

# Embedded vision system for monitoring arc welding with thermal imaging and deep learning

Andrea Fernández, Álvaro Souto, Carlos González, Roi Méndez-Rial  
Robotics and Control Unit - Aimen Technology Centre, Porriño, Spain,  
Email: {andrea.fernandez, alvaro.souto, carlos.gonzalez, roi.mendez}@aimen.es

**Abstract**—We develop a novel embedded vision system for online monitoring of arc welding with thermal imaging. The thermal images are able to provide clear information of the melt pool and surrounding areas during the welding process. We propose a deep learning processing pipeline with a CNN-LSTM architecture for the detection and classification of defects based on video sequences. The experimental results show that the CNN-LSTM architecture is able to model the complex dynamics of the welding process and detect and classify defects with high accuracy. In addition, the embedded vision system implements an OPC-UA server, enabling an easy vertical and horizontal integration in Industry 4.0.

**Index Terms**—Arc welding, embedded vision, deep learning, convolutional neural networks, Industry 4.0.

## I. INTRODUCTION

Arc welding has been used for years in many industrial applications and it is implemented in most automotive, aerospace and energy industries [1]. The introduction of robot-based welding has led to an increased efficiency of the welding processes, enabling the manufacturing of more parts in less time, while minimizing scrap, increasing quality and improving the working environment [2].

On the other hand, the complexity of the physics of the welding process (non-linear and strongly coupled inputs, e.g. gas flux, welding intensity, speed material composition, arc length, etc.) and the robotised welding equipment (composed of different parts and machines that interact with each other), make it challenging to develop automated solutions for flexible manufacturing lines with higher product quality, which are essential for responding to the dynamic behavior of the market.

The main limiting factor regarding manufacturing productivity is the incidence of weld defects, usually at high currents and welding speeds [3]. Besides, a sub-optimal utilization of welding machines implies expensive losses in production, e.g. welding machines with erratic arcs also waste great amount of energy. Besides, due to their long life and durability, routine maintenance is not commonly applied producing unexpected breakdowns and consequently production losses.

A great effort in research has been done in the last years regarding the development of real-time monitoring system for in-process quality control and detection of weld defects during the robotic welding process. Different sensing technologies

have been proposed: electro-optic sensors in the IR/UV range [4], spectroscopy solutions [5] [1], and vision systems based on CCD cameras [6], CMOS NIR [7] and thermal imaging [8] [9] [10] [3]. In these works, automatic identification and classification of weld defects have been performed by means of the traditional approach of feature selection and modeling, using for instance statistical estimation, principal component analysis or artificial neural networks.

In the last years, deep learning have brought revolutionary advances in computer vision and machine learning, outperforming traditional machine learning approaches in many application fields. Recently, vision systems based on Convolutional Neural Networks (CNNs) [11] have also been proposed for quality assessment of arc welding [12], [13]. In both works, CNNs are able to successfully classify welds based on visible images. However, the proposed defects classification models are limited to surface defects and do not exploit the temporal information (process dynamics) and other type of information that might be provided by alternative sensing technologies, e.g. thermal imaging or optical spectroscopy.

In this work, a machine vision system based on thermal infrared imaging for on-line condition monitoring and detection of weld defects in arc-welding processes using CNNs is proposed. The main contributions of this work can be summarized as:

- An edge machine vision system compliant with Industry 4.0 standards (e.g. OPC-UA, a standardized communication protocol for industrial automation) enabling an easy and seamless vertical and horizontal integration. The embedded systems provide a real-time processing pipeline and inference engine to process continuous streams of input data in real time.
- A methodology for proper monitoring and imaging the welding pool during the manufacturing of welding parts in robotised industrial cells. We use a thermal infrared camera sensitive in the long-wavelength infrared range (LWIR) to have a clear vision of the molten pool with two different configurations: a fixed camera position attached to the robot arm and a fixed camera setup.
- A Deep Neural Network combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory-based Networks (LSTMs) [14] is proposed to identify defects in real-time. The CNN-LSTM network architecture is specifically designed for sequence prediction problems

