**Advanced Steel Construction** – Vol. 20 No. 2 (2024) 169–178
**DOI:** 10.18057/IJASC.2024.20.2.9
(Special Issue: *11st International Conference on Advances in Steel Structures, ICASS'2023, Kuching, Sarawak, Malaysia*)

# SUPER-RESOLUTION RECONSTRUCTION AND HIGH-PRECISION TEMPERATURE MEASUREMENT OF THERMAL IMAGES UNDER HIGH-TEMPERATURE SCENES BASED ON NEURAL NETWORK

Yi-Chuan Dong, Jian Jiang [*], Qing-Lin Wang, Wei Chen and Ji-Hong Ye

*Jiangsu Key Laboratory of Environmental Impact and Structural Safety in Engineering, Xuzhou, China*
** (Corresponding author: E-mail: jian.jiang@cumt.edu.cn)*

## ABSTRACT

Accurate temperature readings are vital in fire resistance tests, but conventional thermal imagers often lack sufficient resolution, and applying super-resolution algorithms can disrupt the temperature and color correspondence, leading to limited efficiency. To address these issues, a convolutional network tailored for high-temperature scenes is designed for image super-resolution with the internal joint attention sub-residual blocks (JASRB) efficiently integrating channel, spatial attention mechanisms, and convolutional modules. Furthermore, a segmented method is developed for predicting thermal image temperature using color temperature measurements and an interpretable artificial neural network. This approach predicts temperatures in super-resolution thermal images ranging from 400 to 1200°C. Through comparative validation, it is found that the three-neuron neural network approach demonstrates superior prediction accuracy compared to other machine learning methods. The seamlessly combined proposed super-resolution architecture with the temperature measurement method has a predicted RMSE of 20°C for the whole temperature range with over 85% of samples falling within errors of 30°C.

## ARTICLE HISTORY

| | |
|---|---|
| Received: | 4 June 2024 |
| Revised: | 4 June 2024 |
| Accepted: | 11 June 2024 |

## KEYWORDS

High-temperature scene;
Infrared thermal imaging;
Neural networks;
Image super-resolution;
Color temperature prediction

## 1. Introduction

Steel structures, despite their widespread use, suffer from poor fire resistance, leading to significant performance degradation, as exemplified by the loss of mechanical properties in portal frames at 600°C [1]. To tackle this problem, fire resistance testing is crucial for assessing component performance under high temperatures. During the test, obtaining comprehensive and precise temperature data is paramount. While thermocouples offer reliable measurements, they are limited to single points and prone to failure, compromising efficiency. Infrared thermal imaging cameras offer a solution by capturing the entire two-dimensional temperature field. However, their high cost and low-resolution present challenges, particularly in medium to large-scale fire tests, where detailed temperature data are essential.

The limited resolution of thermal imagers hinders the acquisition of optimal temperature data in experiments, and the small temperature matrix fails to precisely represent spatial temperatures, thus sparking interest in enhancing the resolution of instruments for production and cost challenges [2]. As a result, algorithmic image processing emerges as a cost-effective and efficient alternative. The existing interpolation methods [3] are swift, but their effectiveness is limited, especially when dealing with large scaling factors. Reconstruction-based super-resolution techniques [4,5], which leverage prior knowledge and predefined data dictionaries, offer superior image reconstruction but increase workload and reduce processing speed.

With the advancement of artificial intelligence, the convolutional neural network (CNN) based method has emerged as the preferred approach due to its strong learning ability and flexible adjustability [6,7,8]. Utilizing the CNN architecture, this method seamlessly transforms low-resolution images (LR) into super-resolution versions (SR), ensuring close resemblance to the original high-resolution images (HR).

Multiple studies have focused on refining the CNN super-resolution architecture, incorporating advancements such as residual learning, image sampling, and attention mechanisms [9]. Residual learning, embodied in models such as Very Deep Super-resolution Network (VDSR) [10], Deeply-recursive Convolutional Network (DRCN) [11], Deep Recursive Residual Network (DRRN) [12], and Enhanced Deep Residual Super-resolution Network (EDSR) [13], boosts the learning capacity through modular expansion and skip connections. Image sampling, enhanced by Shi et al. [14] with subpixel convolution, efficiently converts channel data to spatial features. Zhang et al. [15] combined channel attention [16,17] with residual blocks in their model, creating a high-efficiency module (RCAB). This integration maximized the benefits of the attention mechanism but still had an issue by introducing residual scaling.

In comparison to the aforementioned visible light image SR model, research on thermal image processing is more constrained. Initial super-resolution models for thermal images were primarily derived from visible light models: including Thermal enhancement network (TEN) [18], Brightness-Based Network [19], and GAN architecture [20]. A significant hurdle in thermal image processing research is the challenge of acquiring high-quality and highly pertinent datasets. Public thermal image datasets [21,22] predominantly feature infrared grayscale images from everyday settings. These single-channel images offer limited assistance in training three-channel color thermal models, thus yielding subpar outcomes. Additionally, existing models often lack tightly integrated attention and residual learning mechanisms, leading to bulky networks. He et al. [23] employed a two-stage cascade architecture for ×8 super-resolution, progressively restoring the image size from ×2 to the desired target. While effective for learning multi-scale features, its focus on large scaling factors results in a sizable model with reduced operational efficiency. These studies mainly focused on image resolution, ignoring its suitability for temperature measurement. Even with super-resolution techniques, the altered color, temperature logic, and image format make the enlarged images incompatible with thermal analysis software.

Infrared thermal imagers in the domain of image temperature measurement utilize the blackbody radiation correlation formula to determine the color-temperature correspondence for each pixel [24]. However, existing color temperature measurement algorithms designed for high temperatures, such as two-color method [25], laser speckle photography [26], and image fusion techniques [27], often demand sophisticated and costly optical devices along with stable, vibration-free systems. Their intricate setup and restrictive measurement requirements pose significant challenges for direct application in high-temperature environments.

Prediction models based on machine learning only need to be given input characteristics to predict the target, and this method has effectively proven its worth in a range of detection tasks related to fire [28,29,30] and structural engineering [31,32,33]. Sun et al. [34] introduced an enhanced BP network for adaptive spatial data sensing to measure tunnel ceiling temperatures. Its efficacy was validated through tunnel fire tests, with a prediction error of just 0.12 under dimensionless scaling. Chen et al. [35] proposed a machine learning method for structural analysis, incorporating a Physics-Informed Neural Network (PINN) to steer the training process. When dealing with small datasets, the prediction of the neural network closely resembles the results obtained from finite element software. The applicability and accuracy of this method were verified through four sets of examples. Wang et al. [36]

introduced an inverse model utilizing a Graph Neural Network (GNN) to address structural parameter identification from structural responses. This model transforms the truss structure into a GNN graph, delivering precise inversion outcomes even with limited data. Its strong interpretability and high potential in structural analysis make it a valuable tool. Sharifi et al. [37] employed artificial neural networks to assess the load-bearing capacity of various corroded steel beams under localized loads, accurately predicting the failure of these deteriorated beams.

In this paper, a specialized CNN architecture designed for high-temperature environments was proposed to address the challenges of low-resolution thermal images and the incompatibility of super-resolution techniques for temperature readings in fire tests, as is shown in Fig. 1. This architecture was enhanced with the Joint Attention Sub-Residual Block (JASRB), generated precise, high-definition color and temperature mappings. A temperature measurement approach utilizing artificial neural networks was developed for super-resolution images. Fire tests were conducted to validate the accuracy of the separated approach and their combination.

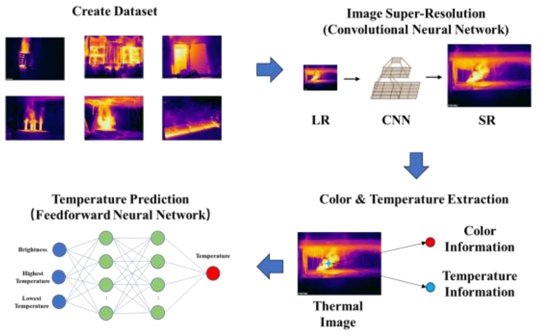


**Fig. 1** Overall architecture of super-resolution network

## 2. Super-resolution of thermal images

The proposed architecture is an innovative thermal image super-resolution (SR) network that features a multi-path structure with JASRBs. It has four main parts: the LR image processing part (Part 1), the preliminary clarification part (Part 2), the fine recovery part (Part 3), and the image reconstruction part (Part 4). The core of the architecture lies in the preliminary clarification and fine recovery parts, which use different JASRB configurations.

### 2.1. Inputs and Outputs

The proposed architecture is designed to enhance thermal imaging resolution by utilizing pairs of low and high-resolution images for network training. To faithfully replicate LR images captured by budget thermal cameras, HR images are down-sampled by a scale factor of four. The degradation function for this process can be stated as follows.

$$I_{LR}=D(I_{HR};\delta) \tag{1}$$

Where $I_{LR}$ denotes the original low-resolution image; while $D$ signifies the Bicubic image interpolation function employed for image degradation mapping; $I_{HR}$ represents the high-resolution image, and $\delta$, represents the scaling factor, is set to 4 in this context.

