④) beyond the conventional OK/NG criteria to mitigate productivity loss due to pseudo-defects and prevent the leakage of assembly defects. This raises the benchmark and requires operator verification when the criteria are not met, enabling the system to be operational even before sufficient training data have been accumulated. The inspection error images selected by the operators (⑤) can then be used as evaluation data for algorithm retraining, which streamlines the algorithm assessment and improvement process. A separate system must be established to maintain and manage the algorithm performance in case environmental changes in the industrial site cause variations in the mobile robot's positioning, potentially degrading the performance of pre-set detection algorithms (⑥–⑫), as suggested in Figure 1. This involves calculating the variance in the images during the cropping process, extracting a T-Matrix (⑦) through feature detection and matching (⑥), and storing the T-Matrix for each robot position (⑧) to calculate the range of variance. By employing this approach in deep learning algorithms, it enhances the algorithm performance through image augmentation techniques (⑨), utilizing the T-Matrix to define the range of variance within which the robot can acquire images, thus improving the detection algorithm performance. This strategy maintains and enhances the algorithm performance (⑩), enables the development of new algorithms through retrain-

ing (11), and uses evaluation data created by operator selection processes to compare and evaluate the performance of old and new algorithms, facilitating algorithm replacement if necessary (12). This approach ensures the continuous improvement in and maintenance of the algorithm performance.

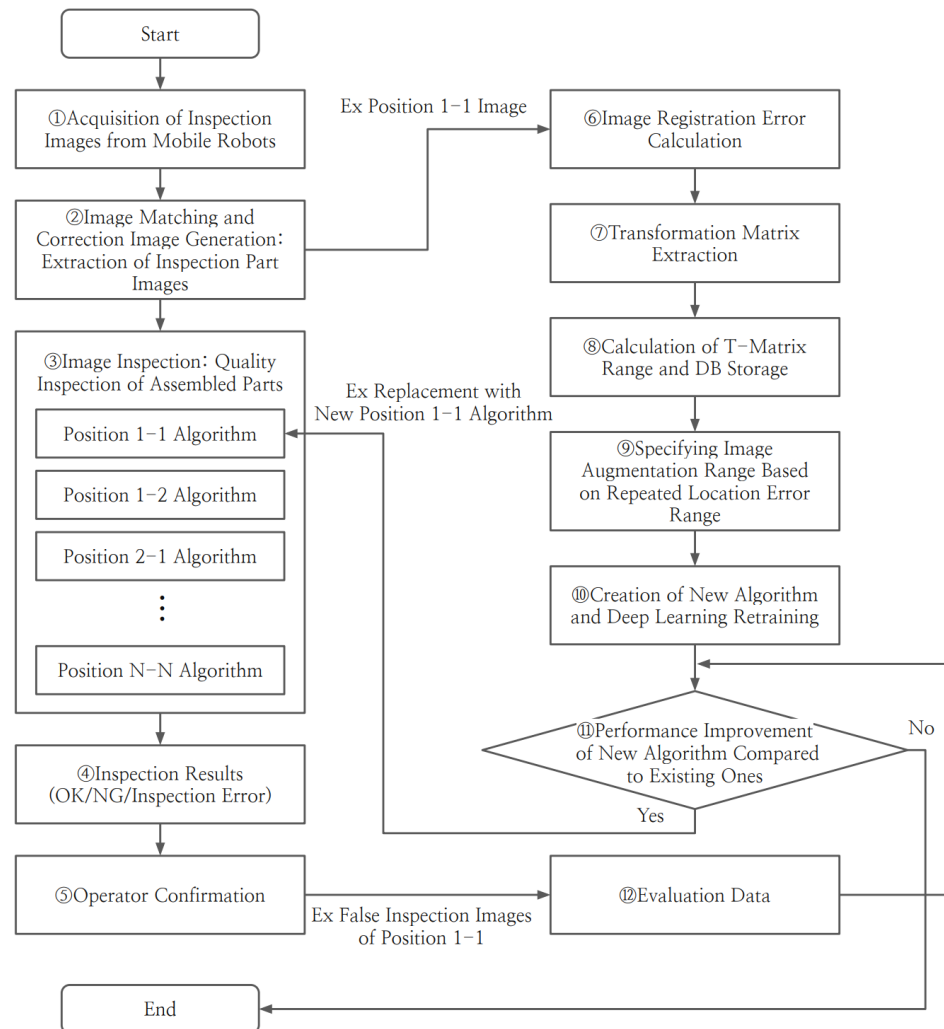


Figure 2. Integrated vision AI inspection system flowchart for quality testing in industrial environments.

4. Detailed Proposed Technology and Test Results

4.1. Acquisition of High-Quality Deep Learning and Inspection Data

In environments such as the manual assembly process in car production factories, which are narrow and complex, and where obstacles like bolts and nuts exist, a visual acquisition device must be developed that can acquire images in real-time and in the same space as the operator. In this study, we utilized a Boston Dynamics' (Waltham, MA, USA) four-legged robot, called SPOT, which has a relatively small body and the ability to move through narrow spaces with its four-legged walking system, as well as the self-SLAM technology that enables it to avoid obstacles. Additionally, the 4K camera attached to the SPOT package's seven-axis robot arm enables the easy acquisition of the part images [9]. However, due to the characteristics of the four-legged walking system, the repeated positioning accuracy exhibits a deviation of more than ± 200 mm from the body base; there is also deviation when acquiring images using the camera attached to the arm. This is a common problem with all mobile robots used in industrial sites, and it can degrade the quality of the data, thereby affecting the performance of AI inspection. Unlike stationary robots that employ positional constraints, mobile robots utilizing methods such as visual SLAM can experience location errors ranging from 10 cm to 1 m. This variation in

the positioning accuracy further complicates data acquisition and can significantly impact the effectiveness of AI-based inspections. Therefore, a solution must be developed to address this issue.