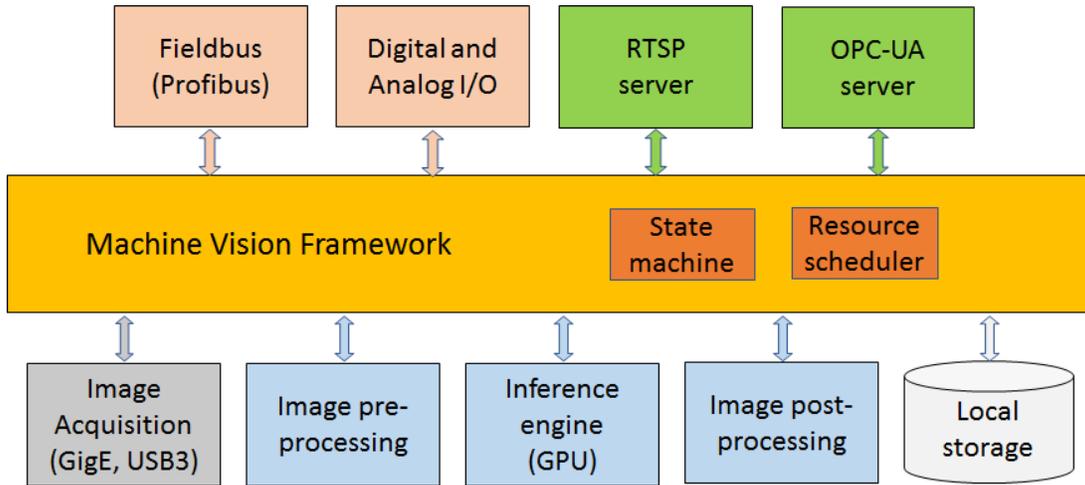


Fig. 1. Embedded machine vision system architecture.

with spatial inputs, like images or videos.

This paper is organized as follows. Section II describes the Methodology applied in this work, including; i) detailed description of the embedded vision system, ii) the experimental setup, iii) the deep neural network architecture. Section III summarizes the results obtained and Section IV draws the main conclusions of our research work.

## II. METHODOLOGY

### A. Embedded Vision System

We have developed a generic machine vision (MV) edge solution ensuring interoperability and easy integration with all the hierarchical levels of the Factory 4.0, e.g. Control Devices, Station and Enterprise, through the use of the open source global accepted standard OPC-UA [15]. The conceptual model of the VDMA OPC-UA vision companion specification has been used as a reference, although the information model has not been exactly implemented as described in the specification. Figure 1 gives a general overview of the system architecture. The bottom layer represents all the inner resources of the vision system, including interfaces to gather data from industrial vision cameras via common vision standards (i.e. GigE, USB3 vision). The resource scheduler manages the control and synchronizes the acquisition of the different resources (field data, configuration files, image pre-processing, AI inference engine, post-processing and local storage). The MV framework implements the logic and state machine of the system and coordinates the acquisition and processing modules. The top layer represents the communication with the outside world. The system is exposed through an OPC-UA interface, providing functionalities for control and configuration, and sharing the machine vision results. In addition, it also provides various interfaces for communicating with other shopfloor devices through field buses (e.g. Profibus) and digital analog I/O signals. Finally, an RTSP allows the operator to see the raw or post-processed video sequence in real-time.

The following programming libraries have been used for the implementation: Aravis (open source based library for video acquisition using Genicam cameras), Hilsher driver for controlling the field bus adapter, Free OPC-UA Library (used for the implementation of the OPC-UA server), and OpenCV and TensorRT for image processing and the implementation of the deep learning inference module. The embedded vision framework is compatible with the powerful family of embedded SoC boards from NVIDIA: Jetson TX1, TX2 and Xavier.

### B. Experimental setup

The proposed vision system has been used to monitor the arc welding process in a robotized welding cell equipped with an ABB welding robot and CMTi FRONIUS welding equipment. To monitor the arc welding process a thermal infrared camera with spectral response in the range of  $8 - 14\mu\text{m}$  and spatial resolution of  $640 \times 480$  pixels has been used, the Xenics Gobi-640-GigE. To obtain a clear IR image of a molten pool and avoid image saturation and lens damage from spatter, a narrow band-pass filter together with a Ge protection window were placed ahead of the camera lens.