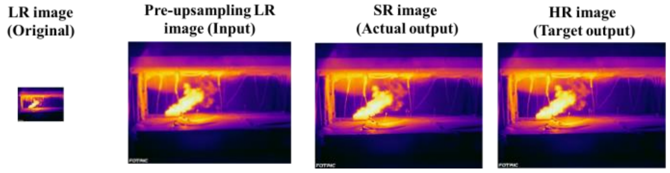


**Fig. 2** Input and Output of the CNN architecture

The input to the CNN is a low-resolution image that has undergone Bicubic interpolation for pre-upsampling (Fig. 2). This interpolation ensures that the image size matches the target high-resolution (HR) image, despite the internal clarity remaining insufficient. The objective of CNN is to produce a high-resolution image, with the actual output being a super-resolution (SR) image. In practical scenarios, the HR image remains unknown. Hence, the primary goal of this undertaking is to align the SR image closely with the HR image, thereby elevating the thermal image quality in unfamiliar settings. Nonetheless, discrepancies often arise between the actual and target outputs, necessitating optimization to narrow this divergence.

### 2.2. Network architecture

As shown in Fig. 3, the LR image processing part preprocesses input imagery, adjusting dimensions with bicubic interpolation to match the specifications of the RGB input layer. It then segments RGB images into 48×48×3 patches for network training. The details of hyperparameter settings for the first convolutional layers in each part are listed in Table 1.

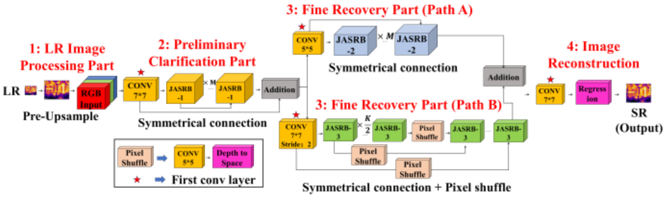


**Fig. 3** Overall architecture of super-resolution network

**Table 1**

Parameters of the first convolutional layer of each part of the network

| Position | Num Filters | Kernel Size | Stride |
|---|---|---|---|
| Part 2 | 48 | 7*7 | 1 |
| Part 3 Path A | 64 | 5*5 | 1 |
| Part 3 Path B | 64 | 7*7 | 2 |
| Part 4 | 3 | 7*7 | 1 |

The preliminary clarification part aims to extract prominent features from thermal images using convolutional layers with few filters for efficiency. The large-scale features, often comprising shallow details like scene structure, can be adeptly extracted via convolutional layers with few filters. It employed a 7×7 receptive field, zero-padding, and stacked $M$ JASRBs with symmetric connections to minimize shallow feature loss. The output is a 48×48×48 feature map.

The third segment restores intricate details using multi-path learning, converging diverse features from varying receptive fields. Learning image details is more challenging than shallow features due to their intricate link with large structures, leading to easy loss or misinterpretation [23]. Therefore, more extensive parameter learning is required. Similar to the preliminary clarification part, Path A amplified the filter and stacked $M$ JASRBs used for deeper abstraction, while Path B halved the image size and introduced $K$ JASRBs (where $K$ is an even number) with pixel shuffling [14] to obtain different details from Path A. Finally, the segment used the element-wise addition layer to fuse the features from both paths, resulting in a 48×48×64 output patch.

The final part, the image reconstruction part, used a 3-channel convolutional layer to resize the output to RGB dimensions. A 7×7 convolution kernel aggregated features and a regression prediction layer produced the SR image.

This architecture used half-mean-square error loss as the loss function of the neural network, which can be expressed as:

$$Loss=\frac{1}{2}\Sigma_{p=1}^{HWC}(t_p\text{-}y_p)^2 \tag{2}$$

where $H$, $W$, and $C$ denotes the height, width, and number of channels of the input image, respectively; $t$ and $y$ denote the target image and the image predicted by the network, respectively; $p$ is linearly indexed into each element of $t$ and $y$.

### 2.3. Joint attention sub-residual block (JASRB)

The performance of the network is primarily defined by the architecture, which consists mainly of JASRBs. Traditional modules such as RCAB, which integrate channel attention and residual modules, often yield limited benefits since channel attention alone may not significantly aid training. This holds especially true in the case of thermal images, where color consistency restricts the amount of information obtainable from color channels, thereby highlighting the significance of spatial attention. Therefore, relying solely on channel attention mechanisms can lead to deficient learning ability, impeding deep network training. To address this, a custom spatial attention module was innovatively combined with the existing channel attention module within the RCAB framework.

The overall structure of the JASRB is shown in Fig. 4, which consists of three parts: the Convolution Module (CM), the Channel Attention Module (CAM, Fig. 4b) [15], and the proposed Spatial Attention Module (SAM, Fig. 4c). The channel attention module compacts the feature graph and generates the channel weight matrix of 1×1×*C*. It demonstrates the significance of every channel in the feature map subsequent to the output of CM. The details of hyperparameter settings for the convolutional layers in each JASRB are listed in Table 2. All kernels in Table 2 are 3×3 in size with a stride of 1, ensuring network consistency and efficiency.

**Table 2**
Number of filters of each JASRB in different parts

| Location | Num Filters | Kernel Size | Stride |
|---|---|---|---|
| CM Conv1 | 192 | 256 | 128 |
| CM Conv2 | 48 | 64 | 64 |
| SAM Conv1 | 64 | 64 | 64 |
| SAM Conv2 | 1 | 1 | 1 |
| CAM Conv1 | 10 | 10 | 10 |
| CAM Conv2 | 48 | 64 | 64 |

A spatial attention extraction block consists of two convolutional layers activated by the ReLU function [38]. The first convolution layer maps the features directly to the higher-dimensional space and learns them, while the second convolution layer integrates them into the one-dimensional space. This combination is straightforward and can generate a weight matrix of size $H$×$W$×1, representing the more important spatial position in the image. The expression for this part is as follows:

$$F_s=\omega_I*max(0,\omega_E*\mathrm{F}_{\mathrm{o}}+b_E)+b_I \tag{3}$$

$$g=\frac{1}{1+e^{-x}} \tag{4}$$

$$s_{CA}=g(F_s) \tag{5}$$

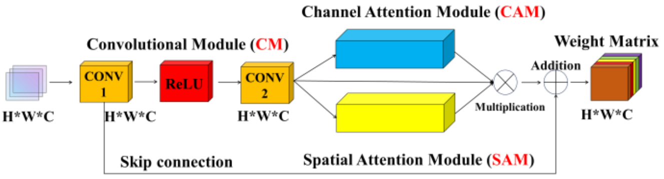
where $F_s$ denotes the features learned through spatial attention; $\omega_E$, $b_E$ and $\omega_I$, $b_I$ respectively denotes the weights and biases of the first and second convolutional layers of the proposed spatial attention extraction block, that is, the internal weights and other information of the feature learning layer and the feature integration layer; $F_o$ represents the original features; $g$ denotes the Sigmoid function and $s_{CA}$ denotes the final spatial attention matrix.

The output of the entire block can then be expressed as:

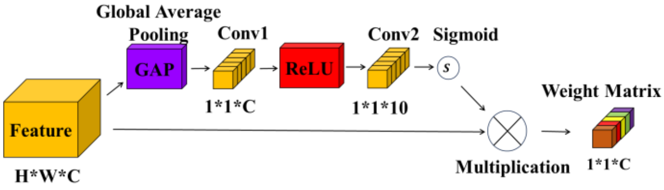
$$X_i=\omega_{i|2}*max(0,\omega_{i|1}*F_o+b_{i|1})+b_{i|2} \tag{6}$$

$$F_i=F_{i\text{-}1}+(s_{CA}*s_{SA}*X_{i\text{-}1}) \tag{7}$$

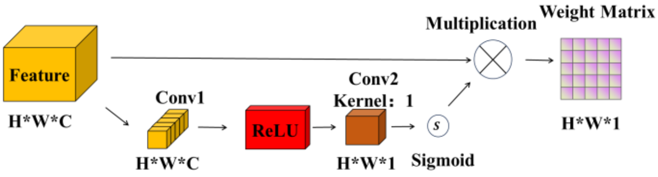
where $\omega_{i|1}$, $b_{i|1}$ and $\omega_{i|2}$, $b_{i|2}$ denotes the weights and biases of the first convolutional layer and the second convolutional layer in the residual block respectively; $X_i$ and $X_{i\text{-}1}$ are the residual information between the current layer and the previous layer after learning; $F_{i\text{-}1}$ and $F_i$ denotes the output features of the current layer and $s_{CA}$, $s_{SA}$ is the channel and spatial attention feature matrix in current layer, respectively.



(a) Joint attention sub-residual block (JASRB)



(b) Channel attention module (CAM)



(c) Spatial attention module (SAM)

**Fig. 4** Architecture of CAM, SAM and JASRB

By integrating both channel and spatial attention mechanisms within the skip connection, improved residual scaling efficiency and enhanced learning of spatial image features can be achieved, as evident in sections 2.4.2.2 and 2.4.3.3.

