

Article

Vision-Based Acquisition Model for Molten Pool and Weld-Bead Profile in Gas Metal Arc Welding

Gwang-Gook Kim ^{1,2} , Dong-Yoon Kim ^{1,*}  and Jiyoung Yu ^{1,*}¹ Flexible Manufacturing R&D Department, Korea Institute of Industrial Technology, Incheon 21999, Republic of Korea; kimgg@kitech.re.kr² Department of Mechanical Convergence Engineering, Hanyang University, Seoul 04763, Republic of Korea

* Correspondence: kimdy@kitech.re.kr (D.-Y.K.); willow@kitech.re.kr (J.Y.)

Abstract: Gas metal arc welding (GMAW) is widely used for its productivity and ease of automation across various industries. However, certain tasks in shipbuilding and heavy industry still require manual welding, where quality depends heavily on operator skill. Defects in manual welding often necessitate costly rework, reducing productivity. Vision sensing has become essential in automated welding, capturing dynamic changes in the molten pool and arc length for real-time defect insights. Laser vision sensors are particularly valuable for their high-precision bead profile data; however, most current models require offline inspection, limiting real-time application. This study proposes a deep learning-based system for the real-time monitoring of both the molten pool and weld-bead profile during GMAW. The system integrates an optimized optical design to reduce arc light interference, enabling the continuous acquisition of both molten pool images and 3D bead profiles. Experimental results demonstrate that the molten pool classification models achieved accuracies of 99.76% with ResNet50 and 99.02% with MobileNetV4, fulfilling real-time requirements with inference times of 6.53 ms and 9.06 ms, respectively. By combining 2D and 3D data through a semantic segmentation algorithm, the system enables the accurate, real-time extraction of weld-bead geometry, offering comprehensive weld quality monitoring that satisfies the performance demands of real-time industrial applications.



Citation: Kim, G.-G.; Kim, D.-Y.; Yu, J. Vision-Based Acquisition Model for Molten Pool and Weld-Bead Profile in Gas Metal Arc Welding. *Metals* **2024**, *14*, 1413. <https://doi.org/10.3390/met14121413>

Academic Editor: Dariusz Rozumek

Received: 13 November 2024

Revised: 6 December 2024

Accepted: 8 December 2024

Published: 10 December 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/>).

Keywords: gas metal arc welding; real-time quality monitoring; molten pool; weld-bead profile; semantic segmentation

1. Introduction

Gas metal arc welding (GMAW) is extensively used across various industries owing to its high productivity, reliability, and ease of automation. However, in sectors such as shipbuilding and heavy manufacturing, certain weld joints remain difficult to automate, requiring skilled operators for manual welding. In such cases, weld quality heavily relies on the operator's expertise, as they must monitor the molten pool directly through a welding helmet. GMAW is susceptible to defects, including burn-through, lack of fusion, slag inclusion, and irregular bead shape. These defects often demand rework to meet quality standards or welding procedure specifications, leading to increased costs and reduced productivity. To address these challenges, vision sensing has become essential in automated welding systems for ensuring weld quality. By emulating a welder's visual perception of the weld pool, vision sensing provides more intuitive feedback than other methods, making it especially advantageous in welding automation.

A variety of vision sensing technologies have been investigated for use in welding processes, with laser vision sensors emerging as the most researched and widely applied. These sensors can capture high-precision bead profile data for different joint geometries, including multi-pass welds, making them highly suitable for welding automation. However, most laser vision sensors on the market are costly and often require offline scanning

of the weld after the process, leading to increased time costs. Huang et al. [1,2] developed a relatively cost-effective and efficient laser vision sensor for nondestructive weld quality inspection, offering an alternative to more expensive options. Despite its benefits, this system cannot perform real-time quality inspection during welding. Han et al. [3] addressed this limitation by designing a laser vision sensor with a shielding plate and narrowband filter to reduce arc light interference, enabling real-time bead measurements during welding. Additional research [4–6] has examined the use of laser vision sensors for bead profile acquisition in quality assessments. However, accurately extracting parameters such as bead width and height from the acquired profiles remains challenging. To overcome this limitation, Ye et al. [7] proposed a method that applies the expectation-maximization algorithm to correct bead profile distortions caused by noise or weld deformation, improving profile accuracy. Similarly, Lee et al. [8] employed global and adaptive thresholding algorithms to capture real-time bead profiles during lap-joint GMAW and used deep neural networks to predict weld quality. Nevertheless, the performance of this method is constrained by specific threshold settings and can be impacted by environmental changes, arc light, spatter, and fumes.

Recent studies have focused on capturing and monitoring molten pool images directly during GMAW. Wang et al. [9] introduced a hybrid model that combines visual features from molten pool images with welding voltage signals to detect weld defects. However, their approach captured molten pool images only during the base current phase of the welding cycle, which limited a complete observation of the molten pool behavior. Additionally, converting welding voltage data into images using a short-time Fourier transform and processing these with convolutional neural networks introduces considerable computational complexity, potentially necessitating high computing power for real-time monitoring. Yu et al. [10] used color charge-coupled device sensors to capture molten pool images and utilized information on pool area, length, and width to predict penetration depth in GMAW. Molten pool images provide critical insights by enabling the real-time observation of dynamic changes in the molten pool surface and arc length variations, aiding in early defect detection. Numerous studies have explored online molten pool monitoring methods for evaluating weld quality.

Chen et al. [11] predicted weld penetration depth during gas–tungsten arc welding using a passive vision method, while Xia et al. [12] developed a model to classify weld states using molten pool images from passive vision as input variables. Zhang et al. [13] explored the use of high-dynamic-range vision to predict penetration depth in GTAW, which Wang et al. [14] later enhanced by developing an improved algorithm for keyhole geometry extraction and monitoring. Lu et al. [15] employed a deep residual network to quantitatively predict bead height and penetration depth in cold metal transfer welding. Xu et al. [16] introduced a method to control penetration depth by visually sensing molten pool width during the GMAW of 304 stainless steel, but its reliance on a fuzzy proportional integral derivative controllers limited its applicability in industrial environments. Nomura et al. [17] proposed a method to predict penetration depth in gas metal arc welding by simultaneously observing both the molten pool and the weld bottom with a high-speed camera and CMOS sensor, analyzed using a convolutional neural network. Although this configuration allowed real-time monitoring, it required a high-speed camera for molten pool capture and an additional camera for bottom weld observation, resulting in a complex setup. However, molten pool images capture only 2D information and lack depth, which limits their ability to fully characterize the 3D geometry of the weld bead. In contrast, using a line laser to obtain the weld-bead profile allows for the precise extraction of 3D data, such as weld-bead width, height, and deposition area, even though it restricts the real-time observation of the welding process itself. These 3D profile data are critical for assessing weld quality and stabilizing automated welding processes. Therefore, an integrated approach combining 2D image data with 3D profile information is necessary to comprehensively characterize weld-bead geometry.

