Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation

Chunyang Xia
*University of Wollongong, cx772@uowmail.edu.au*

Zengxi Stephen Pan
*University of Wollongong, zengxi@uow.edu.au*

Zhenyu Fei
*University of Wollongong*

Shiyu Zhang
*University of Wollongong, sz946@uowmail.edu.au*

Hui Jun Li
*University of Wollongong, huijun@uow.edu.au*

Follow this and additional works at: [https://ro.uow.edu.au/eispapers1](https://ro.uow.edu.au/eispapers1)

Part of the Engineering Commons, and the Science and Technology Studies Commons

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation

Abstract
© 2020 The Society of Manufacturing Engineers As an advanced and highly efficient welding method, Keyhole Tungsten Inert Gas (keyhole TIG) welding has drawn wide interests from the manufacturing industry. In order to improve its manufacturing quality and automation level, it’s necessary to develop an online monitoring system for the keyhole TIG welding process. This study developed a visual monitoring system, which utilized an HDR welding camera to monitor the welding pool and keyhole during keyhole TIG welding process. A state of the art Convolutional neural network (Resnet) was developed to recognize different welding states, including good weld, incomplete penetration, burn through, misalignment and undercut. In order to improve the diversity of training dataset, image augmentation was performed. To optimize the training process, a metric learning strategy of center loss was introduced. Furthermore, visualization methods, including guided Grad-CAM, feature map and t-SNE were applied to understand and explain the effectiveness of deep learning process. This study will lay a solid foundation for the development of on-line monitoring system of keyhole TIG.

Disciplines
Engineering | Science and Technology Studies

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers1/4107
Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation

Chunyang Xia, Zengxi Pan*, Zhenyu Fei, Shiyu Zhang, Huijun Li
School of Mechanical, Materials, Mechatronic and Biomedical Engineering, University of Wollongong, Wollongong, NSW, 2522, Australia

Corresponding author: Zengxi Pan
Email: zengxi@uow.edu.au
Postal address: Northfields Ave Wollongong, NSW 2522 Australia
Telephone numbers: +61 2 4221 5498

Abstract: As an advanced and highly efficient welding method, Keyhole Tungsten Inert Gas (keyhole TIG) welding has drawn wide interests from the manufacturing industry. In order to improve its manufacturing quality and automation level, it’s necessary to develop an online monitoring system for the keyhole TIG welding process. This study developed a visual monitoring system, which utilized an HDR welding camera to monitor the welding pool and keyhole during keyhole TIG welding process. A state of the art Convolutional neural network (Resnet) was developed to recognize different welding states, including good weld, incomplete penetration, burn through, misalignment and undercut. In order to improve the diversity of training dataset, image augmentation was performed. To optimize the training process, a metric learning strategy of center loss was introduced. Furthermore, visualization methods, including guided Grad-CAM, feature map and t-SNE were applied to understand and explain the effectiveness of deep learning process. This study will lay a solid foundation for the development of on-line monitoring system of keyhole TIG.

Keywords: keyhole TIG; welding; Deep learning; defects; visual monitoring; Resnet.

1. Introduction

In the modern manufacturing industry, Tungsten Inert Gas (TIG) welding has been widely applied due to its high welding quality. However, its process
efficiency is low and penetration capability is relatively weak because of its low arc current used. In order to improve the efficiency of heat source, researchers [1] [2] proposed to modify the welding torch and introduce heavier welding current to increase the arc pressure and enhance the penetration capability. As a result, keyhole mode can be produced during the welding process, and this new welding technology was known as keyhole TIG (Keyhole mode TIG) [2]. Compared with conventional TIG welding, keyhole TIG has superiority in energy density, welding efficiency and penetration ability. Additionally, it’s reported that the keyhole formed in keyhole TIG is more stable than that in the PAW process [3]. Due to these advantages, keyhole TIG has a broad prospect of application in the welding industry, especially for welding medium thickness materials, like titanium, stainless steel, low carbon steel and dissimilar metal welding.