Since the level of complexity of a robotised welding system in a production environment may be large, e.g. twin arc welding including work pieces manipulators, it is still challenging to effectively and efficiently apply online sensing techniques in industrial applications. The high temperature, intense light from arc, fume, high current, molten metal, spatter and accessibility to avoid collisions with the fixturing systems and the welding torch, are factors that need to be considered when designing an online monitoring system.

To cope with the diversity of complexity in industrial production welding systems, two different camera setups were considered: i) moving camera configuration attached to the robot arm ii) fixed camera position to monitor the process with panoramic vision. In the moving camera configuration, the camera is attached to the robot arm pointing to the melt pool in  $45^\circ$ . In the fixed camera setup, the imager was placed

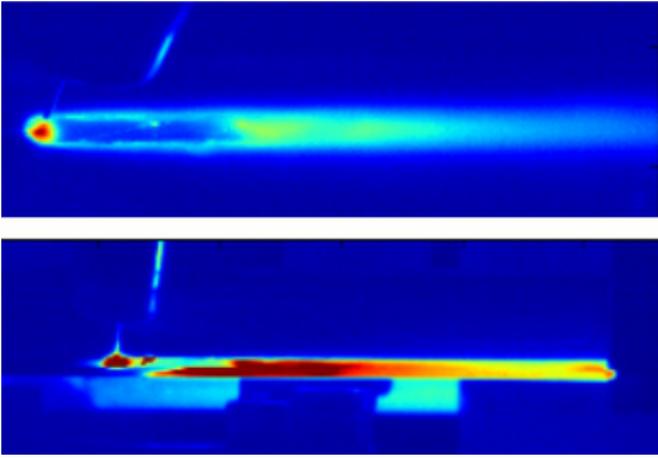


Fig. 2. Thermal image samples with moving camera attached to the robot arm (top), and fixed camera setup (bottom).

in a fixed frame pointing to the weld track perpendicularly, with a field of view covering the entire welding track. While the first configuration is preferred because it provides a higher resolution imaging of the molten pool, its deployment in a robotised welding production cell is not always feasible due to restricted accessibility. The second configuration was proposed as an alternative solution for complex configurations where attaching a camera to the robot arm is not feasible due to collisions.

Two different welding joints were considered, namely butt welding and lap joint welding, to connect two pieces of aluminium together. During the trials, several parts were manufactured for each joint type. Some weld defects and flaw specimens were deliberately produced to train the classification models. The list of intentional defects produced include: lack of fusion, lack of penetration, undercut and porosity. Those were produced using aggressive weld parameters and worn out nozzles. Figure 3 shows some butt joints and lap joints welding samples with and without defects.

Two datasets considering thermal video sequences were built to train the classification model. Additional relevant parameters were also collected from the welding equipment, such as arc-voltage, current and robot speed.

- Dataset I: contains data generated during the manufacturing of 30 butt joint welding parts using the moving camera setup, including good and defected samples. Each image of the video sequences in the dataset was labeled as good or defected joint.
- Dataset II: contains data generated with the lap joint welding and fixed camera configuration, including 30 video sequences with good samples and defected parts with undercut and localised porous. Each image of the video sequence was labeled as good, defect type I and defect type II.



Fig. 3. Weldings samples with and without defects. Top three correspond to dataset I, bottom three to dataset II.

### C. CNN-LSTM Network

In this research work, we propose a classifier based on a CNN and a LSTM-based network for online welding quality assessment.

Due to its intrinsic spatial and temporal nature, this type of networks can be applied to a variety of vision tasks involving sequential inputs and outputs [14]. Our proposed architecture consists of three stages: i) a CNN for the extraction of spatial features, ii) a LSTM-based network for the extraction of temporal features, and iii) a fully connected layer (FC) for the classification of the spatiotemporal features extracted. The detailed explanation of each stage is presented in the following subsections.

1) *Convolutional Neural Network*: The first stage of our proposed architecture comprises a Resnet18 [16] model as the CNN for the extraction of spatial features of each input image. This network consists of a first convolutional layer followed by a max pooling layer, 8 residual units, an average pooling layer and a remaining fully connected (FC) layer with a softmax function used for pre-training purposes. Each residual unit consists of two blocks of the following operators:

- A Batch Normalization (BN) layer, which is a regularization technique to accelerate training and reduce overfitting.
- An activation layer with a ReLU function that add non-linearities to the model
- A convolutional layer with a filter size of  $3 \times 3$  that performs the feature operations on the image.