In this study, we present a deep learning-based monitoring system capable of the real-time observation of a molten pool and capture of the weld-bead profile during GMAW, effectively addressing interference from intense arc light through an optimized optical design. Our vision system overcomes the offline inspection limitations of conventional laser vision sensors by enabling the real-time acquisition of both molten pool images and bead profiles, leveraging 2D and 3D data simultaneously for comprehensive weld quality monitoring. This system combines a vision camera and line laser within an optical setup that facilitates the concurrent observation of the molten pool and laser-projected profiles during GMAW. To achieve accurate 3D extraction from the weld bead's laser-projected profiles, we employ a deep learning model based on semantic segmentation to efficiently capture bead profiles.

2. Methodology

2.1. Experimental Setup

2.1.1. Welding Condition

The base metal used in the experiment was ISO 630 E 355 [18], with dimensions of 300 mm × 100 mm, a thickness of 15 mm, a V-groove angle of 90°, and a root face length of 3 mm (Figure 1). To reduce the deformation of the test plate during welding, tack welding was applied at the start and end points of the weld line before the main welding process. The filler wire used was ISO 636-A W 42 5 2Si [19] (Lincoln Electric Company, Cleveland, OH, USA) with a diameter of 1.2 mm; its chemical composition is shown in Table 1. The welding power source was a Daihen P500L (Daihen Corporation, Osaka, Japan) operating in direct current pulse mode. Welding was carried out using a YASKAWA six-axis robot (Yaskawa, Kitakyushu, Japan) to control the welding torch. The wire feed rate was set at 7.4 m/min, with welding voltages of 22.3, 25.3, and 28.3 V, as detailed in Table 2. The contact tip-to-work distance was 18 mm, and the welding speed was 30 cm/min. Shielding gas was a 9:1 mixture of argon and CO₂, with a flow rate of 15 L/min.

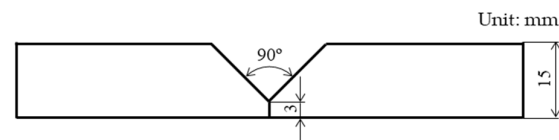


Figure 1. Schematic design of V-groove welding test plate.

Table 1. Chemical composition of the base metal E 355 and filler wire W 42 5 2Si (wt.%).

Content	C	Si	Mn	P	S
E 355	0.1852	0.343	1.355	0.011	0.0044
W 42 5 2Si	0.07	0.65	1.14	0.015	0.01

Table 2. Welding experiment conditions for the weld quality classification and prediction model.

Condition	Wire Feed Rate (m/min)	Welding Voltage (V)
A	7.4	22.3
B	7.4	25.3
C	7.4	28.3

2.1.2. Vision System

The configuration of the vision system used in this experiment is shown in Figure 2. The system included a vision camera, lens, filters, a first-surface mirror, and a line laser. The vision camera, lens, and filters were assembled following the optical design method described in a previous study [20]. The vision camera model was a UI-3271LE (IDS Imaging, Obersulm, Germany), equipped with a fixed 16 mm focal length lens set at an aperture of

f/4. The filter setup included an 830 nm band-pass filter with a 10 nm full width at half maximum (FWHM) and a UV filter. A line laser was chosen to match the camera's filter wavelength range, featuring a non-Gaussian lens with a 30° line angle. It operated on a DC 5 V input, delivering an output power of 200 mW. The vision system was mounted in parallel with the welding torch using a fixed bracket, maintaining a height of 130 mm above the workpiece to align with the torch movement. Due to the substantial heat and fumes generated during welding, the vision system is prone to potential damage, as elevated temperatures can reduce laser output, degrade camera performance, and cause malfunctions. Specifically, the wavelength of the line laser may shift with temperature increase. Although cooling systems are typically unnecessary for lasers with output powers in the tens to hundreds of milli-watts range, the heat and fumes in this experimental setup warranted an air-cooling structure to prevent excessive temperature rise within the vision system. Additionally, a bracket was installed to blow air in the opposite direction of the welding path, forming an air jacket to reduce spatter particle accumulation on the laser-transmitting glass and the vision camera's front glass. The vision module was designed to facilitate the replacement of the glass when a significant amount of spatter particles accumulates on its surface, obstructing the acquisition of clear images. Aligning the vision camera and welding torch is crucial for capturing both the molten pool and line laser shape within the camera's field of view. Preliminary experiments indicated that the optimal angle between the vision camera and welding torch was 45° . However, maintaining this angle while capturing the molten pool and weld bead in a single image increased the system's overall size, which could interfere with jigs securing the weld during the welding process. To reduce the size of the vision system module, the camera was positioned facing away from the weld area, with a first-surface mirror placed inside to reflect the laser and weld images. Unlike conventional mirrors, the reflective coating on a first-surface mirror is applied to the front surface of the glass, preventing refraction and scattering. This feature makes the first-surface mirror critical for precise reflection without distance errors, ensuring measurement accuracy within the vision system.

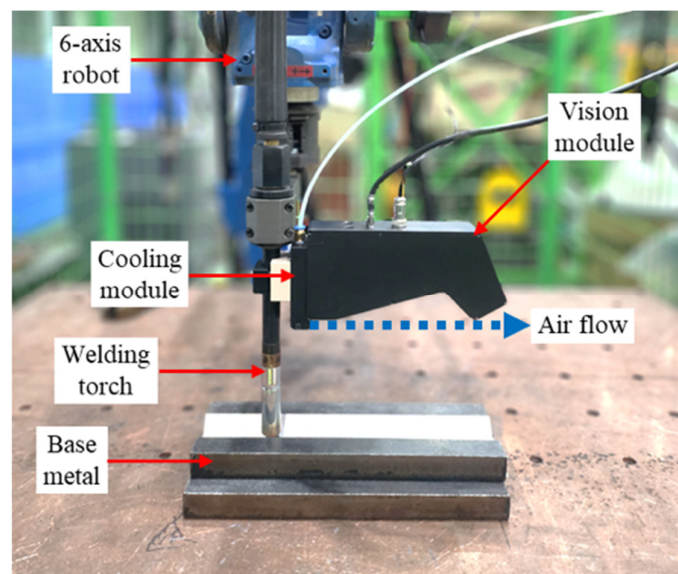
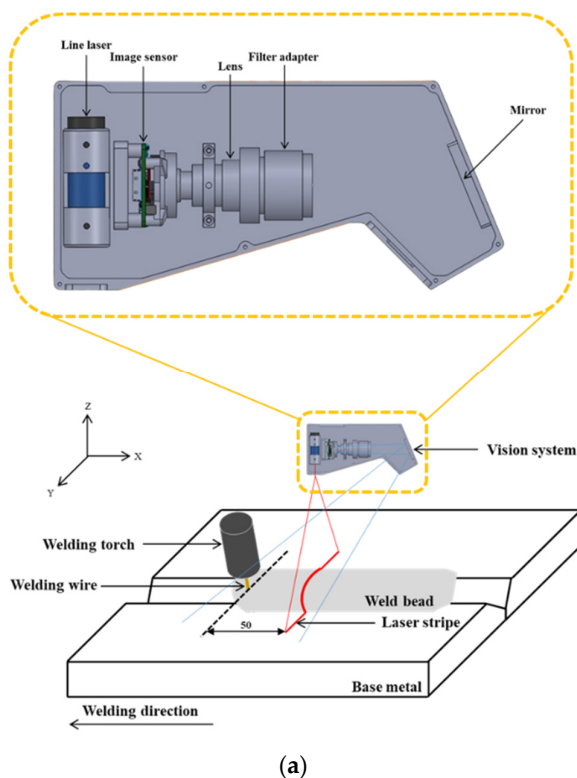


Figure 2. Configuration of vision system module: (a) schematic; (b) experimental setup.