## *2.4. Training of the convolutional neural network*

### *2.4.1. Training details*

The GPU used for training was NVIDIA GeForce RTX3060Ti, and the CPU was Intel Core i7-10700CPU. During the training process, the minibatch size was set to 32, and the optimizer was Adam optimization [39]. For a total of 100 epochs of training (a total of 120000 iterations), the learning rate was set to 0.001 and decreased to half of the original value every ten rounds during the training process. Each training image was randomly cropped into 64 patches, with random shuffling performed in each iteration, and L2 regularization was applied to prevent the gradient from being excessively large. The network training platform was MATLAB R2022b. The initial weight and bias learning rate of the convolution layer were set as 1.

### *2.4.2. Dataset*

In this study, 750 thermal images were employed, which were partitioned into a training set comprising 600 images, a validation set of 50 images, and two test sets, each containing 50 images. The first test set served for the ablation study aimed at refining the architectural hyperparameters, whereas the second test set was utilized for the ultimate model comparison. This approach mitigated the risk of hyperparameter optimization outcomes being exclusively tailored to a subset of the dataset. Fig. 5 displays a schematic overview of the thermal images dataset captured in various fire scenes, along with the shooting equipment. Utilizing thermal images from diverse scenes enhances the generalization of CNN and elevates the quality of training.

The Fotric348 series handheld thermal imager was employed, boasting a 640×480 thermal image resolution. The specially formatted images of the instrument can export the temperature matrix using the software of the manufacturer, AnalyzIR. Its uncooled infrared focal plane detector has a response band of 7-14μm and can detect temperatures between -20°C and 1200°C. Thermal image emissivity is set to 0.95, in this part, emissivity does not affect the quality of the image. The reasons for using this setting are explained in detail in section 3.4.1.

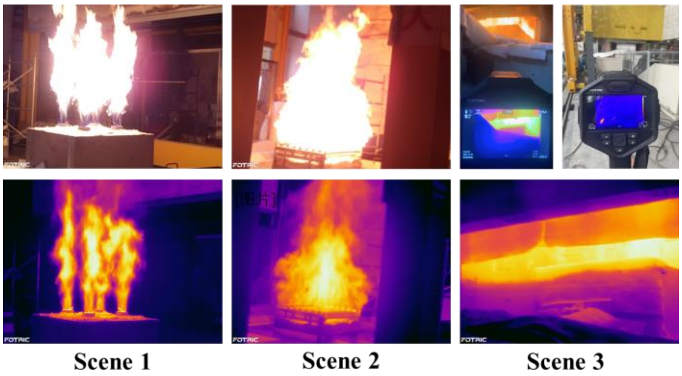


**Fig. 5** Shooting equipment and some scenes

The thermal imager was positioned 4~15m from the fire source to obtain the pictures, and such shooting distance is typical for civil engineering applications. To ensure stability and prevent focusing issues due to external factors, a tripod was used throughout the image capture process.

### *2.5. Experimental simulation*

### *2.5.1. Evaluation index*

The evaluation metrics for the proposed super-resolution network include Peak Signal-to-noise Ratio (PSNR), Structural Similarity (SSIM), and Natural Image Quality Evaluator (NIQE) [10]. PSNR measures the difference between HR and SR images based on the maximum pixel value and root mean square error. The calculation formula of PSNR is:

$$PSNR=10log_{10}\frac{peakval^2}{MSE(A,Ref)} \tag{8}$$

where *peakval* specifies the range of the image data type. For 8-bit images, the *peakval* value is 255; *MSE (A, Ref)* represents the mean square error between the specified image *A* and the reference image *Ref*. A higher PSNR indicates smaller differences, with values between 30 and 40 considered acceptable distortion loss and above 40 indicating near-original quality.

SSIM assesses the similarity of two images by evaluating brightness, contrast, and structure independently, ranging from 0 (no similarity) to 1 (identical images). The calculation formula of SSIM is:

$$SSIM(x,y)=[l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \tag{9}$$

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x{}^2\mu_y{}^2+C_1} \tag{10}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x{}^2\sigma_y{}^2+C_2} \tag{11}$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{12}$$

where $l(x,y)$, $c(x,y)$ and $s(x,y)$ represent brightness term, contrast term and structure term respectively; $\mu_x$, $\mu_y$, $\sigma_x$, $\sigma_x$ and $\sigma_{xy}$ represent the local mean, standard deviation and cross-covariance of the image *x*, *y*. $C_1$, $C_2$ *and* $C_3$ represent three constants respectively, where $C_1=(K_1L)^2$, $C_2=(K_2L)^2$, $C_3=0.5*C_2$. $K_1$ and $K_2$ are 0.01 and 0.03 respectively, and *L* represents the pixel value range, which is 255 for 8-bit images.

NIQE, which does not require a reference image, assesses naturalness by measuring distribution differences against multivariate natural scenes, with lower values indicating better visual quality in line with human perception [40].

### *2.5.2. Ablation study*

To validate the hyperparameter settings, the impact of spatial attention mechanisms, and the use of double residual scaling factors, three ablation tests were conducted to assess the optimal number of network modules, evaluate the significance of JASRB components, and verify the residual scaling approach within JASRB.

### *2.5.2.1. JASRB module quantity analysis*

The number of JASRBs in the two paths in the network determines the prediction accuracy and other critical indicators of the neural network. The output of this super-resolution task will be used in image temperature recognition. Any inaccurate prediction may significantly affect the temperature output, and therefore the SR image should be as clear as possible.

The number *M* of two modules was set from 3 to 5, and the module number *K* of path B was set as 2 or 4. Table 3 shows the prediction results of different module numbers. Since the learnable parameters of the comparison module did not increase significantly, and the images were relatively clear, the changes of indicators were relatively small. The clearer the image, the more difficult it is to restore. The reason for the fact is that only the delicate parts of the image were left, and it was also challenging to discover them. The improvement of index represents the degree of the repair of details, every increase of 0.1 in PSNR is more helpful for detail recovery. These details hardly have a significant impact on the indicators, but they do have a great impact on the image temperature measurement in the third section of the paper.

**Table 3**
Number of JASRBs

| Number of *M* and *K* | PSNR(dB) | SSIM | Time(s) |
|---|---|---|---|
| 3, 4（baseline） | 39.9947 | 0.9986 | 1.55 |
| 3, 2 | 39.9626 | 0.9986 | 1.32 |
| 4, 2 | 40.0961 | 0.9986 | 1.56 |
| 4, 4 | 40.2143 | 0.9987 | 1.57 |
| 5, 2 | 39.7860 | 0.9986 | 1.72 |
| 5, 4 | 0 | 0 | 0 |

As the number of modules increased, the network training effectiveness improved until *M* reached 5, whereupon performance started to deteriorate, signifying the peak capacity of network. Adding more depth beyond this point failed to facilitate meaningful learning. Furthermore, as the module count *K* rose, gradient explosion ensued, leading to unbounded training losses. When considering single-image reconstruction time, using the minimum number of modules yielded the fastest computation. Nevertheless, prioritizing image quality, the model with both *M* and *K* set to 4 emerged as the optimal choice.

### *2.5.2.2. Internal importance analysis of JASRB*

To validate the significant impact of incorporating spatial attention on the network, and to precisely assess the importance of each submodule for guiding parameter adjustment, a submodule with minimal parameters was employed as the baseline model. Subsequently, a comparison was made by elevating the parameters of the three respective submodules to a certain level.

The comparison results, summarized in Table 4, revealed that the SAM was nearly as crucial as the CM, while the CAM had a more negligible contribution. Specifically, the PSNR of the CM and SAM increased by 0.5167 and 0.4973, respectively. The superior performance of the residual block can be attributed to its direct connection with the original information, preventing information loss during interlayer transfers. The improvement of proposed SAM closely approached the optimal value, despite not being directly linked to the original features.

**Table 4**
Comparison of the importance of each part of the sub-module

| Models | PSNR(dB) | SSIM | Parameters |
|---|---|---|---|
| Baseline model | 38.0406 | 0.9976 | 212.3k |
| Add parameters in CM | 38.5573 | 0.9980 | 378.3k |
| Add parameters in CAM | 38.2125 | 0.9977 | 379.1k |
| Add parameters in SAM | 38.5379 | 0.9979 | 380.1k |

Therefore, the CM and SAM should be considered more when assigning learnable parameters. This principle was followed when deciding the parameters within the module in this paper.

### *2.5.2.3. Residual scaling*

Residual scaling, the practice of multiplying residuals by a coefficient ranging from 0 to 1, aims to enhance network stability and accuracy. Typically, based on prior research, this scaling factor was set between 0.1 and 0.3 [13,41]. Nevertheless, this approach lacks adaptability as it uniformly applies the same factors across both shallow and deep network layers, potentially leading to over- or under-scaling that can hinder network performance.

In contrast, JASRB introduce two attention modules that learn the residual scaling value dynamically. These modules serve dual purposes: as weight matrices of attention mechanisms and as residual scaling factors.