Currently, some research effort has been carried out on keyhole TIG. Some studies are focusing on the metallurgical qualification of keyhole TIG welded stainless steel [4], carbon steel [5], titanium alloy [6], dissimilar metal [7]. Fei et al. [8], Xie et al. [9] And Fang et al. [10] researched the process optimization. Liu et al. [11] were studying the process dynamics of the keyhole. In the modern industry, the monitoring systems play an important role in ensuring the quality of production and improving the automation level. Although preliminary studies on monitoring of the keyhole TIG process have been conducted, extensive and thorough researches are still needed. Process monitoring and control are crucial to the welding process, since welding defects like incomplete penetration, lack of fusion and porosity may generate during the welding process. At the same time, human resources can be saved by improving the level of automation in production. Therefore, it's necessary to develop advanced monitoring systems for the keyhole TIG welding process.

In order to monitor the keyhole TIG welding process, researchers have tried several methods. Cui et al. [12] analyzed the frequency feature of arc voltage during keyhole TIG process, and found that the frequency features were able to reflect different welding state. Zhu et al. [13] developed a multi-sensor monitoring system, which combined an acoustic sensor with a CCD camera. Based on this system, the penetration state can be recognized based on the PS0-CV-SVM model during the keyhole TIG process. Although some monitoring systems have been designed to monitor the keyhole TIG welding process, the recognition is limited to the penetration state. Some other welding defects, such as undercut, burn-through and misalignment of plates, cannot be detected via existing monitoring strategies. In addition, interference may be generated in acoustic and electrical signals of existing monitoring systems during practical production.
The vision sensing method has been applied to monitor the conventional welding process, since it could obtain abundant information for pattern recognition. Zhang et al. [14] [15] designed a 3D positive visual sensing system to measure the surface profile of the welding pool of TIG. A dot-matrix laser pattern was projected to the welding pool surface, and the three-dimension shape of the welding pool surface could be reconstructed based on the reflection pattern. Based on the reconstruction of welding pool, the size of welding pool can be measured and controlled online [16] [17] [18]. Xu et al. developed a compact passive visual sensor for robotic seam tracking of the TIG welding process [4]. Fang et al. [19] developed a passive vision system using CCD camera in combination with filters to monitor the weld pool of hump formation during the GMAW process. The Chan-Vese model with a fuzzy C-means algorithm was developed to segment the weld pool image. It can be seen that visual signals contain more intuitive information and have the potential to recognize various welding states.

During the practical welding process, human welders are able to estimate the welding state and make adjustments on process parameters by observing the surface of the melt pool. As a cutting-edge technology, the deep learning algorithm imitates the working of the human brain, and has extensive application prospects in the modern manufacturing industry. Through applying deep learning networks to process monitoring, defects of products can be detected effectively, together with the increased productivity and reduced labor costs. Recently, researchers have proposed to take advantage of deep learning theory in visual inspection of the welding process. For example, Feng [20] et al. applied a deep learning method to implement image denoising and classification of penetration state for TIG. Bacioiu et al. [21] [22] and Zhang et al. [23] developed a novel deep learning architecture to classify welding defects of TIG.

As described above, AI technology has been introduced to the welding industry. However, the feasibility of applying AI technology in visual sensing of the keyhole TIG process has not been investigated. Considering the benefits of vision sensing and deep learning technology, this study developed a deep learning based vision sensing system to implement defects diagnosis for keyhole TIG welding. A high dynamic range camera integrated with a filter system was utilized to monitor the melt pool and keyhole. Furthermore, an advanced deep learning model (Resnet) was applied to extract the image feature of different welding defects deeply, and implement classification for various welding image. Usually, deep metric learning is often utilized to optimize distance loss to provide more discriminative features in image classification [24]. In this study, the center loss [25] was introduced to combine with conventional softmax loss to optimize the learning process.
In order to put more faith in the AI system and explain their decisions, it is necessary to build ‘transparent’ models, which have the ability to interpret what they predict. This issue is especially important for risk-sensitive applications. For example, one straightforward is to visualize the Convolution filters, which reveal the characteristic of the content extracted a layer in CNN. Recently, researchers proposed some novel algorithms to visualize deep learning models and interpret what is learned by them. Zhou et al. [26] proposed Class Activation Mapping (CAM) to identify discriminative regions and locate object. Selvaraju et al. [27] extended this method, known as Grad-CAM, which integrate CAM with existing pixel-space gradient visualization techniques. In this study, in order to understand the content learned by CNN from welding image, a visualization method named Guided Grad-CAM [28] was introduced to interpret the model results.