2.2. Molten Pool State Classification

With the proposed vision system, molten pool images were captured during the GMAW process to classify weld-bead conditions. Similarly to a welder observing the molten pool during manual welding, monitoring the molten pool surface is essential in robotic welding processes. The weld-bead conditions identified from the captured molten pool images fall into several categories: a normal condition with no defects, burn-through (often resulting from insufficient gap management between workpieces or excessive heat input), and spatter, typically due to unclean workpieces or incorrect voltage settings, as illustrated with example images in Figure 3. To assess weld-bead conditions, V-groove test plates were welded using GMAW, and images of the molten pool were captured for classification. The molten pool states observed during welding were categorized into three classes. For this classification, deep learning models ResNet50 [21] and MobileNetV4 [22] were employed. ResNet50, with its deep network architecture, effectively learns intricate image features, allowing for precise classification. Meanwhile, MobileNetV4, a lightweight convolutional neural network optimized for efficiency, minimizes latency, ensuring that the monitoring system meets real-time performance requirements with rapid inference times. All computational tasks in this study were performed using Python (version 3.10) on a system equipped with an Intel Core i7-13700KF processor (Intel, Santa Clara, CA, USA), 64 GB of 5600 MHz DDR5 RAM (Samsung, Suwon, Republic of Korea), and an NVIDIA GeForce RTX 3090 (NVIDIA, Santa Clara, CA, USA).

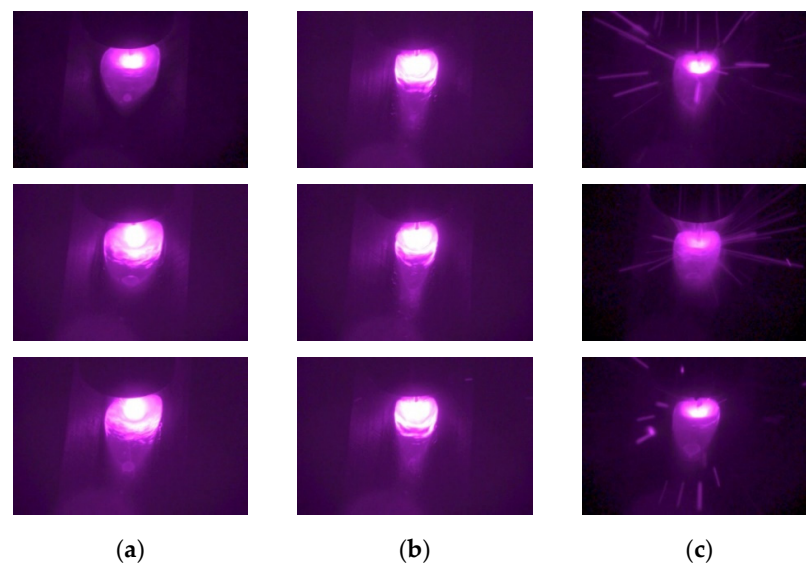


Figure 3. Sample molten pool images during GMAW process: (a) normal; (b) burn-through; (c) spatter.

2.3. Semantic Segmentation-Based Deep Learning Model

This semantic segmentation typically employs an encoder–decoder architecture, where the encoder extracts key features from the input image. The structure of the encoder, or backbone, plays a crucial role in determining the model’s performance. While CNN-based backbones, such as ResNet, perform well in many tasks, they face limitations, such as a strong focus on local information, which can hinder their ability to adapt to new datasets. To address these challenges, this study adopted the SegFormer model [23], a transformer-based architecture that balances efficiency and accuracy in segmentation tasks. The structure of the SegFormer model is illustrated in Figure 4. Unlike traditional CNNs, SegFormer uses the MixVision Transformer (MiT) as its backbone to capture both global and local features effectively. The MiT follows a hierarchical approach, initially learning local information with small patches and progressively incorporating larger patches with varying spatial reduction ratios to capture global features. This approach enhances both computational

efficiency and model generalization. The MiT sets the input embedding dimension of the encoder to 32 and consists of four stages, with each stage containing two layers. In the multi-head attention mechanism, the number of heads per stage is set to one, two, five, and eight, enabling the extraction of features at multiple levels. The patch sizes for each stage are configured as 7, 3, 3, and 3, with corresponding spatial reduction ratios of 8, 4, 2, and 1 to improve computational efficiency. Additionally, a drop-path probability of 0.1 is applied to enhance the model's ability to generalize. The MiT model employs overlapping patch embedding to improve spatial continuity and feature representation by allowing adjacent patches to overlap. The output image size for each layer is calculated as follows:

$$\text{Output height } H' = \frac{H + 2P - D}{S} + 1, \text{Output width } W' = \frac{W + 2P - D}{S} + 1, \quad (1)$$

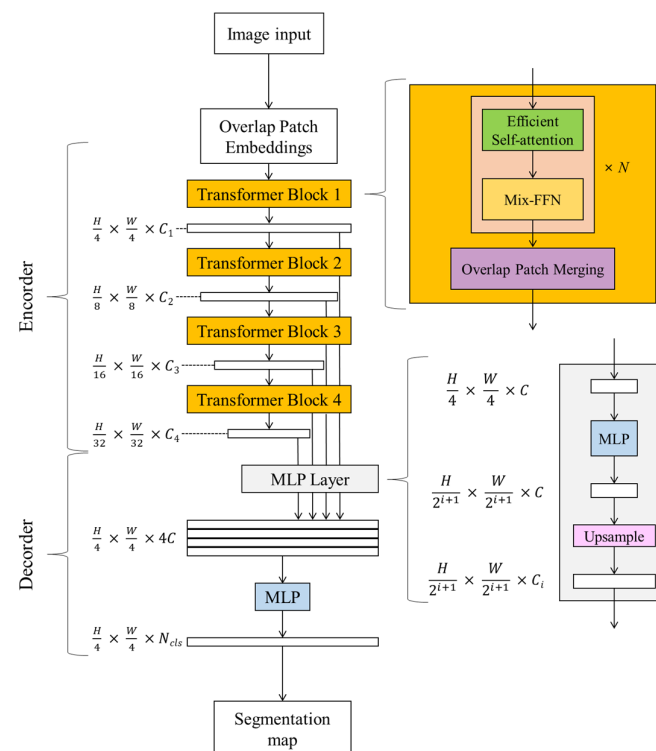


Figure 4. Schematic of the SegFormer model structure.

The patch sizes, strides, and padding values are listed in Table 3.