To validate the superiority of the proposed method over empirical coefficient settings, several comparative experiments were conducted: A. Using learnable residuals scaling of SAM and CAM (the proposed method); B. Replacing the spatial attention with a scaling layer with a fixed parameter of 0.2, which corresponds to a small scaling coefficient; C. Replacing the SAM with a scaling layer with a fixed parameter of 0.8, which corresponds to a large scaling coefficient; D. Without any attention, connect the residuals directly to a scaling layer with a fixed parameter of 0.2. E. Use CAM only (RCAB method). All tests only employed one module to ensure identical amounts of learnable parameters across methods, resulting in minimal variations in RSNR and SSIM, while still illustrating the impact of incorporating various methods.

The results in Table 5 show that method A achieved the best results, indicating that the proposed method was the best. Meanwhile, the comparison of methods B and D revealed that the effect of using attention was consistently better than that of setting residual scaling by experience. The comparison of methods B and C shows that a lower scaling factor was generally better, which was the similar trend in the reference [40]. Method E revealed that single-channel attention struggled to learn image features effectively, and its residual scaling was also inadequate.

**Table 5**

Comparison of different residual scaling methods

| Methods | PSNR(dB) | SSIM |
|---|---|---|
| A | 38.7357 | 0.9981 |
| B | 38.0389 | 0.9977 |
| C | 37.9230 | 0.9976 |
| D | 36.7100 | 0.9975 |
| E | 34.5550 | 0.9954 |

In addition, if only a single type of attention mechanism was used (Method E), the residual scaling coefficient was directly the attention weight matrix. Only one attention matrix may still lead to a high scaling coefficient, and these points may still not be fully scaled. In the proposed approach, two attention matrices were multiplied and the resulting joint attention matrix consisted of very small values. This ensured that the pixels not been fully scaled were further scaled. Essentially, the joint matrix imposed a stronger scaling effect on those features, and this scaling effect was adaptively learned.

Fig. 6 and Fig. 7 depict the training progress of the refined model following ablation test optimization. During the initial training phase, both the loss function and root mean square error undergo fluctuations before steadily declining to zero. Additionally, the high validation frequency of the validation set (every 50 iterations) appears to cause slight oscillations, but these remain within an acceptable range of 0~1. Hence, it can be inferred that the network has been effectively trained.

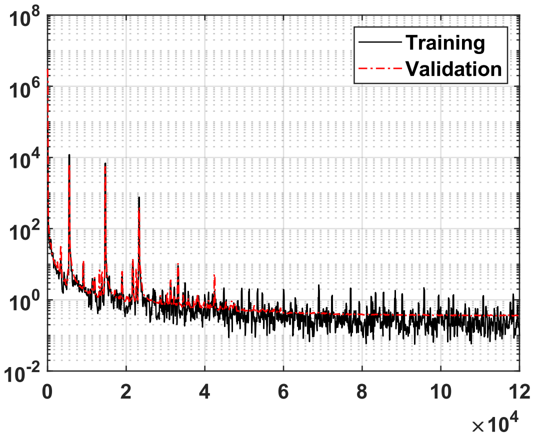


**Fig. 6** Training process: loss function

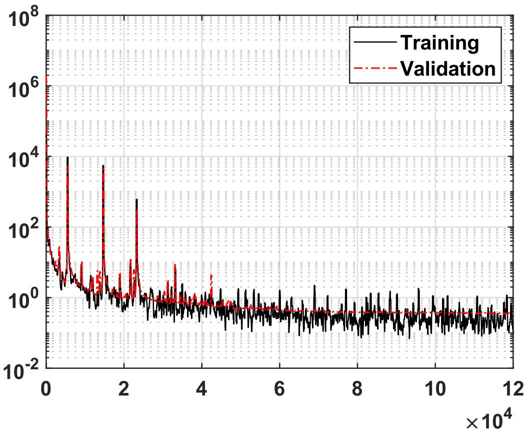


**Fig. 7** Training process: Root mean square error

### *2.6. Training results*

The proposed method was compared with four super-resolution techniques: bicubic [3], EDSR [13], PAN [42], and GSRTI [43] which are renowned for their performance in image enhancement. These methods incorporate various modules such as residual blocks and attention mechanisms.

EDSR, PAN and GSRTI were replicated using Matlab deep network designer, adhering to the architecture of the original paper. The main difference between the original model and our replication is the use of MSE loss, as we prioritized architectural performance and considering that the PSNR and SSIM by MSE loss is comparable to other loss functions [44]. All models used the training and test sets outlined in section 2.4.2, undergoing 100 training rounds totaling 240,000 iterations.

Table 6 and Fig. 8 present reconstruction results alongside others. The proposed method excelled in all evaluation metrics. Bicubic and GSRTI showed limited improvement. While EDSR and PAN offered some clarity, their PSNR and SSIM scores were notably lower due to color correspondence issues. Specifically, in the RGB color model, the same color can be represented by multiple different RGB values. The evaluation metrics revealed that the EDSR and PAN models learned incorrect dolor correspondence relationships. Conversely, the proposed method excelled in indicator results and image clarity with no color correspondence errors and closely resembling the HR image, and exhibited competitive performance.

**Table 6**

Comparison between the SR method in this paper and other methods

| Methods | PSNR(dB) | SSIM | NIQE |
|---|---|---|---|
| Bicubic | 32.4 | 0.9938 | 5.6395 |
| EDSR | 24.457 | 0.9804 | 5.0797 |
| PAN | 21.951 | 0.9476 | 4.9926 |
| GSRTI | 32.635 | 0.9942 | 5.1761 |
| Proposed | 40.129 | 0.9987 | 4.8965 |

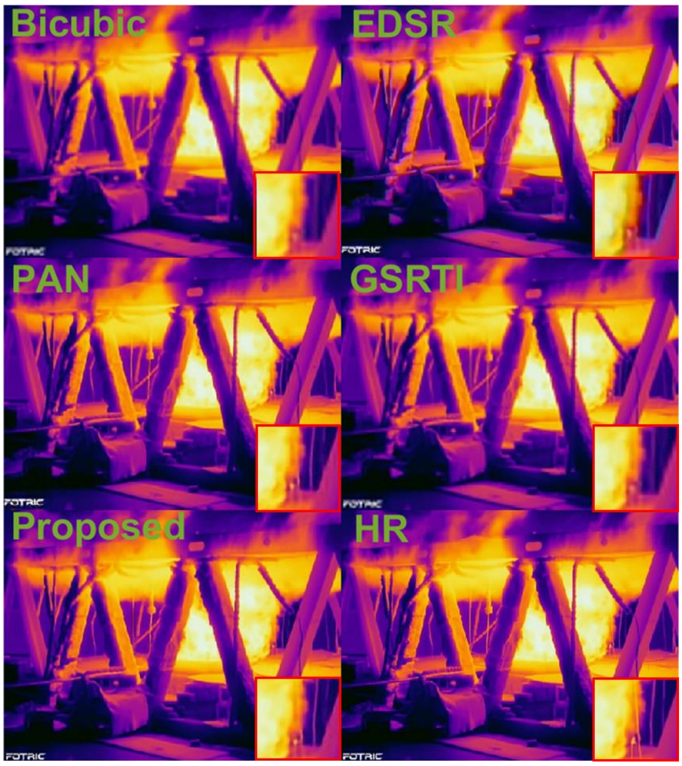


**Fig. 8** Sample of the reconstructed image

## 3. Prediction of image temperatures

### *3.1. Prediction method*

Enhancing image resolution alone is insufficient to justify the use of SR images for temperature measurement. Consequently, there is an urgent need to devise a temperature measurement method for SR images that spans a wide range. In current unenclosed space fire tests, the peak temperature, influenced by fuel, seldom exceeds 1200°C [1,45]. Thus, a temperature measurement range of 400 to 1200 °C was selected.

In thermal imaging, a given color can signify varying temperatures. This is due to the fixed color scale but adjustable temperature range during imaging. This complexity makes it challenging for CNN to predict temperature distributions across various scenes. Additionally, the efficiency of CNN is constrained by its architecture, leading to slow training and application. Contrary to CNN, Artificial Neural Networks (ANNs) only need some pixels from the thermal image along with their corresponding temperatures to learn the color-temperature relationship. This simplified approach can then be directly applied to measure specific points in positional thermal images, enhancing deployment efficiency and streamlining the temperature prediction process compared to CNN.

An ANN-based method was proposed. Since the temperature range involved in high-temperature scenarios is extensive, using one neural network to predict all pixels to be measured within 1200°C directly is tough, and its training requires too many samples. As shown in Fig. 9, the temperature samples were divided into eight segments from 400°C to 1200°C with a temperature difference of 100°C. Correspondingly, eight simple ANN models were established, and each was tailored to correspond with a specific temperature segment.

Meanwhile, datasets were made for each ANN model. The dataset was then divided into many small sections every 100°C. For images falling within a specific temperature difference segment, pixels are consistently classified into corresponding small temperature sections based on their actual temperature, while also saving pixel color and the temperature information of the entire image. This information serves as the input and output for the ANN, as further detailed in section 3.2.

The case in Fig. 10 shows the overall progress of the section of 900°C. The first small square in the upper left corner represents the data with a small temperature segment of 900~1000°C, which is the highest temperature range in this situation. The lowest temperature in the lower right corner is from 0 to 100°C, which represents the lowest temperature range.