This paper will be organized as follows. In section 2, the deep learning model used in this study is introduced. The experimental setup, experiment detail and datasets for model training are presented. In section 3, the training results are presented and visualized by T-SNE, and the training process is visualized by Guided Grad-CAM and feature map to verify and understand the training process. The main findings in this study are summaries in section 4.

2. Methodology

This work aims to monitor the welding state during the keyhole TIG process by utilizing a deep learning model to classify the molten pool image. As a classical deep neural network, the convolutional neural network is composed of convolutional layers, spatial pooling layers, and fully connected layers. The convolutional layers consist of a series of filters, which are responsible for extracting deeper features from input. The pooling layers play a role in reducing the spatial dimensions of input images through conjoining the outputs of neuron clusters at one layer into a single neuron in the next layer. The fully connected layer is used to make a classification decision by imparting probability to each class.

In order to obtain high accuracy for welding state classification tasks, a suitable architecture for CNN need to be chosen. In the welding image, the features between different welding states is complex. Therefore, a deep network with multiple layers should be chosen to extract high-level features. In this study, a Residual Neural Network was applied to implement welding image classification due to its efficiency and capability in extending deeper networks.
2.1 Resnet model

Residual network (ResNet) is an improved deep learning algorithm for CNN, which is firstly proposed by Kaiming He [7] to improve the ability of image classification for the ImageNet dataset. Resnet doesn’t introduce a large number of parameters and large computation while it’s capable of solving the problems of vanishing and exposing gradients.

Resnet is composed of stacked objects known as residual blocks. They can solve the problems of vanishing and exploding gradients in stacked layers through introducing a residual mapping structure. At the same time, in order to accelerate the training process, Resnet also introduced the concept of shortcut connection, which could skip layers when training by skip-connection or residual connection. The basic element of design named “bottleneck” architecture. It mainly complies with two basic rules: (i) there will be the same amount of filters on the layers if the size of the feature maps for those layers are the same; (ii) the number of filters will be doubled when the feature map's size reduces by half. The down-sampling is performed by convolutional layers, which have a stride of 2 and batch normalization is performed after each convolution and before ReLU activation. When the input and output have the same dimensions, the identity shortcut is used. When the dimensions increase, the projection shortcut is used to match dimensions through 1×1 convolutions. In both cases, when the shortcuts go across feature maps of two sizes, they are performed with a stride of 2.

In this study, a Resnet model with 18 convolutional layers (Resnet18) is selected.

In order to improve the training process, a metric learning strategy of center loss was combined with Resnet. The training process was implemented by two modules: the original softmax loss and the introduced center loss. As illustrated in Figure 1, the center loss module was connected with the last average pool layer. At the stage of back propagation, the gradients information will be returned from both conventional softmax module and the newly introduced center loss module. Although conventional softmax loss function could provide separable identification information, it’s somehow not discriminatory for images. Therefore, the network with metric learning is necessary. For the center loss algorithm, the metric signal is the loss between each input image and the corresponding class center [25]. The center of each class is updated with the increasing of samples. The center loss function is calculated as:
\[ L_{cen} = \frac{1}{2} \sum_{i=1}^{N} \| f(x_i) - c_{yi} \|_2^2 \]  

(1)

Where \( f(x_i) \) represents the deep feature in the embedding space, \( c_{yi} \) is the \( y \)th class center and \( N \) is the batch size. The total loss for the network can be considered as the combination of softmax loss and center loss:

\[ L = L_{cls} + \lambda L_{cen} \]  

(2)

Where \( L_{cls} \) is the softmax loss, and \( \lambda \) is the weight of center loss. In this study, the value of \( \lambda \) is selected as 1.