Table 3. Output image size when using SegFormer's patch size, stride, and padding values with an input image size of width (W) 224 × height (H) 224 pixels.

Stage	Patch Size D	Stride S	Padding P	Output Height $H' \times$ Output Width W'
1	7	4	3	56×56
2	3	2	1	28×28
3	3	2	1	14×14
4	3	2	1	7×7

In the transformer block, self-attention allows each position in the input features to evaluate its relevance to all other positions. Self-attention is defined as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{\text{head}}}}\right)V, \quad (2)$$

where Query (Q) contains information about the current position and identifies significant input positions, Key (K) describes each position within the input features and measures its relevance to the Query, and Value (V) provides information based on the alignment of the Key and Query. Here, d_{head} represents the dimension of each attention head, and scaling with $\sqrt{d_{\text{head}}}$ is applied to stabilize the Softmax operation by normalizing the magnitude of the dot product QK^T , ensuring numerical stability during the attention computation. The computational complexity of self-attention is $O(N^2)$, which makes it expensive for high-resolution imaging. To mitigate this, efficient self-attention reduces the dimensionality of the key by a factor of R , thereby decreasing the complexity to $O\left(\frac{N^2}{R}\right)$. The Mix-FFN replaces positional encoding with 3×3 convolutions, integrates spatial information, and uses the GELU function for smooth nonlinearity. The encoder's multilevel features, F_i , are progressively downsampled to capture complex features. F_i in the i -th layer can be expressed as follows:

$$F_i = \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i, \quad (3)$$

where i takes values from 1 to 4, with the feature map F_i at each layer defined relative to the input image resolution, as shown in Table 4. Here, C_i represents the number of channels in the feature map at the i -th layer. As i increases, C_i typically grows to allow the model to capture more complex and abstract features while compensating for the loss of spatial resolution. As the model progresses through subsequent layers, the resolution gradually decreases while the number of channels increases, allowing it to capture increasingly complex features.

Table 4. Resolution of feature map F_i at each layer.

First Layer (Stage 1)	Second Layer (Stage 2)	Third Layer (Stage 3)	Fourth Layer (Stage 4)
$F_1 = \frac{H}{4} \times \frac{W}{4} \times C_1$	$F_2 = \frac{H}{8} \times \frac{W}{8} \times C_2$	$F_3 = \frac{H}{16} \times \frac{W}{16} \times C_3$	$F_4 = \frac{H}{32} \times \frac{W}{32} \times C_4$

The decoder adjusts the channel dimensions using MLP layers, upsamples the feature maps, and concatenates them into a single map, F . The segmentation map M is then generated as follows:

$$M = \text{Linear}(C, N_{\text{cls}})(F), \quad (4)$$

The objective of this study was to optimize and streamline a deep learning model that currently exhibits limited performance in real-time monitoring. The goal was to enhance parameter efficiency in order to achieve a frame rate suitable for real-time applications. To this end, post-processing was applied to the line laser-shaped object data to improve the model's robustness and compactness. Figure 5 illustrates the process of extracting the weld-bead profile from the original image during the GMAW process, captured using a line laser, by applying the proposed model. Initially, a region of interest (ROI) was defined on the line laser-shaped object to isolate the weld-bead profile. The image was then cropped, reducing its pixel dimensions from 600×400 to 400×160 . Next, the SegFormer model was employed to extract a segmentation map M . In the original image shown in Figure 5a, spatter and flare reflected from the welding arc interfere with the line laser-shaped object. However, as shown in the resulting image M in Figure 5b, these interference elements were effectively removed, and the line laser-shaped object was clearly isolated. Figure 5c presents the results after column thinning, reducing the image to one pixel per column to lower computational load while accurately extracting the feature points.

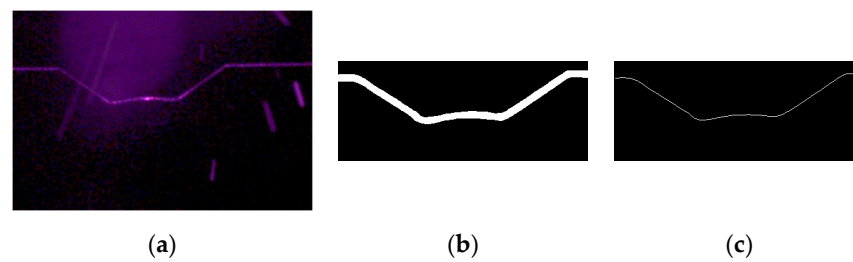


Figure 5. Procedure and results of extracting the weld-bead profile from the line laser pattern objects: (a) captured original image; (b) segmentation map M ; (c) after column thinning image.

A semantic segmentation algorithm was employed to extract the shape of the line laser from the input images. This technique classifies each pixel in an image into specific classes, enabling the model to accurately identify both the shape and location of objects. By assigning every pixel to a class, semantic segmentation provides detailed information about object boundaries and complex shapes, rather than merely detecting the presence or absence of objects. As a result, it enhances the precision of image analysis and is particularly effective for distinguishing objects from intricate backgrounds. This method is particularly well suited for extracting the laser line, even in the presence of noise such as spatter, which can interfere with object detection. During the GMAW process, spatter or interference from the intense arc light can obscure the visibility of the laser line, making it challenging to accurately extract the weld-bead profile. Thus, the model must be trained on labeled data that accounts for various conditions to ensure robust extraction in diverse scenarios. To construct a dataset for semantic segmentation, Roboflow software (version 1.0) [24] was used to label the shapes of line lasers as polygon objects across 1952 images. Figure 6 provides examples of images labeled with polygon annotations for line laser and weld-bead profiles. Polygon labeling involves tracing an object's contours with a series of connected vertices, enabling the precise definition of complex shapes, making it ideal for segmentation tasks. The labeled dataset was then exported in the Common Objects in Context (COCO) JSON format for model training and validation. This widely adopted standard in object detection and segmentation tasks includes detailed information such as coordinates for object boundaries and class labels for each object.

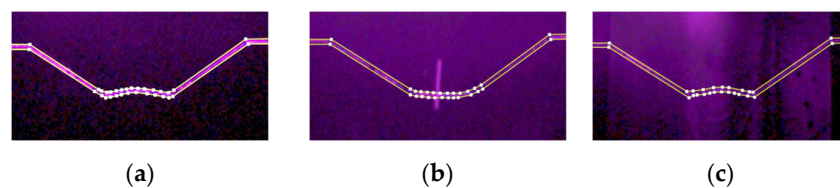


Figure 6. Sample images of polygonal labeling for extracting the weld-bead profile using a line laser during gas metal arc welding: (a) when no spatter occurs; (b) when spatter occurs and overlaps with the line laser pattern; (c) when arc light interferes with the laser line.