According to the type of shortcut connectivity within the residual units, they can be categorized as  $(1 \times 1)$  convolutional units or identity units. These connections, characteristic of ResNet models, solve the saturation problem of very deep networks and improve its learning capabilities. All input images have a resolution of  $64 \times 64$  pixels.

During the training phase, the CNN is used as an image "encoder" by first pre-training it for the image classification task. After completing this process, the output vector of the last

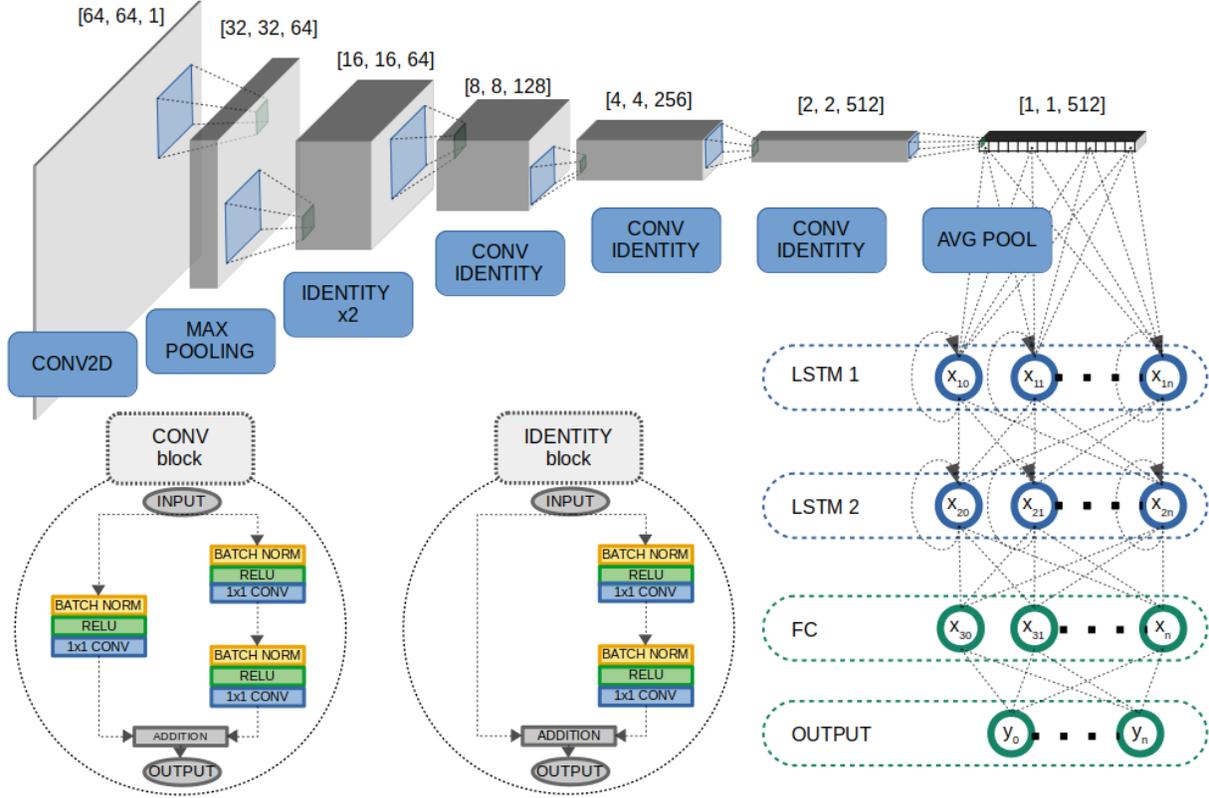


Fig. 4. CNN-LSTM network architecture.

average pooling layer is selected for further temporal feature extraction and classification.

2) *Long Short-Term Memory-based Network*: In the second stage, the LSTM-based network, a variant of a Recurrent Neural Network, is used to extract the temporal characteristics of the feature vectors extracted in the first stage. This network consists of two hidden layers with 4 and 2 units respectively. The output vector of this network contains both spatial and temporal information and is used for the final classification.