![Figure 1 Architecture of Resnet](image)

2.2 Experimental setup

As illustrated in Figure 2, the experimental setup consists of two subsystems: keyhole TIG welding system and monitoring system. The keyhole TIG welding system mainly consists of keyhole TIG 1000 AMP power source, a control cabinet, water-cooling welding torch. The monitoring system mainly includes a vision sensing system and an electrical sensor. The visual sensing system utilized a Xiris
XVC-1000e HDR welding camera combined with a 650 nm central wavelength filter (30 nm of bandwidth). The camera is fixed by a 6 DOF holder above the work flat and aims at the key-hole entrance and welding pool area. The torch was operated in DCEN mode. The camera and the welding torch are stationary, and the workpiece was set to travel at a given speed during the welding process.

Experiments were performed on 250mm×150mm×6mm 304 stainless steel. Pure argon was utilized as the shielding gas with a flow rate of 25 L/min. In order to provide desired electrical characteristics, a 6.4 mm diameter of the tungsten electrode was utilized with a small addition of lanthanum. Offset from tungsten tip end to the workpiece is set to be 3 mm.

![Diagram](image)

**Figure 2 Schematic diagram of the welding experiment system**

### 2.3 Experiment process

<table>
<thead>
<tr>
<th>Welding current (A)</th>
<th>Welding speed (cm/min)</th>
<th>Type of defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>330-360</td>
<td>35</td>
<td>Lack of penetration</td>
</tr>
<tr>
<td>510-550</td>
<td>35</td>
<td>Burn through</td>
</tr>
<tr>
<td>435-465</td>
<td>35</td>
<td>Misalignment</td>
</tr>
<tr>
<td>390-410</td>
<td>35</td>
<td>Undercut</td>
</tr>
<tr>
<td>430-490</td>
<td>35</td>
<td>Good formation</td>
</tr>
</tbody>
</table>

**Table 1 experiment parameters**

Experiments are designed to obtain various welding states. In this study, five types of welding states can be obtained: good weld, incomplete penetration, burn through, seam misalignment and undercut. Table 1 lists the welding process
parameters and corresponding welding states. Each welding state was performed for 5 trials and each trial used different welding currents. 200 images were collected from each trial.

Figure 3 illustrates the surface morphology of the melt pool of different welding states. It can be observed that each welding state presents different melt pool morphology. For example, the morphology of keyhole and melt pool during the good weld state is relatively regular. When seam misalignment happens, the melt pool will slope to the lower side of the workpiece, especially the geometry of keyhole will present asymmetry. During incomplete penetration state, the surface of base metal melts while its bottom side is still solid. Therefore, the melt pool can't flow sufficiently and wrinkles may be generated. When the heat input during the welding process is inadequate, the volume of liquid metal is limited and can’t overspread all the area of weld bead (as shown in Figure 3. (b)). As a result, the undercut may be generated.

![Figure 3 Sample of welding pool dataset](image-url)
2.4 Datasets

Our study aims to classify the image of the weld pool based on a deep learning model. In order to train the deep learning model efficiently, the dataset needs to be enriched by the image augmentation method. When trained by augmented images, the deep learning network is possible to learn the features deeply and robustly. The dataset was split into two sub-datasets: training dataset and validation dataset. In this study, a dataset containing 5000 images (1000 for each classification) is as training dataset, and 1000 (200 for each class) images will be used for validation. The images for those two datasets were selected randomly.

The size of the original image from the camera is 1280×1024 pixels. The welding pool area is located at center of the image, and a large number of black pixels exist on the image. In order to reduce the amount of calculation, the images were cropped to 880×780 pixels, and then resized to 256×256 pixels.
Data augmentation could enhance the diversity of the dataset significantly for the model’s training without collecting additional data actually. Commonly used data augmentation methods for training large neural networks include cropping, padding, and horizontal flipping and so on. Applying those augmentation methods could help in reducing overfitting by introducing slight distortion. In machine learning, as well as in data science, overfitting happens when a model represents unregular noise and error instead of substantial relationships. Moreover, data augmentation could improve the performance of a trained model[29]. As shown in Figure 5, a series of image augmentation methods, including rotation, flipping, and random aspect ratio are implemented on images randomly. Afterwards, the input images were normalized with the mean and standard deviation value of [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225] respectively. The augmentation operation was only performed on the training set, while the validation dataset used the real-world data.