2.4. Camera Calibration

The line laser pattern captured by the proposed vision system is initially extracted as a 2D profile. To accurately measure the dimensions and shapes of the weld bead, it is essential to convert this 2D profile into a 3D profile. To achieve this, camera calibration was performed to estimate both the intrinsic and extrinsic parameters of the camera and correct for any geometric distortions. The camera calibration process is described by Equation (5) [25]:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = A[R|t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (5)$$

where A represents the intrinsic parameters of the camera that describe its inherent characteristics. k_u and k_v represent the focal lengths in the x and y directions of the camera, respectively. The coordinates u_0 and v_0 denote the principal point, which is the point on the image sensor where the camera's optical axis intersects the image plane. Parameter s represents the skew coefficient, which indicates the degree of asymmetry when the x and y axes of the image sensor are not perfectly orthogonal. The extrinsic parameters consist of the rotation matrix R and translation vector t , which represent the relationship between the world and camera coordinate systems. Lens distortion must be considered in this projection process, and the distortion correction equations given as Equation (6) are applied using the radial distortion coefficients k_1, k_2, k_3 and the tangential distortion coefficients p_1, p_2 .

$$\begin{aligned} x_{distorted} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + [2p_1 xy + p_2(r^2 + 2x^2)], \\ y_{distorted} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + [p_1(r^2 + 2y^2) + 2p_2 xy], \end{aligned} \quad (6)$$

where r represents the distance from the image center to the corresponding pixel, defined as $= \sqrt{x^2 + y^2}$. With intrinsic camera parameter A , the corrected image coordinates $(x_{distorted}, y_{distorted})$ are transformed into the actual pixel coordinates (u, v) as follows:

$$u = k_u \cdot x_{distorted} + u_0, v = k_v \cdot y_{distorted} + v_0. \quad (7)$$

The intrinsic and extrinsic parameters estimated through the camera calibration process alone are insufficient for accurately converting a 2D profile into 3D world coordinates. This is because multiple points in 3D space are projected onto a single point on the image sensor. However, if the equation of the laser plane created by the line laser in the camera coordinate system is known, it is possible to convert the 2D image coordinates into 3D coordinates. The coefficients of the laser plane equation are calculated using the intrinsic and extrinsic parameters determined during the camera calibration. The laser plane equation, which represents the laser plane projected into 3D space, is given by Equation (8).

$$aX_c + bY_c + cZ_c + d = 0, \quad (8)$$

where X_c, Y_c , and Z_c denote the 3D camera coordinates, and a, b, c , and d are the coefficients of the laser plane. To calculate these coefficients, the line laser was first projected onto the checkerboard, as shown in Figure 7b, with the checkerboard itself depicted in Figure 7a. Next, as illustrated in Figure 7c, the intersection points between the feature lines connecting the corner points of the checkerboard and the laser-projected line were extracted. The 2D normalized image coordinates of these intersection points were computed using the intrinsic parameters. Next, the 3D camera coordinates of the intersection points were determined by applying invariance to the cross ratio. A nonlinear least-squares method was employed to fit the laser plane based on the 3D camera coordinates of each intersection point. Once the coefficients of the laser plane were calculated, validation was conducted using a calibration block with a height difference of 2.5 mm. The accuracy of the calibration was verified by solving the simultaneous equations of the laser plane using Equations (8) and (9) to calculate the 3D camera coordinates. With the calibration results obtained, converting 2D images into 3D coordinates and acquiring 3D information, such as the height and width of the weld bead, becomes possible. The red points represent the intersection points extracted from each calibration image, while the plane illustrates the laser plane calculated using the nonlinear least-squares method.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \\ 1 \end{bmatrix}. \quad (9)$$

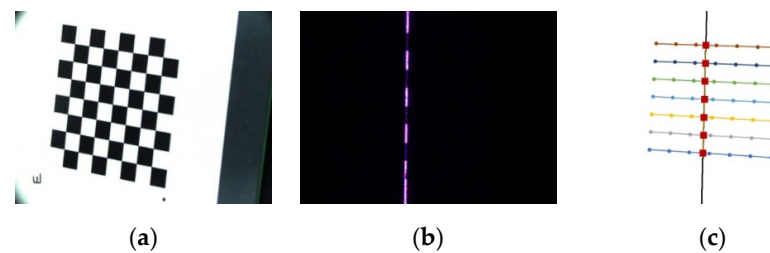


Figure 7. Camera calibration process: (a) checkerboard; (b) line laser projection; (c) extracting corner points and line laser plane.

3. Results and Discussion

3.1. Molten Pool State Classification

With the proposed vision system, molten pool images were captured during the GMAW process and classified using convolutional neural network models to assess molten pool quality. Meeting real-time monitoring performance requirements necessitates a careful selection of deep-learning models and parameters. Table 5 compares the performance of the molten pool classification models, ResNet50 and MobileNetV4, and Figure 8 shows the training accuracy for each model. Both models achieved accuracy above 99%, demonstrating not only the high performance of the deep learning models but also the precision of the proposed vision system's optical design. This accuracy suggests that the system was meticulously calibrated to capture complex molten pool characteristics with high fidelity, including optimal wavelength range, lenses, filters, and focal length, enabling the accurate detection of subtle changes in molten pool states. The high classification accuracy further supports that the captured images were of high quality and largely noise-free, facilitating distortion-free data collection. These results confirm that the vision system effectively captures various molten pool states, consistently providing models with data rich in detail and precision for classification. In conclusion, the proposed vision system reliably reflects changes in molten pool states, producing high-quality images that maintain accuracy and stable classification performance as input for deep learning models. Figure 9 displays a graph of the average inference time for each model. To meet real-time monitoring requirements, MobileNetV4 was chosen owing to its potential for faster inference; however, its average inference time was recorded at 9.06 ms, approximately 39% slower than ResNet50's 6.53 ms. This difference likely arises from MobileNetV4's lightweight structure, which requires additional computations during inference to maintain accuracy. This result underscores the tradeoff between accuracy and inference time. Furthermore, because a GPU accelerator was utilized, ResNet50 demonstrated efficient parallel computation, achieving high accuracy with a faster inference time. Consequently, ResNet50 was confirmed as the most suitable model for the real-time monitoring of molten pool states during the GMAW process.

Table 5. Comparison of the performance of each model when molten pool images are used as input data for the deep learning model to classify the molten pool state.

Test Number	Learning Model	Test Accuracy	Test Loss	Inference Time (ms)
1	ResNet50	0.9976	0.0078	6.53
2	ResNet50	0.9976	0.0081	6.46
3	ResNet50	0.9976	0.0102	6.60
4	MobileNetV4	0.9976	0.0056	9.06
5	MobileNetV4	0.9902	0.0095	8.86
6	MobileNetV4	0.9902	0.2158	9.26

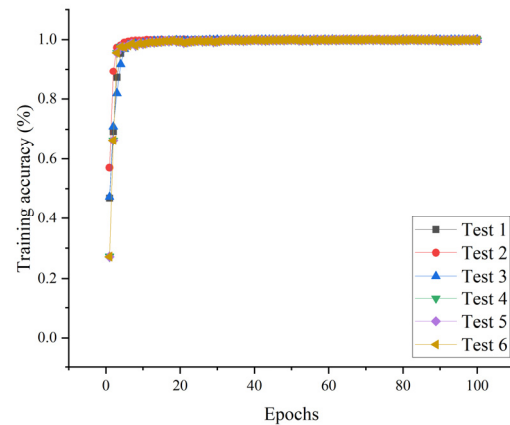


Figure 8. Training accuracy of each model when molten pool images are used as input data for the deep learning model to classify the molten pool state.