Datasets were made and trained separately for HR and SR images, and the corresponding model was trained using the corresponding sample sets to ensure the accuracy of the network.

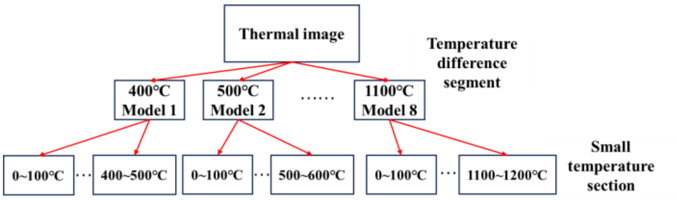


**Fig. 9** Temperature division scheme

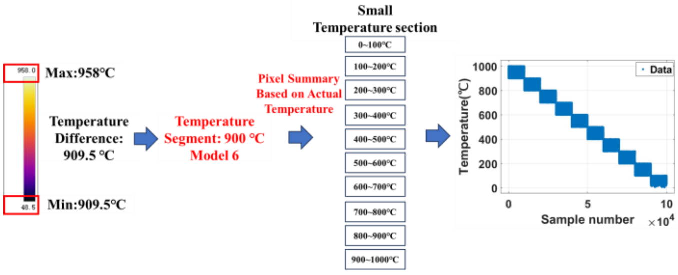


**Fig. 10** Sample selection process of each small temperature range in the 900°C temperature difference section

## 3.2. Network architecture

In every ANN model, the input consists of the brightness value of a specified pixel in the thermal image, as well as the maximum and minimum temperature values from the corresponding thermal image, and the model outputs the estimated temperature of that specified point. Temperature recognition was performed using a single hidden layer feedforward neural network, the structure of which is shown in Fig. 11.

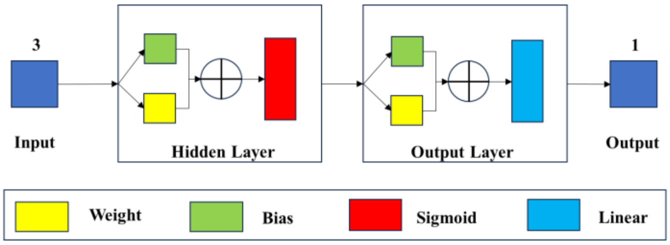


**Fig. 11** Artificial neural network architecture

Given the wide temperature range and the need to train multiple neural networks with large sample sizes, efficiency becomes crucial. The Levenberg-Marquardt backpropagation algorithm [46] was utilized in this study, combining the best gradient descent and Gauss-Newton methods. This approach ensures both rapid convergence near the optimal solution and robustness even when far off, and it uses the MSE loss as the loss function. The algorithm is calculated as follows:

$$X_{k+1}=X_k-[J^TJ+\mu I]^{-1}J^Te \tag{13}$$

where $X_k$ and $X_{k+1}$ represent the current and next solution vectors respectively; $J$ represents the Jacobian matrix; $\mu$ is the damping factor; $I$ represents the Identity matrix; $e$ represents the error vector.

The cornerstone of this approach lies in the utilization of a damping factor $\mu$ for adaptive adjustment during iterations. When $\mu$ is set to zero, the algorithm transforms into the Newton method, renowned for its rapid and precise convergence near the error minimum. Conversely, as $\mu$ increases, the algorithm shifts towards the gradient descent method. The ultimate objective is to transition to the Newton method as swiftly as possible, with $\mu$ undergoing adaptive reduction in each iteration, thereby ensuring efficiency and accuracy in neural network training.

## 3.3. Interpretation of neural network model

The black box characteristic of artificial neural networks limits their transparency, as they can only produce predictions without explaining how those predictions are generated. ANNs cannot reveal the specific influence and rationale behind each effect of the input on the prediction outcomes, thereby challenging the trustworthiness of the model [47]. The Shapley additive interpretation proposed by Lundberg et al. [48] can quantify the contribution of each feature in the ANN and consider the mutual influence between features. The Shapley value of the $i$-th feature of query point $x$ is defined by the value function $v$:

$$\varphi_i(v_x)=\frac{1}{M}\sum\nolimits_{S\subseteq\mu\{i\}}\frac{v_x(S\cup\{i\})-v_x(S)}{\frac{(M-1)!}{|S|!(M-|S|-1)!}} \tag{14}$$

where $M$ is the number of features; $\mu$ is the set of all features; |S| is the number of elements in set $S$; $v_x(S)$ is the value function of the feature of measurement point $x$ in set $S$, and its value represents the expected contribution of the features in $S$ to the prediction of measurement point $x$.

The Shapley value is primarily utilized for explaining local samples, which involves multiple neural networks with tens of thousands of samples each, incurs extremely high computational costs. To circumvent this, four extreme temperature scenarios was used to interpret the relationship of the input arguments and the output argument in the proposed neural network models: a. A neural network temperature segment operating within 400°C (spanning from 0~500°C), specifically measuring temperatures within the range of 400 to 500°C. b. The same neural network temperature segment as in scenario a, but now measuring temperatures only within the range of 0~100°C. c. A neural network temperature segment operating within 1100°C (spanning 0~1200°C), particularly attending to temperatures within the bracket of 1100~1200°C. d. The neural network temperature segment from scenario c, but with a focus on temperatures within the interval of 0~100°C.

Fig. 12 shows the Shapley values for the features of the SR neural network model in the four scenarios. These values revealed the significance of the three inputs of the neural network. The color, crucial in thermal imaging for temperature detection, was reflected by brightness, which had high Shapley values. The network also used max and min temperature values to control the temperature measurement range. While these two features cannot determine temperature directly, they played a key role in controlling the temperature output results in different temperature ranges.

In scenarios a and c, where the temperature measurement ranges align with the highest temperatures reachable by their respective neural networks, the neural network inputs positively impact the output, driving the network to output high-temperature predictions. Conversely, in the extreme low-temperature cases of b and d, the Shapley values are uniformly negative, indicating that the input features decrease the output temperature.

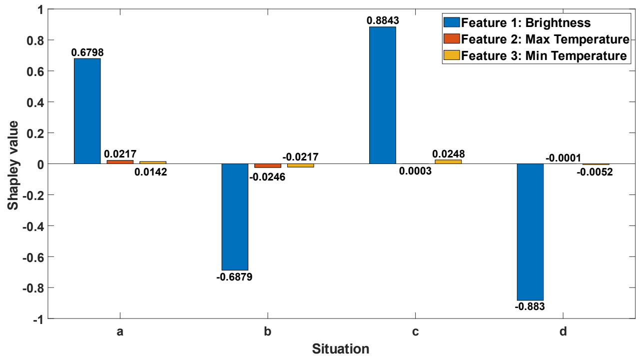


**Fig. 12** Shapley values of SR model under four extreme temperature scenarios

### *3.4. Training of the artificial neural network*

#### *3.4.1. Training details*

Unlike super-resolution architecture training, the image temperature measurement dataset demands thermal images spanning various temperature ranges. A total of 233 high-temperature flame scene images from the fire tests in section 2.4.2 were selected, ensuring 10,000 samples covering all small temperature segments within each range.

The emissivity of the thermal image remains set at 0.95. This is because, in this method, instead of focusing on a specific object measurement point, the entire spatial temperature field is measured, with the emissivity representing that of the air. Moreover, the primary aim of the methodology of this paper is to present a color temperature prediction approach, and a fixed emissivity effectively streamlines the temperature measurement process.

For each sub-segment, the data was split into training (70%), validation (15%), and testing (15%) sets. This section presents prediction results using these test sets. Both HR and SR image datasets were created using identical methodologies and images, with the HR dataset facilitating network optimization.

Before training, the normalization function was used in Matlab to normalize the data set within ±1. The feedforward neural network was carried out in the fitting toolbox in Matlab R2022b, and the network was trained for a maximum of 1000 rounds. Several early termination conditions were set, and these specific termination conditions can be found in the function “trainlm” in Matlab. All trainings were performed three times to minimize the deviation caused by different initial random parameters.

#### *3.4.2. Evaluation indicators*

Four indicators were used to comprehensively compare the overall error, prediction stability, and number of outliers in the dataset. RMSE serves as a widespread metric for assessing regression models, efficiently identifying substantial errors. A minimal RMSE is desirable for an optimal model. The percentage of sample errors within ±30°C provides a comprehensive overview of the prediction sample, emphasizing that a high-quality prediction requires a balance between low error rates and consistent predictability. The MAE indicator resembles RMSE in measuring network errors but offers a simpler calculation and reduced sensitivity to outliers. The point where the sampling error is beyond ±100°C is called the flawed point. These points cannot be accurately predicted by the network model, which can indicate the approximate error limit of the network.

#### *3.4.3. Optimization of neuron number*

The number of neurons is crucial to the network prediction results, and an appropriate number of neurons can improve the network prediction accuracy. In this section, the number of neurons was selected as 3, 5, and 10 to represent fewer, medium, and more neurons, respectively. Tests were performed on the HR image dataset, and the evaluation indicators are the main indexes. Figs. 13 and 14 compare the training results. The 3-neuron model showed the best accuracy and stability in all evaluation indicators. Using more neurons for simple tasks can easily cause a mismatch between network and task complexity, thereby reducing the training effect. In this task, using a low-complexity three-neuron model demonstrated the highest accuracy and stability.