4.1.1. Landmark (Fiducial Mark) Centering Technique for Improving the Repeat Positioning Accuracy of SPOT

The SPOT robot uses five ToF cameras on its body for visual SLAM-based position movement. However, the reference vehicle moves using an AGV or conveyor for component imaging, causing position dispersion. Furthermore, the characteristics of quadrupedal walking result in poor repeat positioning accuracy with errors of over ± 200 mm, as shown in Figure 3a. To address this problem, short-range communication devices, such as UWB technology, have been used to improve the repeat positioning accuracy in the industrial field [10,11]. However, this method incurs additional costs for installing infrastructure such as UWB transceivers in the surrounding environment, as well as AGV or vehicle attachment and removal of UWB. The location accuracy may also be reduced in environments such as car factory structures that can cause wireless signal fading. In this study, a landmark (Fiducial Mark) was attached to the AGV to consider the characteristics of the industrial field and minimize investment costs, which serves as the reference for position movement. A vertical reference point was specified between the SPOT body and the F-Mark to align the body accurately. This method does not require additional modification or cost even when the process changes or when new vehicles are introduced, as only the F-Mark must be attached without requiring additional infrastructure installation. Using this proposed method, the SPOT robot utilizes its front-facing camera to recognize the size of the attached F-Mark and measures its size and angle upon reaching the inspection site, as depicted in Figure 3b. By centering the SPOT body to be perpendicular to the F-Mark and adjusting the pre-set distance values, the positional error was reduced from ± 200 mm to as low as ± 14 mm. However, even if the error of the SPOT body is small, the positional error of the camera at the end of the arm that acquires the image accumulates based on the arm pose, causing a large deviation in the acquired image.

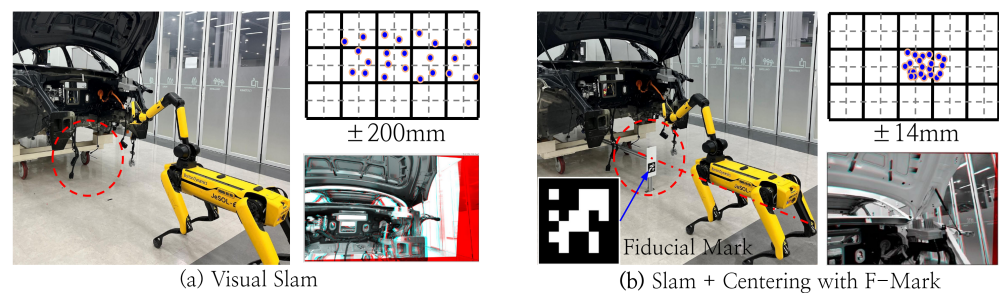


Figure 3. Improvement in repeat positioning accuracy with the centering technique using the Fiducial Mark.

4.1.2. Automatic Correction Algorithm for Image Matching Deviation Caused by Positional Precision Error

Additional hardware improvements to increase the positional accuracy would require excessive cost and time to reduce the deviation in the images acquired from the mobile robot or arm. In the industrial field, there is a trade-off between performance improvement and cost; therefore, an appropriate performance improvement method must be developed. In this study, we propose a visual software algorithm that can correct image deviation with relatively low investment cost. We used the speeded-up robust feature (SURF) algorithm to detect feature points between the first image and the repeatedly acquired images, and corrected the deviation using affine, projective, and other transformation techniques [12,13]. However, during feature point detection, there were problems with recognizing the background as a feature point instead of the inspection area, or recognizing parts incorrectly installed as the same feature point, resulting in degraded performance, as shown in Figure 4

(top). To address this issue, we limited the feature point search area to the vehicle body, as shown in Figure 4 (bottom). This prevents the recognition of the background as a feature point outside the vehicle body. Additionally, the image-matching performance is improved by masking the part area and excluding it from the search area to prevent the recognition of incorrect feature points when parts are installed incorrectly.

$$\operatorname{argmax}(x, y) \in R \cap M \sum i, j H_{ij} L_{\sigma}(x + i, y + j) \quad (1)$$

$$\operatorname{argmax}(x, y) \in M \sum i, j H_{ij} L_{\sigma}(x + i, y + j) \quad (2)$$

$$\operatorname{argmax}(x, y) \in I \sum i, j H_{ij} L_{\sigma}(x + i, y + j) \quad (3)$$

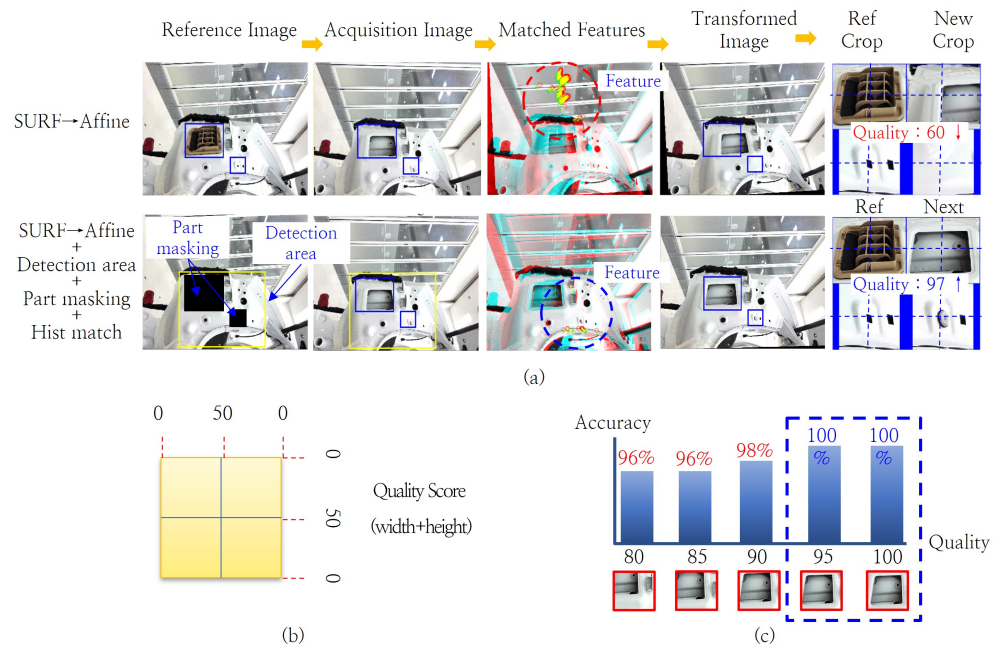


Figure 4. (a) Image registration and cropping using the proposed algorithm, (b) quality scoring of cropped images, (c) changes in AI accuracy according to cropped image quality.