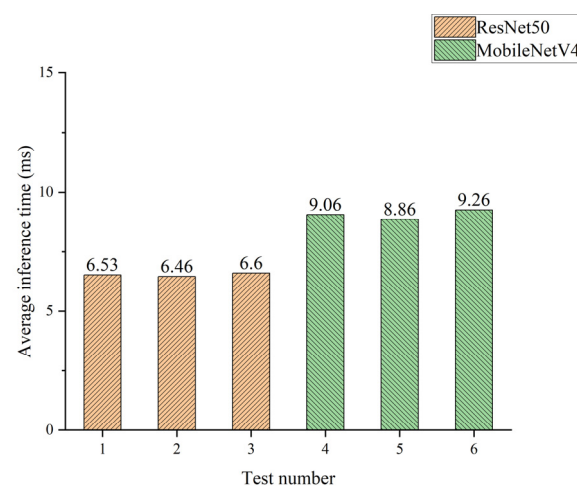


Figure 9. Comparison of the average inference time of each model when molten pool images are used as input data for the deep learning model to classify the molten pool state.

3.2. Weld-Bead Profile Acquisition

The GMAW process was conducted under conditions A, B, and C, as specified in Table 2, with images captured at 30 fps. Due to the varying frequency and intensity of interference factors, such as strong arc light, spatter, and fumes, for each welding condition, 1200 frames from a 40 s recording were selected for consistent image processing performance comparison. Figure 10 shows the line laser shape object captured under each welding condition (A, B, and C). Under the welding voltage of 22.3 V in Figure 10a, the arc length is shorter than the set welding power source voltage of 25.3 V used in the experiment, resulting in deposition transfer closer to the root face of the test plate groove. This condition leads to frequent short circuits and excessive spatter, producing significantly less weld metal compared to other welding voltage conditions, with the spatter interfering with the clear extraction of the line laser-shaped object. As shown in Figure 10b, under the welding voltage of 25.3 V, the shape can be observed to slightly settle toward the groove at approximately 22.90 s, indicating that the metal deposited from the filler wire has excessively transitioned toward the backside of the groove face. In Figure 10c, under the welding voltage of 28.3 V, the arc length increases, causing the arc plasma to concentrate more on the groove walls than on the root. This reduces penetration and results in weld metal deposition on the upper groove area. An analysis of the image sets for each welding condition demonstrated that each set accurately reflected the changes in the line laser shape object based on welding voltage and arc characteristics. These findings suggest that the shape obtained from the line laser during the gas metal arc welding process can be

used as input data to predict weld properties and penetration status. Figure 11 shows the results extracted as M from the captured images for each welding condition, while Figure 12 presents the final output images after column thinning. Under all conditions, the weld-bead profile was extracted without breaks in the line and was free from interference elements, such as arc light, spatter, and fumes. Additionally, the profile was robustly extracted, accommodating size variations (e.g., height and width) and adapting well to shape variations, including flat, convex, and concave profiles.

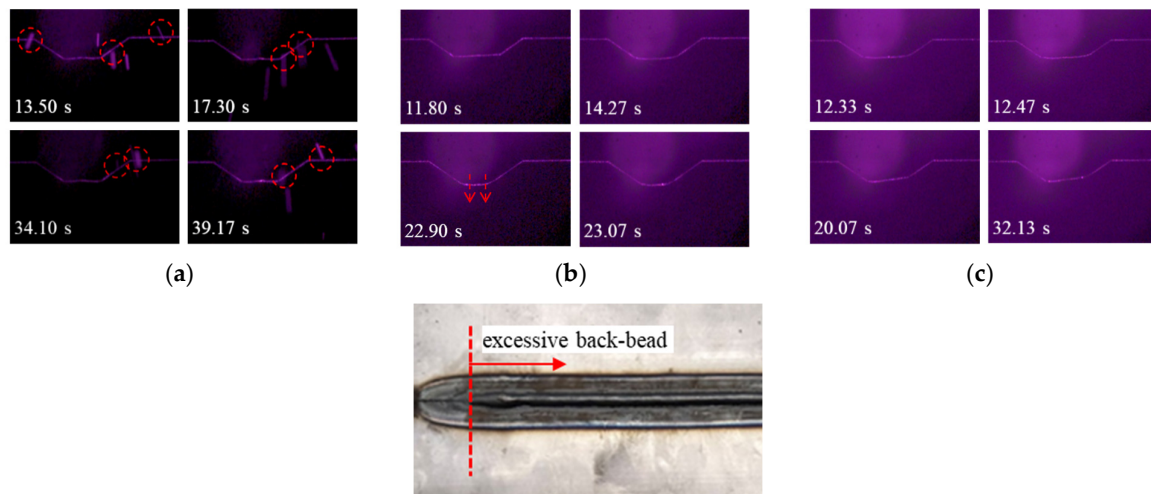


Figure 10. Object images of line laser captured at each specified time during GMAW process: (a) condition A; (b) condition B; (c) condition C. Red circles indicate areas where spatter interferes with line laser objects. (The red arrow in (b) represents the profile of the molten metal as it passes through the test plate and emerges at the back bead).

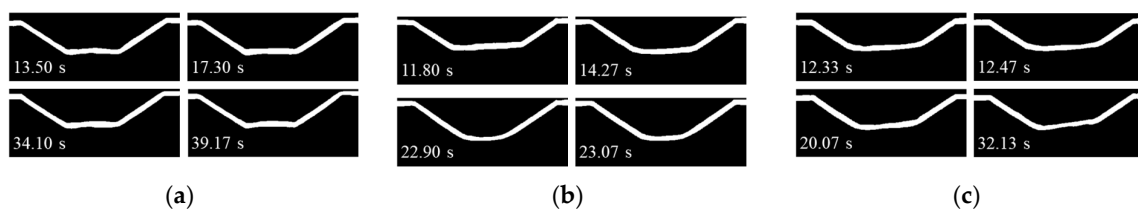


Figure 11. Segmentation map M images extracted at each specified time during GMAW process: (a) condition A; (b) condition B; (c) condition C.

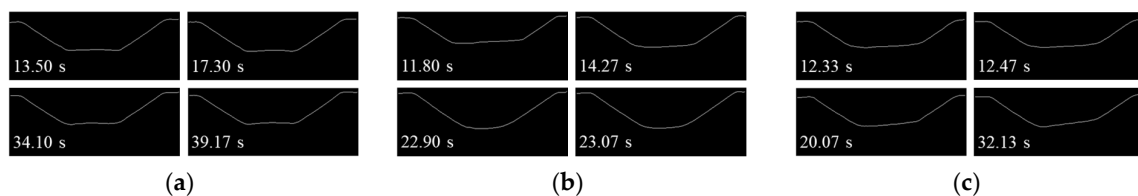


Figure 12. Post-processing images extracted at each specified time during GMAW process: (a) condition A; (b) condition B; (c) condition C.