![Image transformations used for augmentation](a) original image; (b) rotations; (c) random aspect ratio; (d) flipping

Compared to the conventional welding method, keyhole TIG generates stronger arc light during the welding process. Although some optical filters can be used to constrain the arc light, the welding images still inevitably become hazy and the region far from arc center may be lack of illumination. The poor visibility may weaken the features of an image, and prevent the deep learning network from extracting features efficiently. Therefore, Retinex algorithm is applied to enhance the welding image. Retinex algorithm is a classical image enhancement algorithm that can be applied to generate better visual representations of images. It can
perform nonlinear spatial transforms that integrate strong color constancy and local contrast enhancement.

Retinex algorithm is able to separate illumination from the reflectance in images, and compensate for the disunited lighting in image[30]. Therefore, Retinex algorithm has superiority in processing contrast and detail information. In this study, Retinex is applied to defog and improve illustration in the keyhole TIG welding image. From Figure 6, it can be seen that the fuzziness can be removed effectively and the detail in the melt pool area becomes clearer.

![Figure 6 Retinex based image enhancement: (a) before enhancement (b) after enhancement](image)

3. Results and discussion

3.1 Modelling results

In this study, a ResNet18 model was selected as the base model, and the center loss was introduced to optimize the learning process. The aim of the training stage is to extract sufficient feature representation and find the probability distribution which describes the dataset. The probability distribution is parametrized by the weights, which comprise the kernels and fully connected layers in the architecture. Adaptive Moment Estimation (Adam) [31] algorithm will be utilized for training convergence. During the initial training period, a relatively larger value of 0.001 for the learning rate was used to avoid to trap in a local optimum. During the later stage, the learning rate decayed by a factor of 0.1 every 7 epochs so that the training process could converge to the optimum. A batch size of 12 samples, momentum $\beta=0.9$ were used for training. The deep learning framework and its visualization in the next section were implemented by Pytorch. Pytorch provides a deep learning library for rapid and convenient development, adoption and application of deep learning model architectures [24]. The model training was implemented by GPU on Google Colab.

The accuracy curve and loss curve for training and validation epochs are shown in Figure 7 (a) and (b), respectively. As we can see in the figures, both loss curves
and training curves converge rapidly and an accuracy above 98% was achieved for validation dataset after 20th iterations. It can be also observed that the accuracy of testing is lower than that of training, while the loss of testing is higher than that of training during the convergent period [32].

![Figure 7 Curve of accuracy and loss: (a) accuracy curve (b) loss curve](image)

The confusion matrices of classification based on Center loss-Resnet is presented in Figure 8. The elements in each row represent true class while the element in columns represent the predicted class. The higher the values of diagonal elements are, the better classification performance can be obtained. From the confusion matrices, we can observe that Resnet model has high classification accuracy for the welding image. The accuracy of classification achieved 0.955, 0.950, 1.000, 0.990, 0.985 for good weld, incomplete penetration, burn through, misalignment and undercut respectively. This result verifies the potentiality of deep learning to be applied in realistic K-TIG welding production.

From Figure 8, it can be observed that compared to under penetration and burn through class, the classification accuracy for good weld, misalignment and undercut was relatively low, which means mismatch may happen among them. The reason for this phenomenon is that the feature of shape and texture for the later three classes are relatively closed to each other (it can be observed in Figure 3 ). Additionally, due to the fluctuation during welding process, the divergences may be generated in testing images, and the feature for pattern recognition will be weakened, which may lead to the misclassification between similar classes.
For comparison, a conventional SVM classifier was also developed to implement classification for the welding image. The same training and testing dataset were utilized. The main parameters of SVM include instant C and kernel parameter gamma. In this study, an RBF kernel function was utilized, and the instant C and kernel parameter gamma were selected to be 0.01 and $1 \times 10^{-8}$ respectively. The SVM classifier was developed by Python language using the Libsvm package [33]. The accuracy and precision of Center loss-Resnet, Resnet and SVM were compared and presented in Table 2. The accuracy is simply the ratio of correct prediction to the total observations, while precision means the percentage of correct prediction in positive predictions. They can be calculated as:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$  \hspace{1cm} (1)

$$\text{Accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{True Negtive} + \text{False Positive} + \text{False Negtive}}$$  \hspace{1cm} (2)