Equation (1) is used to detect features within the intersection of the search area specified by the reference image's feature search area, R , and the image mask, M , that includes the areas outside the feature search area. H_{ij} represents the Gaussian kernel used in the Harris corner detector, and $L_{\sigma}(x + i, y + j)$ represents the result of differentiation and smoothing using the image's Rob operator with a Gaussian filter. $(x + i, y + j)$ represents the position of the kernel. Conversely, there are two ways to perform a feature search on a newly acquired image. If the image distortion is small, Equation (2) can be used, which utilizes the image mask, M , that includes the areas outside the feature search area. In this case, the area for feature detection is reduced since the same image masking area is included, resulting in an increase in the speed. However, it was observed that the accuracy of feature detection decreases with the increase in the image distortion. This is because the difference between the reference image and the masked area caused by the image distortion is severe. To improve this, I represents the entire area of the newly acquired image, enabling the feature search area to be assigned without a separate mask, as shown in Equation (3). This improves the alignment performance. However, it was also observed that the expansion of the search area increases the time required by approximately 30%. Therefore, Equation (2) must be used when the image distortion is small, and Equation (3) must be used appropriately when the distortion is significant. Consequently, only the feature points of the vehicle body that do not change before and after mounting the parts were detected, as shown in Figure 5; further, the image registration algorithms can be

implemented through the image conversion methods, affine and Projective Transform, using the feature points between the two detected images [14]. The inspection image of the part was extracted from the registered image by using the ROI coordinates of the part set in the initially acquired image of SPOT, and high-quality learning data necessary for AI algorithm development could be obtained, as shown in Figure 5.

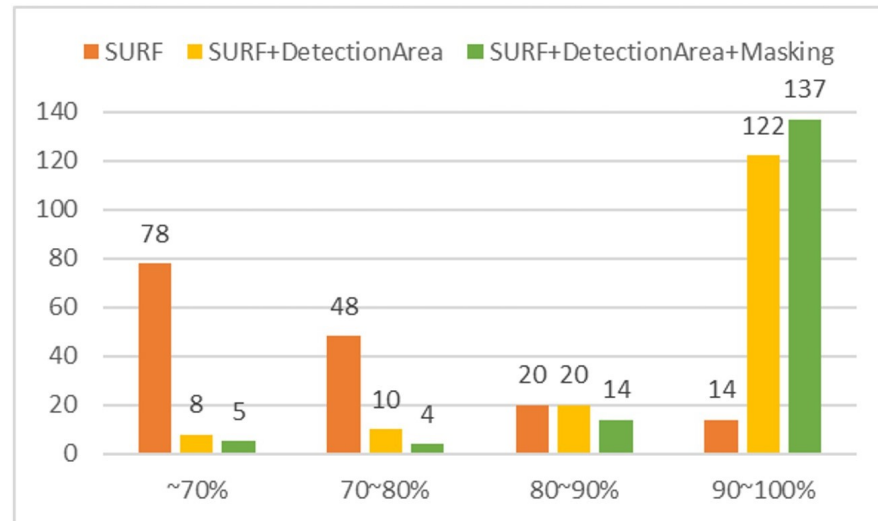


Figure 5. Evaluation of SURF algorithm performance with mixed image search area and image search exclusion area.

4.2. Image Pre-Processing Strategy for Minimizing Lighting Changes Caused by Environmental Variations

Lighting control is possible for small parts, but it is difficult to control lighting in automobile assembly processes due to the large size of the vehicle. Frequent changes in illumination occur due to the inability to control lighting. Additionally, the camera shooting angle changes due to the positional deviation of the mobile robot during image acquisition, causing variations in the gain, exposure, brightness, gamma, and other image features. Contrast changes in the image can degrade the performance of feature point detection through image comparison. Furthermore, changes in the brightness can cause excessive variations at the edges of the image, causing the performance degradation of the CNN network. To address these issues, an algorithm was applied to match the histogram of the acquired image to that of the initial reference image, thereby compensating for overexposure and brightness changes in the image [15,16].

The formula for matching each pixel value, $p_{new}(i, j)$, in the new image, I_{new} , to the histogram of the reference image, I_{ref} , as shown in Figure 6, is given as follows:

$$p_{new}(i, j) = \sum_{k=0}^{L-1} \frac{h_{ref}(k)}{h_{new}(k)} \cdot \max(0, \min(p_{new}^{max}, k + \frac{p_{new}^{max}}{L} - p_{new}(i, j)))$$

Here, $h_{ref}(k)$ and $h_{new}(k)$ represent the histograms of I_{ref} and I_{new} , respectively. L denotes the range of pixel values and p_{new}^{max} denotes the maximum pixel value of I_{new} . This formula matches the histogram of I_{new} to that of I_{ref} , thereby improving the contrast of I_{new} . In software-based image processing after image acquisition, the original image is fixed and the range for change is set. Registration cannot be performed if there is a large difference from the original image. To solve this problem, it is more effective to acquire an image similar to the initially acquired image while changing the camera parameters during image acquisition. However, this is not suitable for automobile production plants

where production cycle time is important since image acquisition time increases. Therefore, the operating time must be considered when developing a system in an industrial setting.