In the following section, we assess whether the proposed integrated image processing deep learning model achieves a processing speed suitable for real-time monitoring. To enhance processing speed, cropping and pruning techniques were applied. Cropping involves defining an ROI in the original image of the line laser-shaped object to remove unnecessary areas. As described earlier, the image was cropped from its original size of 600×400 pixels to 400×160 pixels. Pruning reduces the number of parameters in the deep learning model, decreases memory usage, and improves inference time. However,

selecting an appropriate pruning ratio is crucial, as excessive pruning can degrade model performance. In this study, L1 unstructured pruning was applied, which sets the weights with the smallest L1 norm in the Conv2D layer to zero. The L1 norm is the sum of the absolute values of the weights. This approach preserves important weights while removing less significant ones, making the model lighter and enhancing inference time. The Conv2D layer is used in MiT for patch embedding and to generate multi-resolution feature maps. Table 6 compares the inference quality of the weld-bead profile at pruning ratios ranging from 10% to 60%, while Table 7 lists the inference time and the number of images processed per second. A total of 1328 images were used for inference. At 10% pruning, despite the relatively low pruning ratio, the model failed to accurately extract the weld-bead profile from 41 images, indicating that removing some weights negatively impacted performance. At 20% and 30% pruning, the results were favorable, with no broken or distorted profile lines. The machined surface of the welding test plate was accurately represented by straight lines, and the weld beads were correctly inferred to have curved shapes. After 40% pruning, one image could not be inferred correctly; however, the overall performance remained satisfactory. In contrast, at 50% pruning, the model failed on 71 images, and at 60% pruning, it failed on all images. In conclusion, the model achieved the best balance between lightweight design and performance at pruning ratios of 20% to 30%. It delivered high-quality image inference and improved processing efficiency.

Table 6. Comparison of the inference quality of the weld-bead profile according to the pruning ratio.

Pruning Ratio (%)	Example Image		Failed Image (ea)	Result
	A	B		
10			41	No good
20			0	Good
30			0	Good
40			1	Good
50			72	No good
60			1328	No good

Table 7. Comparison of the inference time and the number of images that can be inferred per second according to the pruning ratio of the weld-bead profile.

Pruning Ratio (%)	Average Inference Time (ms)	Image Per Second (ea)
10	16.11	62
20	16.30	61
30	16.20	61
40	16.23	61
50	16.16	61
60	16.14	61

3.3. Camera Calibration

Table 8 lists the camera calibration results, and Figure 13 visualizes them in 3D. The red points represent the intersection points extracted from each calibration image, while the plane shows the laser plane calculated using the nonlinear least-squares method. The intrinsic camera parameters were estimated as follows: $k_u = 4665.10034$, $k_v = 4621.14792$, $u_0 = 1042.78592$, and $v_0 = 841.17693$. The radial distortion coefficients were calculated as $k_1 = 0.29379$, $k_2 = -17.71131$, and $k_3 = 0$. The tangential distortion coefficients were estimated as $p_1 = 0.00752$ and $p_2 = -0.00194$. The coefficients of the laser plane equation were estimated as $a = 3.47926$, $b = 0.21426$, $c = -3.65155$, and $d = 1247.85$. With the calibrated results obtained through these steps, the 2D image coordinates were converted into 3D world coordinates, enabling the extraction of precise measurements such as the height and width of the top bead.

Table 8. Results of camera calibration.

Intrinsic parameter (A)	$A = \begin{bmatrix} 4665.10034 & 0 & 1042.78592 \\ 0 & 4621.14792 & 841.17693 \\ 0 & 0 & 1 \end{bmatrix}$
Distortion coefficients	$[0.29379 \quad -17.71131 \quad 0.00752 - 0.00194 \quad 0]$
Laser plane equation	$3.47926X_c + 0.21426Y_c - 3.65155Z_c + 1247.85 = 0$
Validation value	2.4493, 2.4488, 2.4409, 2.4742, 2.5090, 2.5667, 2.4412
Mean error	45.9 μm

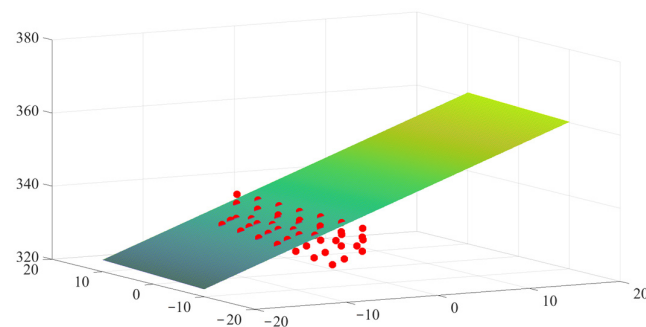


Figure 13. Three-dimensional visualization of calibration results.

4. Conclusions

In this study, we propose a deep learning-based vision system capable of the real-time observation and monitoring of the molten pool and weld-bead profile during GMAW. The proposed system effectively addresses the interference from intense arc light and overcomes the limitations of offline inspection methods that rely on conventional laser vision sensors. With the integration of an optimized optical design with the simultaneous use of 2D and 3D information, the system enables the real-time acquisition of molten pool images and bead profiles. It incorporates a vision camera and a line laser, allowing concurrent observation of the molten pool and laser line during the GMAW process. To extract accurate 3D information from the laser-projected profile of the weld bead, a deep learning-based semantic segmentation model was applied. The experimental results show that the molten pool classification models using ResNet50 and MobileNetV4 achieved high accuracy. Notably, ResNet50 demonstrated a rapid inference time of 6.53 ms and an accuracy of 99.76%, meeting the real-time monitoring requirements. MobileNetV4, while achieving 99.02% accuracy, exhibited a slightly slower inference time of 9.06 ms compared to ResNet50. The calibration of the proposed vision system yielded a mean error of 45.9 μm , validating the system's accuracy in 3D coordinate extraction from 2D images. The proposed segmentation model for extracting the weld-bead profile showed consistent performance

across various welding conditions, effectively removing interference elements such as spatter and arc light. By applying ROI cropping and L1 unstructured pruning to optimize the inference time and lightweight design of the model, we achieved optimal performance at pruning ratios of 20–30%, resulting in a processing speed of 61 fps while maintaining high-quality image inference.

These results demonstrate that the proposed vision system meets both the processing speed and accuracy requirements for real-time monitoring, highlighting its strong potential as a robust tool for automated weld quality management in various welding environments. Future research will involve integrating the proposed vision system with collaborative robots to establish a solution capable of monitoring and controlling welding quality in confined spaces, where challenges arise from limited access or complex weldment geometries. Enhancements to the model will focus on expanding the dataset to encompass a broader spectrum of welding environments and process conditions, aiming to develop a more precise and reliable monitoring system.