3) *Classification*: The last classification stage follows the LSTM-based network and consists of a FC layer with a SoftMax function that returns a probability and assigns it to one of the classes considered, which in our case represent the type of defects.

To summarize, the architecture proposed assigns a final class label representing the "state" of a given sequence of successive images from an arc welding recording by exploiting both their spatial and temporal information. Figure 4 shows the complete CNN-LSTM architecture.

### III. EXPERIMENTAL RESULTS

Two experiments were performed to evaluate the performance of the vision system and the proposed CNN-LSTM network architecture for defect classification using dataset I and dataset II, each corresponding to butt welding and lap joint welding with the fixed and moving camera setup respectively.

The CNN part of the models was first pre-trained for solving a binary classification problem with dataset I (i.e. 0-no defect, 1-defect) and a multi-class classification problem with dataset II (i.e. 0-no defect, 1-defect type I, 2-defect type II) in an image by image basis, minimizing the categorical cross entropy loss function. Finally, the last layers of the architecture were replaced by two LSTM layers with an input sequence length of 15 frames. The models were jointly trained with the same datasets to learn the temporal dynamics and convolutional perceptual representations to improve the classification.

In the first experiment, we split the dataset I in training and validation sets with 90% and 10% of the samples respectively. Table I and Table II show the confusion matrix and the classification metrics achieved in the validation set. An overall

accuracy of 98.9% was obtained.

TABLE I  
NORMALIZED CONFUSION MATRIX WITH CNN-LSTM IN THE  
VALIDATION SET (DATASET I).

	Good	Defected
Good	0.997	0.002
Defected	0.022	0.978

In the second experiment, we split dataset II into training, validation and test sets with 65%, 25% and 10% of the samples respectively to have an independent test set for the final evaluation of the model. In order to assess the impact of the LSTM layers in the classification performance, we show the results obtained using only the CNN part of the architecture (i.e. ResNet18) and the final results achieved with the complete CNN-LSTM model. Table III and Table IV show the confusion matrices and the classification metrics achieved in the test set.

Overall, the results show that the proposed CNN-LSTM network is able to properly classify good and defected welds with high accuracy. In the second experiment, we compare the performance of the CNN network and the complete CNN-LSTM architecture in a more challenging 3-fold classification problem. While both models provide good overall results, the ResNet18 fails to classify the samples labeled as defect type II. This can be due to the reduced number of samples used in the training and the difficulty to capture the process dynamics causing the defect in a image by image basis. In this regard, a significant improvement is achieved with the CNN-LSTM. Although more experiments and much larger datasets will be needed to further train and validate the robustness of the model, the proposed CNN-LSTM architecture shows promising results and seems capable of jointly learn the complex temporal dynamics of the arc welding process and the convolutional perceptual representations.

#### IV. CONCLUSION

A novel IR embedded vision system has been developed for monitoring the arc welding process in industrial robotised cells. Infrared imaging in the LWIR provides clear information

TABLE II  
CLASSIFICATION METRICS WITH CNN AND CNN-LSTM IN THE  
VALIDATION SET (DATASET I).

Accuracy	Precision	Recall	F1
0.989	0.997	0.984	0.990

TABLE III  
NORMALIZED CONFUSION MATRICES WITH CNN AND CNN-LSTM IN  
THE TEST SET (DATASET II).

	CNN			CNN-LSTM		
	Good	Defect I	Defect II	Good	Defect I	Defect II
Good	0.952	0	0.612	0.972	0	0.25
Defect I	0	1	0	0	1	0
Defect II	0.053	0	0.313	0.039	0	0.64

TABLE IV  
CLASSIFICATION METRICS WITH CNN AND CNN-LSTM IN THE TEST SET  
(DATASET II).