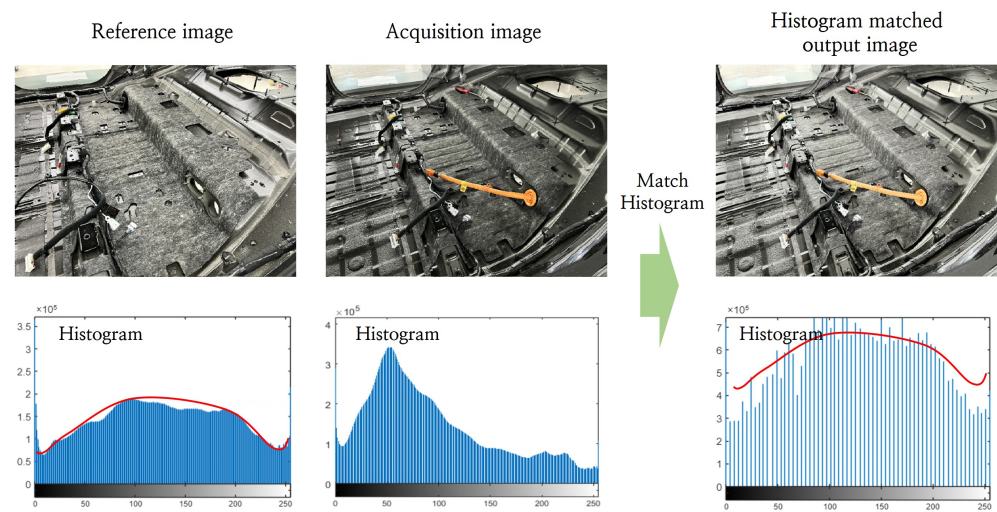


Figure 6. Image brightness correction using histogram matching algorithm.

4.3. Development Plan for Vision-Based AI Algorithms Enabling Maintenance and Continuous Management

In the manual assembly process of an automobile manufacturing plant, the new vehicle release cycle is fast, due to which the parts subject to inspection are frequently changed. Therefore, the cost of developing an inspection algorithm is high, and it is crucial to reduce the development period for inspection from the beginning of production. Additionally, several algorithms must be developed that can inspect a large number of parts due to mixed production, and excessive costs are incurred to maintain the performance. To effectively apply AI technology at industrial sites such as automobile manufacturing plants, maintaining appropriate development costs and reducing the development period are critical issues. In industrial settings, excessive investment costs are incurred for the re-development of algorithms when changing the inspection targets, and there are often cases where equipment is unused because it does not satisfy the required detection performance. Consequently, AI algorithms have a negative perception. In this study, we propose a development plan that can reduce the cost of developing multiple algorithms and drastically reduce the development period to effectively implement AI algorithms in industrial sites.

4.3.1. Cropping Technique to Reduce Learning Data Acquisition Time

The period for acquiring the training images must be reduced to reduce the development period of the deep learning algorithm. However, in industrial settings where product change cycles are fast, it is practically impossible to acquire learning images within a short period. Therefore, a method must be developed to acquire learning data within a short period. In this study, we aimed to maximize the reuse of learning data even when the product changes. Specifically, image pre-processing was performed to crop the inspection parts as large as possible, to exclude the highly variable vehicle structure and color from the images. This ensures that learning and inspection images can be reused even if the vehicle is changed during automobile production. Thus, even if the vehicle is changed, since the structure of parts such as the wiring, clips, connectors, and bolts assembled in the vehicle is similar, a deep learning inspection algorithm can be developed by using the learning data collected from previous vehicles. Figure 7 shows that visual inspection can be applied quickly during new car production by using the deep learning algorithm created for the previous car since the type of clip used for fixing the wiring is similar. Moreover, since the learning data are continuously accumulated and diversity is secured, the performance of the algorithm can be continuously upgraded. To verify this, when launching a new

car with a similar but not the same part type, a short algorithm development test was conducted through transfer learning after securing the minimum quantity of data for the new parts that are similar. The detection performance decreases when performing deep learning with only a small number of new part images of less than 20, as shown in Figure 8. However, the results of transfer learning using the existing algorithm of similar parts exhibit higher accuracy.



Figure 7. Inspection part unit cropping images for reuse of car assembly part types and learning images.

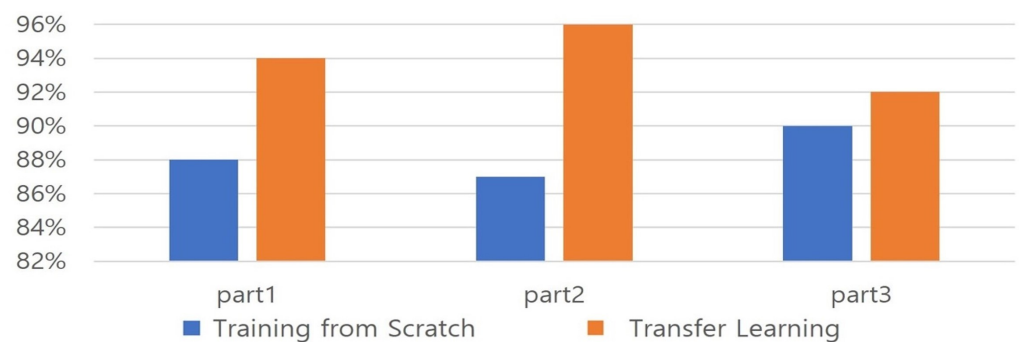


Figure 8. Comparison of results between training from scratch and transfer learning using OK (20 images) and NG (20 images) training data with the resnet101 model.