From Table 2, it can be seen that the overall accuracy of Center loss-Resnet is 0.984, while the accuracy of Resnet and SVM is 0.978 and 0.846 respectively. Also, Center loss-Resnet performs better than Resnet and SVM in terms of Precision. Furthermore, it can be observed that although SVM can achieve relatively high accuracy when recognizing under penetration, good weld and misalignment and undercut, poor performance was obtained in recognizing image of burn through. In contrast, Center loss-Resnet and Resnet can both achieve 100% accuracy when recognizing image of burn through, which demonstrates the superiority of Resent in classifying welding images.
Table 2 Performance of classification of Resnet and SVM

<table>
<thead>
<tr>
<th></th>
<th>Center loss-Resnet</th>
<th>Resnet</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>precision</td>
<td>Accuracy</td>
<td>precision</td>
</tr>
<tr>
<td>Under penetration</td>
<td>0.995</td>
<td>0.955</td>
<td>0.995</td>
</tr>
<tr>
<td>Good weld</td>
<td>0.974</td>
<td>0.950</td>
<td>0.959</td>
</tr>
<tr>
<td>Burn through</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Misalignment</td>
<td>0.990</td>
<td>0.990</td>
<td>0.980</td>
</tr>
<tr>
<td>Undercut</td>
<td>0.961</td>
<td>0.985</td>
<td>0.956</td>
</tr>
<tr>
<td>Average value</td>
<td>0.984</td>
<td>0.984</td>
<td>0.978</td>
</tr>
</tbody>
</table>

(a) Figure 9 visualization of deep feature: (a) Combination of Softmax loss and Center loss (b) Softmax loss only

To intuitively investigate the effectiveness of center loss on Resnet learning, the resulting 2-D deep features are plotted in Figure 9. These deep features are the output of average pool layers. It can be observed that although the features extracted from the deep neural network trained based on softmax loss are separable, they are still not discriminative enough. After introducing Center loss together with softmax cross-entropy loss to train the model, the discriminative ability has been enhanced. This could help improve the learning process for Resnet.
3.2 Visual explanation

3.2.1 Feature map

Through visualizing a feature map for a specific input image, it will be able to interpret what features in the input image are detected or preserved. Features maps are the output of various deep convolutional kernels, which is able to provide insight into the internal representation of local feature extractors. Normally, feature maps were considered to contain more detailed information when the network becoming deeper. In order to investigate the performance of the deep learning model of Resnet, the feature map is visualized in Figure 10. Four most significant features in various layers are selected and presented when the input images belong to incomplete penetration and burn through respectively. In features images, the highly activated locations suggest that around them it contains the visual patterns depicted in the convolutional layer filter. It can be seen that the CNN only extracts intensity, shape, textures and edges information in lower convolutional layers. If the structure of CNN becomes deeper and the number of network parameters increases, the features will become more abstract.

Through visualizing the feature maps of Resnet, the features in the image can be well interpreted in terms of their physical meaning and extraction pattern. It can be observed that in Conve1, the main features of shapes and edges can be extracted and highlighted, such as the shape, size and angle of keyhole, as well as morphology and texture of the welding pool. It’s obvious that the main morphological differences of each welding status can be recognized in Conve1 layer. This means the Resnet models are capable of imitating the work mode of the human brain when estimating the welding state based on visual information. When the layers become deeper, the feature maps become more abstract and hard to be described. As can be seen in Layer1, the shape of the keyhole is somewhat visible, but after that it becomes unrecognizable. The reason is that deeper feature maps encode high-level concepts like “keyhole” or “melt pool” while lower-level feature maps detect simple edges and shapes.
### 3.2.2 Guided Grad-CAM