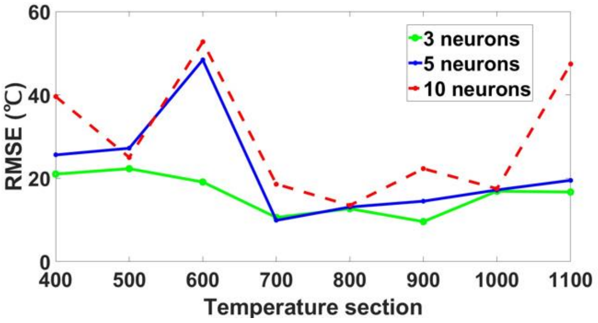


**Fig. 13** RMSE of different neurons

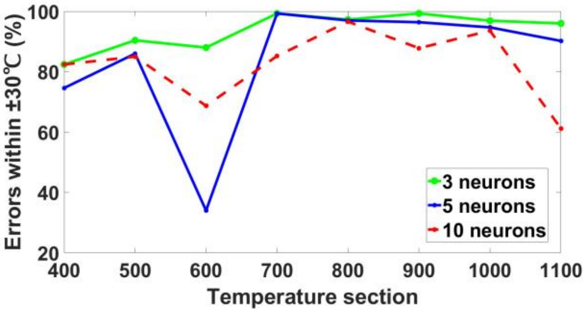


**Fig. 14** Proportion of samples within ±30°C of prediction error in different neurons

Fig. 15 and Fig. 16 depict the synchronized convergence of the training, test, and validation sets after optimizing the neuron count in both the lowest (400°C) and highest (1100°C) temperature difference segments. No adverse effects, such as overfitting, were observed. The training process of the remaining temperature difference sections was roughly the same. Due to space limitations, only the training process of the lowest and highest temperature difference sections is shown here.

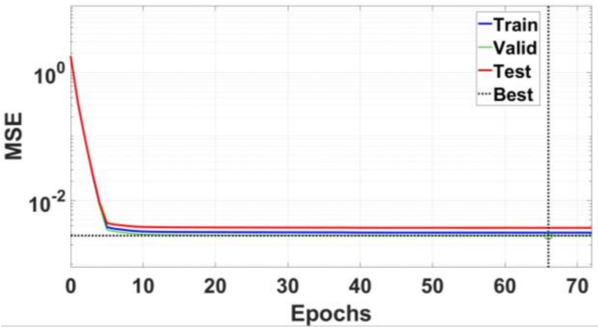


**Fig. 15** Proportion of samples within ±30°C of prediction error in different neurons

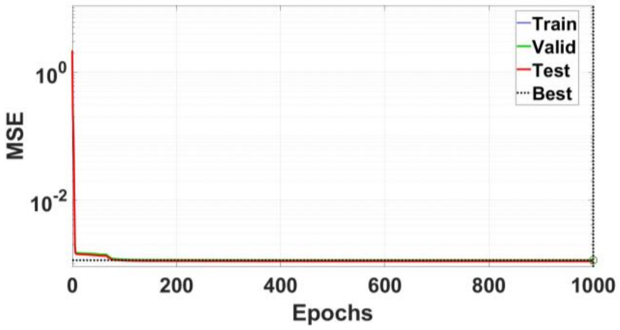


**Fig. 16** Proportion of samples within ±30°C of prediction error in different neurons

### 3.5. Training results

The results in Table 7 show the prediction accuracy of both the HR and SR images using the proposed method. Please note that the accuracy of the neural network, trained using HR images, solely serves as a benchmark for comparison with the network trained using SR images. Simultaneously, the HR image training outcome represents the peak prediction accuracy attainable by the proposed color temperature prediction method for high-definition images. In reality, the HR image remains unknown, yet the SR image generated by this method demonstrates prediction accuracy similar to the HR image, which was revealed by the "Deviation" in Table 7. This suggests that the CNN super-resolution framework described in section 2 has successfully learned to achieve remarkable accuracy in representing color temperature relationships.

The average MAE was 14.2°C, with over 80% of samples falling within an error range of ±30°C. Please note that the temperature measurement range examined in this paper reaches up to 1200°C, and an error of this magnitude is clearly tolerable. This underscores the efficacy of our proposed method and the optimized ANN for high-temperature prediction tasks.

The error in color temperature prediction by the neural network may stem from the utilization of distinct prediction principles: formulas are typically employed for thermal cameras, whereas neural networks are used in the proposed method, potentially causing color-temperature inconsistencies. In addition, prediction accuracy was slightly lower above 900°C, likely due to environmental factors affecting the thermal camera.

Regression analysis on the neural network reveals that the predictions align accurately with target results, as is shown in Fig. 17. Most samples fell within the 95% confidence interval, indicating good responsiveness. In summary, models trained with SR images showed no significant degradation in prediction results.

**Table 7**
HR/SR image prediction accuracy

| Segemnts （℃） | RMSE(°C) | | MAE(°C) | | Error in ±30(%) | | Number of faults | |
|---|---|---|---|---|---|---|---|---|
| | HR | SR | HR | SR | HR | SR | HR | SR |
| 1100 | 21 | 22.4 | 16.1 | 17.4 | 82.4 | 81.5 | 8 | 24 |
| 1000 | 22.3 | 26.5 | 19.5 | 16 | 90.4 | 81.6 | 134 | 129 |
| 900 | 19.1 | 23.7 | 15.5 | 18.6 | 88 | 80.6 | 1 | 4 |
| 800 | 10.6 | 15.1 | 8.3 | 11.8 | 99.3 | 95.4 | 0 | 1 |
| 700 | 12.7 | 19.1 | 9 | 13.8 | 97.3 | 90.3 | 6 | 24 |
| 600 | 9.6 | 16.3 | 7.5 | 11.5 | 99.3 | 93.4 | 0 | 9 |
| 500 | 16.9 | 15.4 | 12.3 | 10.3 | 96.9 | 94.1 | 0 | 5 |
| 400 | 16.7 | 16.9 | 14.6 | 13.9 | 96 | 94.8 | 0 | 3 |
| Average | 16.1 | 19.4 | 12.8 | 14.2 | 93.7 | 89 | 19 | 25 |
| Deviation | 3.3 | | 1.4 | | 4.7 | | 6 | |

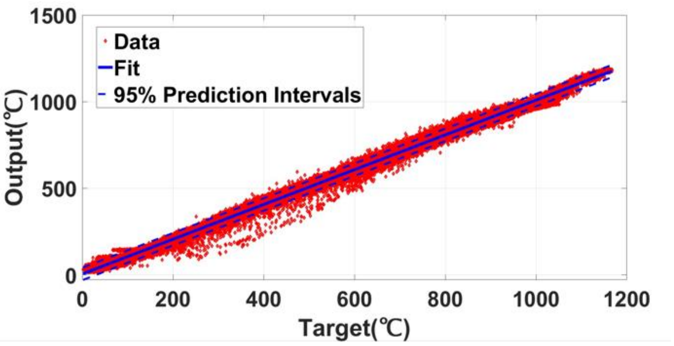


**Fig. 17** Linear regression of SR image model

### 3.6. Model comparison

In section 3.5, the image temperature measurement technique exhibited good accuracy when being trained via neural networks. Nevertheless, distinct machine learning models often excel in specific scenarios. To underscore the robustness of the proposed temperature prediction approach, a comprehensive evaluation was conducted using various machine learning models, including SVM [49], kernel regression [50], regression tree [51], and linear regression [52]. For fairness, the same optimization algorithms, training, and test sets were employed as outlined in sections 3.4.

Figs. 18 and 19 revealed that the SVM and kernel regression exhibited higher errors and regression tree and linear regression demonstrated reasonable accuracy. However, considering prediction accuracy and stability across all temperature segments, a notable disparity remained between existing models and the proposed neural network. The poor performance of these four models stemmed from a mismatch between their complexity and the actual task complexity, resulting in either redundancy or waste.

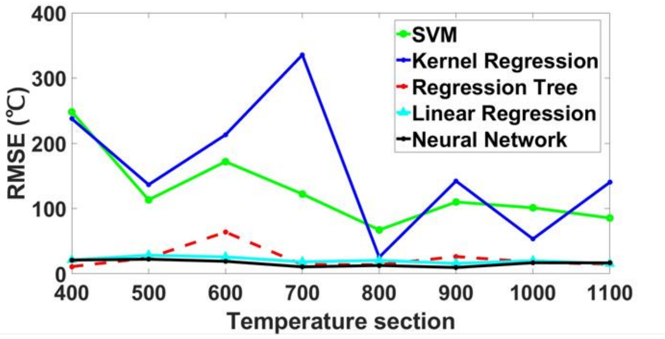


**Fig. 18** RMSE of different machine learning models

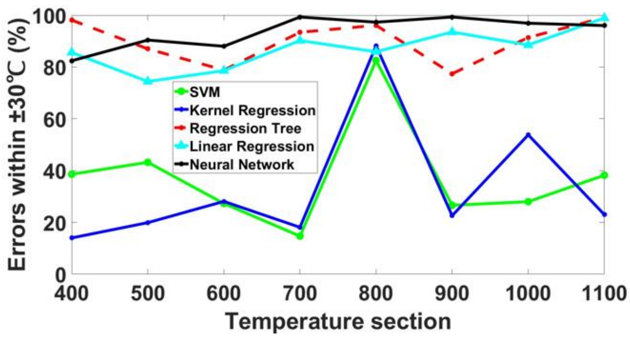


**Fig. 19** Proportion of samples within the error of different models within ±30°C

## 4. Test validation

To validate the wide applicability of the proposed method in high-temperature environments, an extra fire test was conducted an extra fire test to assess the performance of both the super-resolution and prediction networks.