4.3.2. Minimizing the Development Period and Investment Cost of Algorithm Development Plan

Deep learning techniques are broadly classified into three categories, based on their detection performance and resource utilization: image classification, object detection, and segmentation [17]. The accuracy of the analysis increases in the order of classification, object detection, and segmentation, but the processing resources required also increase with the increase in the amount of data to be processed. Additionally, as the accuracy increases, the cost and labeling period for the training data also increase, which are crucial issues in industrial sites where there are frequent changes in the inspection items. Extensive research is being conducted on auto-labeling to address this issue [18–20]. Although this demonstrates a certain level of performance that can be achieved in industrial sites, separate confirmation is required since even one or two mislabeled data points can significantly impact the algorithm's performance. In this study, classification was applied to the data format used in manual assembly processes for inspecting the automobile parts, considering cost and data acquisition time. Since the classification technique exhibits a lower accuracy than other techniques, it must be improved. To improve the inspection accuracy, we focused on implementing image pre-processing algorithms that can obtain high-quality images and improve the algorithm performance.

4.4. Automation Technology for Maintaining the Performance of AI Algorithms

4.4.1. Automatic Image Augmentation Technology That Accounts for Deviation in Mobile Robot's Shooting Position

The image augmentation technique involves creating new data by appropriately transforming the original image during CNN deep learning training. It is effective in making the model robust when there is insufficient training data, and is important to improve the performance of deep learning algorithms [21,22]. When acquiring images using a mobile robot in an automobile assembly process, the range of image acquisition deviations caused by robot position errors can be statistically calculated. Using the calculated image deviation range value, the image augmentation range can be specified during deep learning training. Thus, the image can be augmented within the same range as the image deviation that can occur due to the positional error of the mobile robot.

Therefore, learning data with the same range of deviation as that of the inspection image can be additionally created, thereby improving the detection performance of the deep learning algorithm. Additionally, it is very effective at maintaining the algorithm performance when the error range changes due to environmental changes and robot deterioration since the algorithm can be automatically re-learned within the calculated deviation range for a certain period of time without the need for an engineer. The T-Matrix value of the image deviation information measured in the image error registration SW of Figure 1 is stored. The image acquisition deviation range for each robot position can be calculated as shown in Figure 9. It can be observed that the deviation of the error caused by the different position of the robot and the different pose of the arm for part shooting is different. Essentially, when training the deep learning algorithm using different robot position deviations for each part shooting position, an appropriate image augmentation value for each position can be used as shown in Figure 10. This method can prevent the degradation of the algorithm performance by learning with an image that is completely different from the inspection image during algorithm learning. Additionally, the error range analyzed by the image error registration software can be automatically parameterized for image augmentation without the intervention of an engineer.

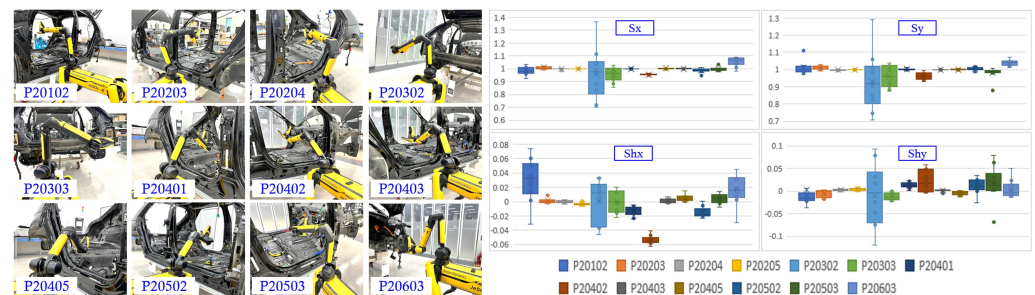


Figure 9. For each set of 100 images captured by the robot at each position, T-Matrix can be used to extract the range of augmentations.

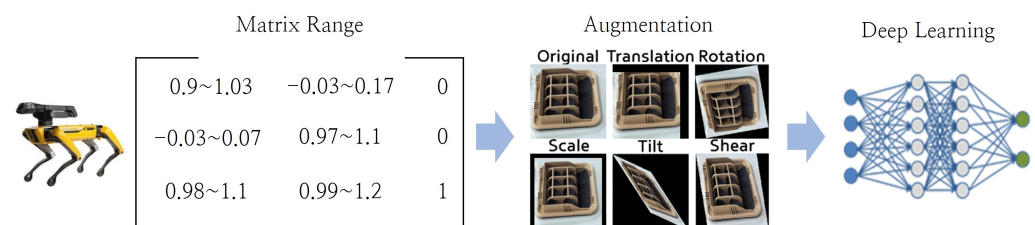


Figure 10. Using the range of image deviation caused by robot position errors as image augmentation parameters.

To verify the effectiveness of the image augmentation technique using T-Matrix values, we compared the results of learning with parameters set by engineers to the results of learning with parameters set automatically using the T-Matrix values. The results indicated

equivalent algorithm performance, as shown in Figure 11. Conversely, we observed that the algorithm's performance deteriorated when the image was augmented with a difference of $\pm 5\%$ or more from the error range for each robot position. This implies that incorrect parameter settings by engineers during deep learning can degrade the algorithm performance, and there is a high possibility of algorithm performance deviation based on the engineer's ability. Additionally, maintenance costs can be minimized because the system can automatically learn the algorithm's performance despite environmental changes or robot deterioration, without requiring an engineer.

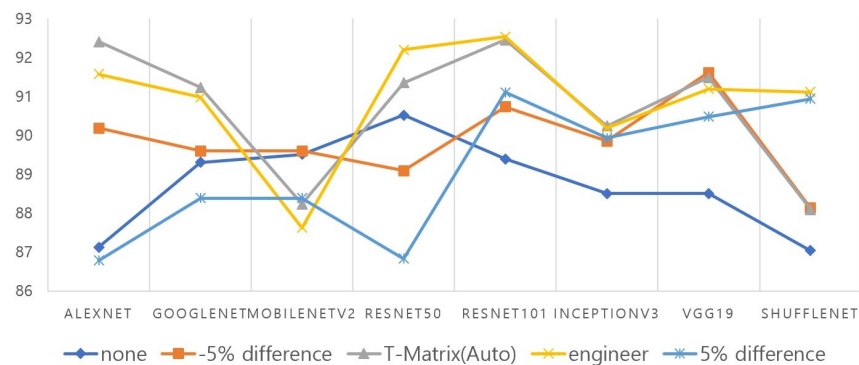


Figure 11. Comparison graph of learning accuracy for each representative network according to image augmentation error range of $\pm 5\%$, T-Matrix (auto), and engineer's experience level.