During the training process of a deep learning network, some visualization techniques can be utilized to understand and verify the training process. Guided gradient-class activation mapping (guided Grad-CAM) highlight the regions that the deep learning model considers important for the classification, which provides an effective verification for the training process of the deep learning network. Based on class-specific gradient information from the last convolutional layer of a CNN, a coarse localization map of the important regions in the image can be generated by Guided Grad-CAM. Consequently, the vital area in the image, where corresponds to decisions of interest, can be visualized with high resolution.

To obtain Grad-CAM $L_{Grad-CAM}^c$, the gradient of class score $y^c$ with respect to map activation $A^k$ need to calculate firstly. Based on this gradient information, the neuron weight ($\alpha_k^c$) of the feature map $k$ for the target class $c$, can be calculated as:

$$
\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}
$$

(3)
Then, the forward activation maps are combined linearly through a ReLU block to obtain $L_{Grad-CAM}^c$:

$$
L_{Grad-CAM}^c = ReLU \left( \sum_k \alpha_k^c A_k \right)
$$

Through fusing Guided Backpropagation [34] and Grad-CAM via element-wise multiplication, Guided Grad-CAM can be obtained. More details can be found in relevant literatures [28] [27]. Figure 11 shows the procedure of generating Guided Grad-CAM through combining Grad-CAM and Guided Backpropagation using pointwise multiplication.

As shown in Figure 12, CAM and Guided Grad-CAM were used to visualize the important regions, which support the Resnet18 model to predict different welding image class. One example image for each class of welding images in the first row were correctly classified using Resnet. The second and third rows present the CAM and Guided Grad-CAM map for each image. In CAM, the important regions highlighted in red reveal that the Resnet can use the most informative features to correctly identify the class of the image. It can be observed that the keyhole and surrounding melt pool area was highlighted, which means the most important information is extracted from this area. Furthermore, the Guided Grad-CAM visualizations can present finer details of localization, shape, and texture of keyhole and melt pool area. For example, it can be seen that the texture of melt pool was extracted as the main feature for ‘incomplete penetration’ class. For ‘burn through’ class, the valid features in the whole collapse area was utilized to determine its class. Therefore, it is believed that Resnet can effectively capture the class-discriminative features.

<table>
<thead>
<tr>
<th>Incomplete penetration</th>
<th>Burn through</th>
<th>Undercut</th>
<th>Good weld</th>
<th>Misalignment</th>
</tr>
</thead>
</table>
4. Conclusion

In this study, a novel visual sensing system for keyhole TIG is designed. The image of the melt pool and keyhole can be collected by this visual sensing system. This visual sensing system effectively filters out the powerful arc light while balancing the image and improve the details in the melt pool. Unlike the traditional processing method for the welding image, we seek to apply human intelligence in visual information recognition. Data augmentation was utilized to improve the diversity of the original dataset, and a Resnet was applied to classify different types of welding images. Furthermore, a metric learning strategy of center loss was introduced to optimize the learning process. The modelling results show that the deep learning framework is capable of learning key features from welding image, and classify different welding states, including good weld, incomplete penetration, burn though, misalignment and undercut. An overall accuracy above 98% was achieved, which can meet the requirements of practical production. Feature maps of various layers were visualized to interpret the physical meaning of the deep learning features. Moreover, the deep learning features were visualized through the guided grad-CAM algorithm to interpret the model's decision. It can be observed that the features extracted by deep learning mainly relied on shapes of keyhole and morphology of the melt pool surface.

This study investigates the feasibility of applying deep learning networks in keyhole TIG welding process monitoring. In the future, more welding states should be considered and the dataset will be expanded to improve the robustness of deep learning model. Furthermore, a deep learning algorithm will be optimized to improve accuracy and real-time performance.
Acknowledgements

The authors gratefully acknowledge the China Scholarship Council for financial support (NO. 201704910782) and UOW Welding and Industrial Automation Research Centre.
Reference