	CNN		
	Precision	Recall	F1
Good	0.95	0.95	0.95
Defect I	1	1	1
Defect II	0.34	0.31	0.33
Accuracy	-	-	0.94
Macro avg.	0.76	0.76	0.75
Weighted avg.	0.94	0.94	0.94
	CNN-LSTM		
	Precision	Recall	F1
Good	0.96	0.97	0.97
Defect I	1	1	1
Defect II	0.72	0.64	0.68
Accuracy	-	-	0.95
Macro avg.	0.89	0.87	0.88
Weighted avg.	0.95	0.95	0.95

of the melt pool and surrounding areas, such as geometry and temperature distribution and cooling profiles, the contact tip temperature or gap. The information can be exploited for process monitoring with two aims: condition monitoring of the welding equipment and quality check of the manufactured part. A CNN-LTSM has been developed to identify weld defects and classify the samples as good or defected with high accuracy. The results show that the CNN-LTSM is capable of exploiting the complex temporal dynamics of the process to improve the classification accuracy with respect to a standard CNN using only spatial information.

#### REFERENCES

- [1] J. Mirapeix, P. B. García-Allende, A. Cobo, O. M. Conde, and J. M. López-Higuera, "Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks," *NDT&E International*, vol. 40, no. 4, pp. 315–323, 2007.
- [2] T. S. Hong, M. Ghobakhloo, and W. Khaksar, "Robotic Welding Technology," *Comprehensive Materials Processing*, vol. 6, no. February, pp. 77–99, 2014.
- [3] S. C. Alfaro, J. A. R. Vargas, G. C. De Carvalho, and G. G. De Souza, "Characterization of "humping" in the GTA welding process using infrared images," *Journal of Materials Processing Technology*, vol. 223, pp. 216–224, 2015.
- [4] P. Sforza and D. De Blasiis, "On-line optical monitoring system for arc welding," *NDT&E International*, vol. 35, no. 1, pp. 37–43, 2002.
- [5] J. Mirapeix, A. Cobo, O. M. Conde, C. Jaúregui, and J. M. López-Higuera, "Real-time arc welding defect detection technique by means of plasma spectrum optical analysis," *NDT&E International*, vol. 39, no. 5, pp. 356–360, 2006.
- [6] J. J. Wang, T. Lin, and S. B. Chen, "Obtaining weld pool vision information during aluminium alloy TIG welding," *International Journal of Advanced Manufacturing Technology*, vol. 26, no. 3, pp. 219–227, 2005.
- [7] X. Gao, D. You, and S. Katayama, "Seam tracking monitoring based on adaptive Kalman filter embedded elman neural network during high-power fiber laser welding," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4315–4325, 2012.
- [8] P. Ghanty, M. Vasudevan, D. P. Mukherjee, N. R. Pal, N. Chandrasekhar, V. Maduraimuthu, A. K. Bhaduri, P. Barat, and B. Raj, "Artificial neural network approach for estimating weld bead width and depth of penetration from infrared thermal image of weld pool," *Science and Technology of Welding and Joining*, vol. 13, no. 4, pp. 395–401, 2008.
- [9] M. Vasudevan, N. Chandrasekhar, V. Maduraimuthu, A. K. Bhaduri, and B. Raj, "Real-time monitoring of weld pool during GTAW using infrared thermography and analysis of infra-red thermal images," *Welding in the World*, vol. 55, no. 7-8, pp. 83–89, 2011.

- [10] S. Chokkalingham, N. Chandrasekhar, and M. Vasudevan, "Predicting the depth of penetration and weld bead width from the infra red thermal image of the weld pool using artificial neural network modeling," *Journal of Intelligent Manufacturing*, vol. 23, no. 5, pp. 1995–2001, 2012.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Neural Information Processing Systems*, 2012, vol. 25, pp. 1097–1105.
- [12] P. Zheng, J. Chen, S. Ye, P. Ott, and L. Wang, "Classification of arc welding joints images based on convolutional neural network," in *IEEE International Conference on Anti-Counterfeiting, Security and Identification, ASID*, 2018, pp. 31–34.
- [13] Z. Zhang, G. Wen, and S. Chen, "Weld image deep learning-based on-line defects detection using convolutional neural networks for Al alloy in robotic arc welding," *Journal of Manufacturing Processes*, vol. 45, no. March, pp. 208–216, 2019.
- [14] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, 2017.
- [15] D. Strutzenberger, T. Frühwirth, T. Trautner, R. Hinterbichler, and F. Pauker, "Communication interface specification in OPC UA," in *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019, pp. 1329–1332.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.