4.4.2. Automatic Data Acquisition Method for Re-Learning AI Algorithms

It is difficult to implement the AI vision inspection system in the industrial field because the performance of the algorithm deteriorates due to changes in the inspection environment. When the algorithm performance degrades, it is essential to manage performance through algorithm re-learning immediately. However, it is difficult to maintain the performance of the algorithm owing to the high time and cost requirements incurred during the labeling task to transform the acquired data into learning data. Essentially, an automatic data labeling method is required to maintain algorithm performance for the implementation of AI technology in industrial sites. In this study, a classification score was used to automatically label the acquired data. To automatically secure the re-learning data, the cross entropy score value was set as high as possible in the Softmax step of the image classification process to ensure that it was OK, as shown in Figure 1. The OK data is automatically classified without operator intervention, and NG and NA, which require operator correction, can be labeled as data by clicking through a GUI that enables the operator to determine whether it is OK, NG, or NA.

4.5. Industrial Field AI Algorithm Development Plan

To improve the performance of deep learning algorithms, obtaining sufficient-quality training data is the most important aspect. However, obtaining sufficient learning data for developing vision AI algorithms within a short period that corresponds to the time of product production in industrial settings such as automobile assembly processes is extremely challenging. Since adequate data cannot be used when developing AI algorithms in industrial settings, a method must be developed to satisfy the algorithm's performance requirements using only a small amount of data that can be initially acquired. Acquiring learning data in industrial settings is very difficult, as mentioned earlier, and problems can arise due to the imbalance of the OK and NG data. The anomaly detection technique is being analyzed to solve this problem; however, it is still inadequate for implementation in industrial settings with diverse inspection images. In this study, we developed algorithms with various combinations of similar or different part images to improve the AI algorithm performance using a small amount of data, as shown in Figure 12. By selecting and using the inspection algorithm for parts with high accuracy among the developed algorithms,

a high-performance detection algorithm was developed within a short period using only a small amount of learning data. To improve the algorithm detection performance, a large amount of learning data with diversity must be obtained. However, the best and easiest way to increase the diversity and accuracy of learning data is to mix between automobile parts to reduce the variance and bias, which is practically impossible to achieve under industrial settings.

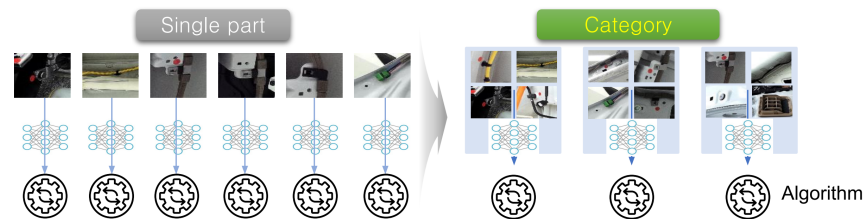


Figure 12. Create categories of similar parts, repeatedly learn with mixed categories, evaluate the accuracy of each part, and use the algorithm of the part with the highest performance.

The Boosting technique is one of the methods that can be used to solve the generalization problem of machine learning algorithms. Boosting assigns more weight to misclassified samples from previous training results on the same data to predict more accurate results [23,24]. However, Boosting has dependencies between data and network models, requiring various tuning operations to achieve the optimal performance. Conversely, the proposed learning method is very simple and intuitive, making it easy to use in industrial automation systems. Additionally, mixing similar parts of data can increase the generalization performance while reducing the variance and bias by increasing the diversity and quantity of the training data. This algorithm takes three primary inputs: the total number of car parts (N), an accuracy threshold (t) for effective model evaluation, and the number of iterations (I) to refine the models through repeated training. The output obtained is the trained models (M) for each car part that satisfy the accuracy threshold. The core process includes initializing a tracking table (T), selecting random subsets of parts for model training, comparing their accuracy against t , and iterating until the most accurate algorithms for each part are identified. After randomly mixing inspection part images as shown in Algorithm 1 and training the model, only the parts with high inspection accuracy are used in the AI algorithm. For parts with low accuracy, the algorithm is retrained using a different mix of parts until the inspection accuracy is sufficiently high to be used in the AI algorithm. Using this method, we were able to improve the performance of the AI algorithm with minimal data and resources within a short period. Using the proposed AI algorithm development method, the results of training on the same data, same network model, and same parameter setting presents higher accuracy than training for a single part, as shown in Figure 13. In particular, applying the category technique to parts with an accuracy of 0.5 or less increased the accuracy to 0.9 or higher.

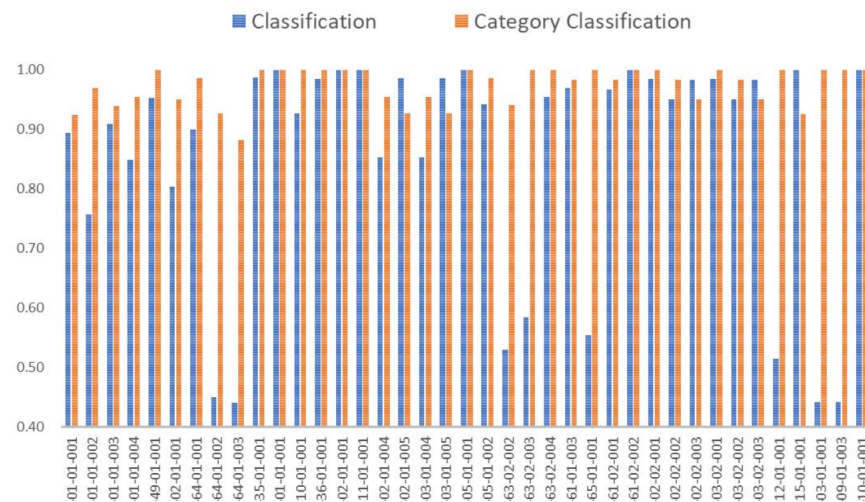


Figure 13. Performance improvement in algorithms through finding optimal AI algorithm by proposed similar part mixing.