### 4.1. Test environment

Fig. 20 depicts the entire test scene for a set of cables coated with fire-resistant and flame-retardant sealant and with two ends on support. An oil basin filled with 30L of diesel oil was centrally positioned at the bottom and ignited. The thermal imager, situated 5 meters opposite the cable (Fig. 20d), was set to record the scene. The environmental temperature was set to 20°C with external optical temperature, and the relative humidity of 0.5, emissivity of 0.95, and picture noise reduction were used. Throughout the test, a thermal camera was applied to observe temperatures ranging from 0 to 1250°C.

Test environment Shooting scene Thermal image during test

**Fig. 20** Realistic fire test scenario of bridge cables

All SR images underwent processing based on the method outlined in section 2, while test sets for HR and SR images were generated using the color temperature measurement model as detailed in section 3.1. Given the brief shooting duration, each experiment produced a limited dataset. The 200 samples per temperature range were selected for testing. This resulted in a total of 800 test samples for a temperature difference segment of 400°C and 2400 samples for a segment of 1100°C.

### *4.2. Test results*

By comparing the predicted outcomes with the actual thermal image outputs, the discrepancies are delineated in Table 8 and illustrated in Figs. 21 and 22. A slight decrease in predicted performance was found compared to previous scenarios. While RMSE, MAE, and the percentage of errors were within 30°C showing minor changes in degradation. Specifically, RMSE remained stable at approximately 20°C, and the proportion of samples with prediction errors within 30 °C only dropped slightly in certain sections, ranging between 70% and 80%. This performance degradation was expected in new scenarios as the network model may not have learned all relevant information. Nevertheless, the minimal decrease demonstrates the good generalization and accuracy of the proposed HR and SR temperature prediction model.

**Table 8**
HR/SR image prediction accuracy under the new scenario

| Segemnts（℃） | RMSE(°C) | | MAE(°C) | | Error in ±30(%) | | Number of faults | |
|---|---|---|---|---|---|---|---|---|
| | HR | SR | HR | SR | HR | SR | HR | SR |
| 1100 | 25.5 | 21.5 | 19.8 | 18.1 | 71.2 | 85.7 | 0 | 0 |
| 1000 | 22.9 | 25.0 | 18.7 | 19.6 | 83 | 75.8 | 0 | 0 |
| 900 | 11 | 20.5 | 8.6 | 16.1 | 98.9 | 85 | 0 | 0 |
| 800 | 14.8 | 18.2 | 12.2 | 15.1 | 97.6 | 90.6 | 0 | 0 |
| 700 | 12.7 | 19.8 | 9.9 | 15.6 | 97.5 | 87.9 | 6 | 0 |
| 600 | 13 | 17.7 | 9.7 | 13.6 | 96.6 | 80 | 0 | 0 |
| 500 | 25.6 | 15.4 | 23.3 | 11.4 | 70.2 | 94.5 | 0 | 0 |
| 400 | 20.9 | 22 | 19.8 | 19.4 | 94.4 | 84.4 | 0 | 0 |
| Average | 18.3 | 20 | 15.2 | 16.1 | 88.7 | 85.5 | 0 | 0 |
| Deviation | 1.7 | | 0.9 | | 3.2 | | 0 | |

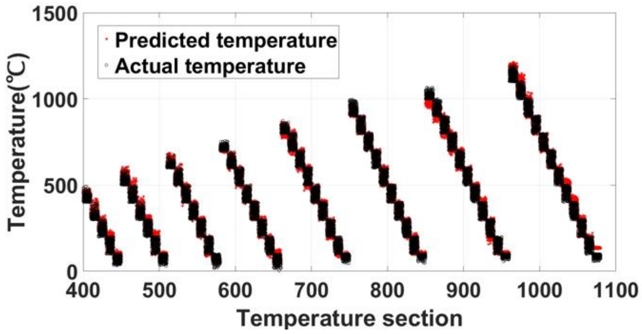


**Fig. 21** HR model prediction results in new scenarios

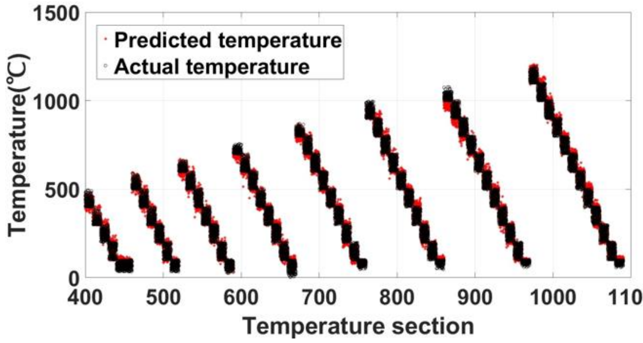


**Fig. 22** SR model prediction results in new scenarios

### *4.3. Discussion*

In section 3.6, it was observed that both the SVM regression model and the linear regression model can predict temperatures, albeit with slightly lower accuracy than the proposed model in the test set. To further evaluate the performance of the proposed method, the SR models were employed to specifically compare the errors of these ANN models.

The results can be found in Figs. 23 and 24. The proposed method consistently demonstrated the highest accuracy across most temperature ranges, with minimal and stable error fluctuations. Although the regression tree model showed good accuracy in certain temperature ranges, it exhibited large prediction errors in the 600°C and 1000°C ranges, indicating potential instability during flame ignition or sustained burning. Given reliable temperature data, the linear regression model was proved to be significantly less accurate than the regression tree model, rendering them unsuitable for practical applications.

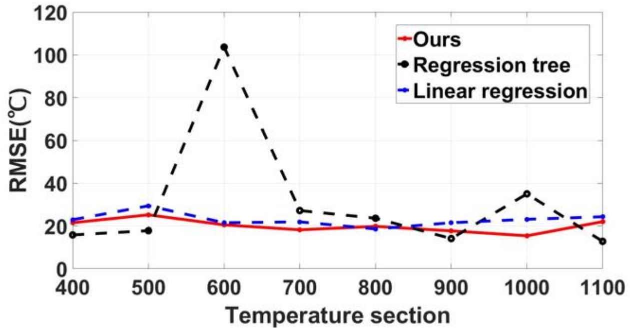


**Fig. 23** RMSE of different machine learning models in new scenarios

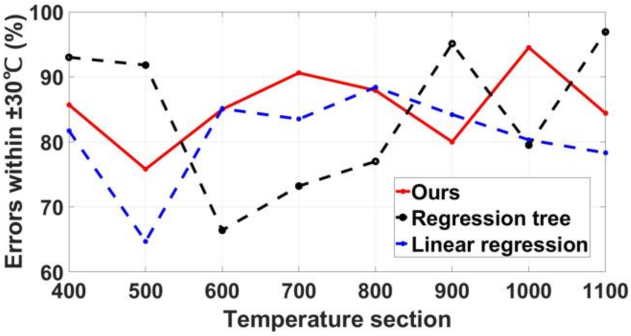


**Fig. 24** Proportion of samples within the error of different machine learning models within ±30°C in new scenarios

## 5. Conclusions

This paper proposed a CNN super-resolution architecture tailored for thermal images in fire scenes, along with a color temperature prediction method for enhanced images. The superiority of the proposed method was validated against fire tests results. The following conclusions can be drawn:

(1) The JASRB module, integrating spatial and channel attention, was proposed for constructing the CNN architecture. This integration notably enhanced the super-resolution of images. With four modules in each core part, the architecture optimized super-resolution, achieving a PSNR of 40.214 dB and an SSIM of 0.9987. It not only refined image clarity but learned accurate color-temperature mappings.

(2) A color temperature measurement method was proposed for wide temperature ranges, utilizing an interpretable ANN. This network incorporated brightness and temperature data from designated points to enhance temperature predictions. Shapley analysis revealed that brightness significantly influenced the predictions of the network. Optimal performance was observed with three neurons in the ANN, yielding RMSE of 16.1°C and 19.4°C for HR and SR images, respectively. Notably, nearly 90% of test sample errors fell within ±30°C.

(3) By integrating super-resolution with the image temperature measurement model, the network effectively accomplished temperature prediction tasks. Additional fire test validated that the unknown thermal images maintained high prediction accuracy after super-resolution

processing. Furthermore, it is worth noting that the three-neuron neural network approach demonstrated superior prediction accuracy compared to other machine learning methods.