Author Contributions: Conceptualization, G.-G.K. and J.Y.; investigation, G.-G.K. and D.-Y.K.; methodology, G.-G.K.; data curation, G.-G.K. and D.-Y.K.; validation, G.-G.K. and D.-Y.K.; formal analysis, G.-G.K. and D.-Y.K.; writing—original draft preparation, G.-G.K.; writing—review and editing, G.-G.K. and J.Y.; visualization, G.-G.K.; supervision, J.Y.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Material Parts Technology Development Project (20022489, Development and Substantiation of Joining Equipment & Smart Joining Lines for Manufacturing Electric Vehicle Chassis and Battery Cases with High-Strength Materials) funded by the “Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea)” and “Korea Planning & Evaluation Institute of Industrial Technology (KEIT)”.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Huang, W.; Kovacevic, R. A Laser-Based Vision System for Weld Quality Inspection. *Sensors* **2011**, *11*, 506–521. [\[CrossRef\]](#)
- Huang, W.; Kovacevic, R. Development of a Real-Time Laser-Based Machine Vision System to Monitor and Control Welding Processes. *Int. J. Adv. Manuf. Technol.* **2012**, *63*, 235–248. [\[CrossRef\]](#)
- Han, Y.; Fan, J.; Yang, X. A Structured Light Vision Sensor for On-Line Weld Bead Measurement and Weld Quality Inspection. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 2065–2078. [\[CrossRef\]](#)
- Chu, H.-H.; Wang, Z.-Y. A Vision-Based System for Post-Welding Quality Measurement and Defect Detection. *Int. J. Adv. Manuf. Technol.* **2016**, *86*, 3007–3014. [\[CrossRef\]](#)
- Li, Y.; Li, Y.F.; Wang, Q.L.; Xu, D.; Tan, M. Measurement and Defect Detection of the Weld Bead Based on Online Vision Inspection. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 1841–1849. [\[CrossRef\]](#)
- Kim, J.-W.; Lee, J.-Y. A Control System for Uniform Bead in Fillet Arc Welding on Tack Welds. *J. Mech. Sci. Technol.* **2008**, *22*, 1520–1526. [\[CrossRef\]](#)
- Ye, G.; Guo, J.; Sun, Z.; Li, C.; Zhong, S. Weld Bead Recognition Using Laser Vision with Model-Based Classification. *Robot. Comput.-Integr. Manuf.* **2018**, *52*, 9–16. [\[CrossRef\]](#)
- Lee, K.; Hwang, I.; Kim, Y.-M.; Lee, H.; Kang, M.; Yu, J. Real-Time Weld Quality Prediction Using a Laser Vision Sensor in a Lap Fillet Joint during Gas Metal Arc Welding. *Sensors* **2020**, *20*, 1625. [\[CrossRef\]](#)
- Wang, J.; Zhang, Z.; Bai, Z.; Zhang, S.; Qin, R.; Huang, J.; Wen, G. On-Line Defect Recognition of MIG Lap Welding for Stainless Steel Sheet Based on Weld Image and CMT Voltage: Feature Fusion and Attention Weights Visualization. *J. Manuf. Process.* **2023**, *108*, 430–444. [\[CrossRef\]](#)
- Yu, R.; Han, J.; Zhao, Z.; Bai, L. Real-Time Prediction of Welding Penetration Mode and Depth Based on Visual Characteristics of Weld Pool in GMAW Process. *IEEE Access* **2020**, *8*, 81564–81573. [\[CrossRef\]](#)
- Chen, C.; Lv, N.; Chen, S. Welding Penetration Monitoring for Pulsed GTAW Using Visual Sensor Based on AAM and Random Forests. *J. Manuf. Process.* **2021**, *63*, 152–162. [\[CrossRef\]](#)
- Xia, C.; Pan, Z.; Fei, Z.; Zhang, S.; Li, H. Vision Based Defects Detection for Keyhole TIG Welding Using Deep Learning with Visual Explanation. *J. Manuf. Process.* **2020**, *56*, 845–855. [\[CrossRef\]](#)
- Zhang, B.; Shi, Y.; Cui, Y.; Wang, Z.; Hong, X. Prediction of Keyhole TIG Weld Penetration Based on High-Dynamic Range Imaging. *J. Manuf. Process.* **2021**, *63*, 179–190. [\[CrossRef\]](#)

14. Wang, Z.; Shi, Y.; Hong, X.; Zhang, B.; Chen, X.; Zhan, A. Weld Pool and Keyhole Geometric Feature Extraction in K-TIG Welding with a Gradual Gap Based on an Improved HDR Algorithm. *J. Manuf. Process.* **2022**, *73*, 409–427. [[CrossRef](#)]
15. Lu, J.; Shi, Y.; Bai, L.; Zhao, Z.; Han, J. Collaborative and Quantitative Prediction for Reinforcement and Penetration Depth of Weld Bead Based on Molten Pool Image and Deep Residual Network. *IEEE Access* **2020**, *8*, 126138–126148. [[CrossRef](#)]
16. Xu, X.; Wang, Y.; Han, J.; Lu, J.; Zhao, Z. Online Monitoring and Control of Butt-Welded Joint Penetration during GMAW. *Metals* **2022**, *12*, 2009. [[CrossRef](#)]
17. Nomura, K.; Fukushima, K.; Matsumura, T.; Asai, S. Burn-through Prediction and Weld Depth Estimation by Deep Learning Model Monitoring the Molten Pool in Gas Metal Arc Welding with Gap Fluctuation. *J. Manuf. Process.* **2021**, *61*, 590–600. [[CrossRef](#)]
18. ISO 630:1995; Structural Steels—Plates, Wide Flats, Bars, Sections and Profiles. ISO: Geneva, Switzerland, 1995.
19. ISO 636:2017; Welding Consumables—Rods, Wires and Deposits for Tungsten Inert Gas Welding of Non-Alloy and Fine-Grain Steels—Classification. ISO: Geneva, Switzerland, 2017.
20. Kim, G.-G.; Kim, Y.-M.; Kim, D.-Y.; Park, J.-K.; Park, J.; Yu, J. Vision-Based Online Molten Pool Image Acquisition and Assessment for Quality Monitoring in Gas–Metal Arc Welding. *Appl. Sci.* **2024**, *14*, 5998. [[CrossRef](#)]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Qin, D.; Lechner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B.; et al. MobileNetV4—Universal Models for the Mobile Ecosystem. *arXiv* **2024**, arXiv:2404.10518. [[CrossRef](#)]
23. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–14 December 2021; Curran Associates, Inc.: New York, NY, USA, 2021; Volume 34, pp. 12077–12090.
24. Dwyer, B.; Nelson, J.; Solawetz, J. Roboflow (Version 1.0) [Software]. 2024. Available online: www.roboflow.com (accessed on 19 September 2024).
25. Gan, Z.; Tang, Q. *Visual Sensing and Its Applications: Integration of Laser Sensors to Industrial Robots*; Advanced Topics in Science and Technology in China; Zhejiang University Press: Hangzhou, China; Springer: Heidelberg, Germany, 2011; ISBN 978-3-642-18287-7.

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.