Algorithm 1. Finding optimal algorithm for car parts

Input: Number of car parts N , accuracy threshold t , number of iterations I .

Output: Trained models M for each part.

1: Initialize an empty table T to store the part number, algorithm number, and accuracy.

2: **for** $i \leftarrow 1$ to I **do**

3: **for** $j \leftarrow 2$ to N **do**

4: Randomly select j parts to form a training dataset.

5: Train a model M_j on the training dataset.

6: **for** each part p in the j parts **do**

7: Evaluate the model M_j on the test dataset FOR part p .

8: **if** the accuracy of M_j for part p is above the threshold t **then**

9: Store the part number, algorithm number, and accuracy in the table T .

10: **end if**

11: **end for**

12: **end for**

13: **end for**

14: Group the parts by the algorithm number that achieved the highest accuracy for each part.

15: Save the trained model for each part and its corresponding algorithm number.

16: Display the table, T , with part numbers, algorithm numbers, and accuracies.

5. AI System Empirical Evaluation

In this study, we present the development and application of a vision AI system that can acquire images of the assembly process and inspect assembly defects using a vision AI algorithm, by utilizing the mobile robot ‘SPOT’. This robot can operate in conjunction with workers on the manual assembly lines of automobile manufacturing plants. In previous studies, there was no device capable of acquiring images while operating alongside workers, and developing a solution to inspect dozens of assembly parts incurred excessive costs. Furthermore, high maintenance expenses post-system-implementation presented challenges for mass production. This paper addresses and resolves these issues. Currently, the system is being applied and operated in the prototype phase 1 process at the Singapore plant of Hyundai Motors, where it has successfully identified multiple instances of defects caused by the assembly mistakes of inexperienced workers at the initial stages of vehicle production in real-time. Additionally, this paper proposes a method for the continuous automatic collection of training data to address the lack of training data during the initial development phase. Consequently, while the average performance of the algorithm for 39 parts was initially 88%, continuous data collection and the proposed algorithm learning method have enhanced the performance to 97.4%.

6. Future Work

Several studies have been conducted on the application of AI technology in various industries. However, most of these studies approach technology development under the assumption that AI can solve everything, and this presents a major obstacle to implementing AI technology in industrial applications. To implement AI technology for improving industrial productivity, it is important to develop good deep learning networks as well as to collect training data, improve algorithm detection performance in constrained environments, maintain algorithms, and reduce the time and cost. If AI algorithms are implemented in industrial fields without such strategies, there is a high risk of failure due to real problems. Research is currently being conducted to obtain training data using 3D data because it is difficult to obtain training data at industrial sites; however, this approach is impractical [25,26]. Therefore, more realistic solutions must be developed. To enable the widespread application of AI technology in industrial fields, continuous research is required to realistically reduce the costs of developing and maintaining AI technology.

7. Conclusions

This paper analyzed the problems of industrial sites where the sustainable application of a visual AI inspection system was difficult due to frequent changes in the environmental conditions and inspection targets, and developed technologies to solve these problems. We developed image acquisition technology using the mobile robot SPOT to obtain high-quality learning data and inspection images for the real-time visual inspection of the manual assembly process of automobiles. We proposed an AI system development architecture that could be effectively applied to industrial sites. We also improved the development of AI inspection algorithms by applying technologies to reduce the learning data acquisition period, save investment costs, improve algorithm performance, and automate the algorithm maintenance. This helped in drastically reducing the existing problems. In particular, the similarity of vehicle parts was used to develop the algorithm for new B-vehicle parts by utilizing the algorithm developed for A-vehicle parts, as shown in Figure 14. If the new B-vehicle parts were identical to the A-vehicle parts, the inspection could be performed using the A-vehicle part algorithm. If the new B-vehicle parts were not identical to the A-vehicle parts but similar, the development period of the algorithm could be reduced by transferring the learning of the B-vehicle parts to the A-vehicle part algorithm. Consequently, a vision AI system with the required detection performance during the early stages of production could be applied to detect assembly defects. Furthermore, it was possible to detect many defects caused by the low skill level of workers during the initial production of new cars in automobile production factories, which could significantly improve the quality of automobile assembly. Lastly, the effective AI technology development method proposed in this study will serve as a useful guide for the implementation of AI technology in industrial sites.

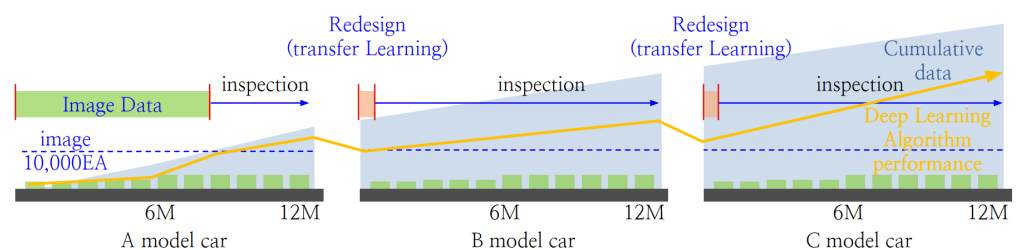


Figure 14. Algorithm development is shortened through transfer learning with same/similar part algorithms and AI algorithm improvement through continuous accumulation of automobile assembly part image data.