## Acknowledgments

The work presented in this paper was mainly supported by the National Natural Science Foundation of China with grants 52127814 and 52078478, and the National Fire and Rescue Administration with grant 2023XFCX18.

## References

[1] Lou, G. B., Hou, J., Qi, H. H., Jiang, Y., Zhong, B., Pang, Y., and Li, G. Q. Experimental, numerical and analytical analysis of a single-span steel portal frame exposed to fire. Engineering Structures, 302, 117417, 2024.

[2] Chudasama, V., Patel, H., Prajapati, K., Upla, K. P., Ramachandra, R., Raja, K., and Busch, C. Therisurnet - a computationally efficient thermal image super-resolution network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 86-87), 2024.

[3] Keys, R. Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing, 29(6), 1153-1160, 1981.

[4] Kim, K. I., and Kwon, Y. Single-image super-resolution using sparse regression and natural image prior. IEEE transactions on pattern analysis and machine intelligence, 32(6), 1127-1133, 2010.

[5] Irani, M., and Peleg, S. Improving resolution by image registration. CVGIP: Graphical models and image processing, 53(3), 231-239, 1991.

[6] Li, J., Brown, C., Dzikowicz, D. J., Carey, M. G., Tam, W. C., and Huang, M. X. Towards real-time heart health monitoring in firefighting using convolutional neural networks. Fire Safety Journal, 140, 103852, 2023.

[7] Jiang, Y., Liu, Y., Zhan, W., and Zhu, D. Improved Thermal Infrared Image Super-Resolution Reconstruction Method Base on Multimodal Sensor Fusion. Entropy, 25(6), 914, 2023.

[8] Hodges, J. L., Lattimer, B. Y., and Luxbacher, K. D. Compartment fire predictions using transpose convolutional neural networks. Fire Safety Journal, 108, 102854, 2019.

[9] Wang, Z., Chen, J., and Hoi, S. C. Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3365-3387, 2020.

[10] Kim, J., Lee, J. K., and Lee, K. M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654), 2016.

[11] Kim, J., Lee, J. K., and Lee, K. M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637-1645), 2016.

[12] Tai, Y., Yang, J., and Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3147-3155), 2017.

[13] Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144), 2017.

[14] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... and Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883), 2016.

[15] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 286-301), 2018.

[16] Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19), 2018.

[17] Hu, J., Shen, L., and Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141), 2018.

[18] Choi, Y., Kim, N., Hwang, S., and Kweon, I. S. Thermal image enhancement using convolutional neural network. In 2016 IEEE/RSJ international conference on intelligent robots and systems (IROS) (pp. 223-230). IEEE, 2016.

[19] Lee, K., Lee, J., Lee, J., Hwang, S., and Lee, S. Brightness-based convolutional neural network for thermal image enhancement. IEEE Access, 5, 26867-26879, 2017.

[20] Liu, S., Yang, Y., Li, Q., Feng, H., Xu, Z., Chen, Y., and Liu, L. Infrared image super resolution using gan with infrared image prior. In 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP) (pp. 1004-1009). IEEE, 2019.

[21] Rivadeneira, R. E., Sappa, A. D., and Vintimilla, B. X. Thermal Image Super-resolution: A Novel Architecture and Dataset. In VISIGRAPP (4: VISAPP) (pp. 111-119), 2020.

[22] Hwang, S., Park, J., Kim, N., Choi, Y., and So Kweon, I. Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1037-1045), 2015.

[23] He, Z., Tang, S., Yang, J., Cao, Y., Yang, M. Y., and Cao, Y. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. IEEE transactions on circuits and systems for video technology, 29(8), 2310-2322, 2018.

[24] Miklavec, A., Pušnik, I., Batagelj, V., and Drnovšek, J. A large aperture blackbody bath for calibration of thermal imagers. Measurement Science and Technology, 24(2), 025001, 2012.

[25] Pelzmann, T., Dupont, F., Sauté, B., and Robert, É. Two-color pyrometry for backface temperature and emissivity measurement of burning materials. International Journal of Thermal Sciences, 184, 107894, 2023.

[26] Zamarripa-Ramírez, J. C. I., Moreno-Hernández, D., and Martinez-Gonzalez, A. Simultaneous measurement of temperature and color spectrum of axisymmetric premixed flames using digital laser speckle photography and an image processing approach. Measurement Science and Technology, 32(10), 105903, 2021.

[27] Shan, L., Huang, H., Hong, B., Zhao, J., Wang, D., and Kong, M. Temperature measurement method of flame image fusion with different exposures. Energies, 13(6), 1487, 2020.

[28] Zhong, Y., Li, X., Yang, Z., Liu, X., and Yao, E. A model for minimum ignition energy prediction of sugar dust clouds based on interactive orthogonal experiments and machine learning. Fire Safety Journal, 104111, 2024.

[29] Qu, N., Li, Z., Li, X., Zhang, S., and Zheng, T. Multi-parameter fire detection method based on feature depth extraction and stacking ensemble learning model. Fire Safety Journal, 128, 103541, 2022.

[30] Özyurt, O. The effect of time correlation function, Sauter mean diameter and asymmetry ratio on machine learning classifiers for particle discrimination by using light scattering. Fire Safety Journal, 141, 104002, 2023.

[31] Qu, X. S., Deng, Y. X., and Sun, G. J. INVESTIGATION ON BEHAVIOR OF STEEL CABLES SUBJECT TO LOCALIZED FIRE IN LARGE-SPACE BUILDINGS. ADVANCED STEEL CONSTRUCTION, 20(1), 1-11, 2024.

[32] Nan, Z., Orabi, M. A., Huang, X., Jiang, Y., and Usmani, A. Structural-fire responses forecasting via modular AI. Fire Safety Journal, 140, 103863, 2023.

[33] Couto, C., Tong, Q., and Gernay, T. Predicting the capacity of thin-walled beams at elevated temperature with machine learning. Fire Safety Journal, 130, 103596, 2022.

[34] Sun, B., Liu, X., Xu, Z. D., and Xu, D. BP neural network-based adaptive spatial-temporal data generation technology for predicting ceiling temperature in tunnel fire and full-scale experimental verification. Fire Safety Journal, 130, 103577, 2022.

[35] Chen, L., Zhang, H. Y., Liu, S. W., and Chan, S. L. SECOND-ORDER ANALYSIS OF BEAM-COLUMNS BY MACHINE LEARNING-BASED STRUCTURAL ANALYSIS THROUGH PHYSICS-INFORMED NEURAL NETWORKS. Advanced Steel Construction, 19(4), 411-420, 2023.

[36] Wang, Z. Q., Zheng, Z. Mengxiang, J. W., and Jiang, W. Q. INVERSION METHOD OF UNCERTAIN PARAMETERS FOR TRUSS STRUCTURES BASED ON GRAPH NEURAL NETWORKS. ADVANCED STEEL CONSTRUCTION, 19(4): p. 366-374, 2023.

[37] Sharifi, Y., and Tohidi, S. Ultimate capacity assessment of web plate beams with pitting corrosion subjected to patch loading by artificial neural networks. Advanced Steel Construction, 10(3), 325-350, 2014.

[38] Nair, V., and Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814), 2010.

[39] Kingma, D. P., and Ba, J. Adam: A method for stochastic optimization. arxiv preprint arxiv:1412.6980, 2014.

[40] Mittal, A., Soundararajan, R., and Bovik, A. C. Making a “completely blind” image quality analyzer. IEEE Signal processing letters, 20(3), 209-212, 2012.

[41] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1), 2017.

[42] Zhao, H., Kong, X., He, J., Qiao, Y., and Dong, C. Efficient image super-resolution using pixel attention. In Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16 (pp. 56-72). Springer International Publishing, 2020.

[43] Ismail, I., Yefriadi, Y., Yuhefizar, Y., and Hendri, Z. The Generating Super Resolution of Thermal Image based on Deep Learning. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6(2), 289-294, 2022.

[44] Wang, Y., Zhao, L., Liu, L., Hu, H., and Tao, W. URNet: a U-shaped residual network for lightweight image super-resolution. Remote Sensing, 13(19), 3848, 2021.

[45] Jiang, J., Wang, C., Lou, G., and Li, G. Q. Quantitative evaluation of progressive collapse process of steel portal frames in fire. Journal of Constructional Steel Research, 150, 277-287, 2018.

[46] Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2), 431-441, 1963.

[47] Liu, X., and Aldrich, C. Assessing the influence of operational variables on process performance in metallurgical plants by use of Shapley value regression. Metals, 12(11), 1777, 2022.

[48] Lundberg, S. M., and Lee, S. I. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017.

[49] Andrew, A. M. An introduction to support vector machines and other kernel-based learning methods. Kybernetes, 30(1), 103-115, 2001.

[50] Huang, P. S., Avron, H., Sainath, T. N., Sindhwani, V., and Ramabhadran, B. Kernel methods match deep neural networks on timit. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 205-209). IEEE, 2014.

[51] Loh, W. Y. Regression tress with unbiased variable selection and interaction detection. Statistica sinica, 361-386, 2002.

[52] Street, J. O., Carroll, R. J., & Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. The American Statistician, 42(2), 152-154.