Author Contributions: Conceptualization, C.Y., D.K. and J.K.; Software, C.Y.; Validation, C.Y.; Formal analysis, J.K.; Investigation, J.K.; Resources, D.K.; Writing—original draft, C.Y.; Writing—review & editing, D.-S.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Hyundai Motor Company under project number 2021_CSTG_0174.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the first author due to Hyundai Motor Company’s internal policies and therefore are not publicly available. Data access requests can be directed to the first author.

Acknowledgments: We thank the Hyundai Motor Company for their support and resources provided for conducting this research.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could appear to influence the work reported in this paper. However, Changmo Yang, Dongweon Kang, and JinSeok Kim are employees of Hyundai Motor Company, which provided funding and technical support for this work. The funder had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
SLAM	Simultaneous Localization and Mapping
SW	Software
ToF	Time-of-Flight
UWB	Ultra-Wideband
SURF	Speeded-Up Robust Features
ROI	Region of Interest
CNN	Convolutional Neural Network
OK	Acceptable or Correct
NG	Not Good or Incorrect
NA	Not Applicable or Not Available
AGV	Automated Guided Vehicle
T-Matrix	Transformation Matrix
F-Mark	Fiducial Mark

References

1. Yang, J.; Li, S.; Wang, Z.; Dong, H.; Wang, J.; Tang, S. Using deep learning to detect defects in manufacturing: A comprehensive survey and current challenges. *Materials* **2020**, *13*, 5755. [CrossRef] [PubMed]
2. Block, S.B.; da Silva, R.D.; Dorini, L.B.; Minetto, R. Inspection of imprint defects in stamped metal surfaces using deep learning and tracking. *IEEE Trans. Ind. Electron.* **2020**, *68*, 4498–4507. [CrossRef]
3. Mazzetto, M.; Teixeira, M.; Rodrigues, É.O.; Casanova, D. Deep learning models for visual inspection on automotive assembling line. *arXiv* **2020**, arXiv:2007.01857.
4. Hemamalini, V.; Rajarajeswari, S.; Nachiyappan, S.; Sambath, M.; Devi, T.; Singh, B.K.; Raghuvanshi, A. Food quality inspection and grading using efficient image segmentation and machine learning-based system. *J. Food Qual.* **2022**, *2022*, 5262294. [CrossRef]
5. Lang, W.; Hu, Y.; Gong, C.; Zhang, X.; Xu, H.; Deng, J. Artificial intelligence-based technique for fault detection and diagnosis of EV motors: A review. *IEEE Trans. Transp. Electrification* **2021**, *8*, 384–406. [CrossRef]
6. Zhou, Q.; Chen, R.; Huang, B.; Liu, C.; Yu, J.; Yu, X. An automatic surface defect inspection system for automobiles using machine vision methods. *Sensors* **2019**, *19*, 644. [CrossRef] [PubMed]
7. Wang, J.; Fu, P.; Gao, R.X. Machine vision intelligence for product defect inspection based on deep learning and Hough transform. *J. Manuf. Syst.* **2019**, *51*, 52–60. [CrossRef]
8. Chalapathy, R.; Menon, A.K.; Chawla, S. Anomaly detection using one-class neural networks. *arXiv* **2018**, arXiv:1802.06360.
9. Boston Dynamics. SPOT. Available online: <https://dev.bostondynamics.com/> (accessed on 21 March 2024).

10. Cheng, T.; Venugopal, M.; Teizer, J.; Vela, P. Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Autom. Constr.* **2011**, *20*, 1173–1184. [CrossRef]
11. Karedal, J.; Wyne, S.; Almers, P.; Tufvesson, F.; Molisch, A.F. A measurement-based statistical model for industrial ultra-wideband channels. *IEEE Trans. Wirel. Commun.* **2007**, *6*, 3028–3037. [CrossRef]
12. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
13. Bansal, M.; Kumar, M.; Kumar, M. 2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimed. Tools Appl.* **2021**, *80*, 18839–18857. [CrossRef]
14. Wiki. 2D Affine Transformation Matrix. Available online: https://en.wikipedia.org/wiki/Affine_transformation (accessed on 21 March 2024).
15. Rother, C.; Minka, T.; Blake, A.; Kolmogorov, V. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 1, pp. 993–1000.
16. Chen, H.M.; Varshney, P.K. Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. *IEEE Trans. Med. Imaging* **2003**, *22*, 1111–1119. [CrossRef]
17. Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M.P.; Shyu, M.L.; Chen, S.C.; Iyengar, S.S. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv.* **2018**, *51*, 1–36. [CrossRef]
18. Fischl, B.; Salat, D.H.; Busa, E.; Albert, M.; Dieterich, M.; Haselgrove, C.; Van Der Kouwe, A.; Killiany, R.; Kennedy, D.; Klaveness, S.; et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **2002**, *33*, 341–355. [CrossRef] [PubMed]
19. Gildea, D.; Jurafsky, D. Automatic labeling of semantic roles. *Comput. Linguist.* **2002**, *28*, 245–288. [CrossRef]
20. Mei, Q.; Shen, X.; Zhai, C. Automatic labeling of multinomial topic models. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 490–499.
21. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. 2009. Available online: <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf> (accessed on 21 March 2024).
22. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
23. Chen, C.; Xiong, Z.; Tian, X.; Zha, Z.J.; Wu, F. Real-world image denoising with deep boosting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 3071–3087. [CrossRef] [PubMed]
24. Chen, C.; Xiong, Z.; Tian, X.; Wu, F. Deep boosting for image denoising. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–18.
25. Dosovitskiy, A.; Springenberg, J.T.; Tatarchenko, M.; Brox, T. Learning to generate chairs, tables and cars with convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 692–705. [CrossRef] [PubMed]
26. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [PubMed]

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.