ACRM: Attention Cascade R-CNN with Mix-NMS for Metallic Surface Defect Detection

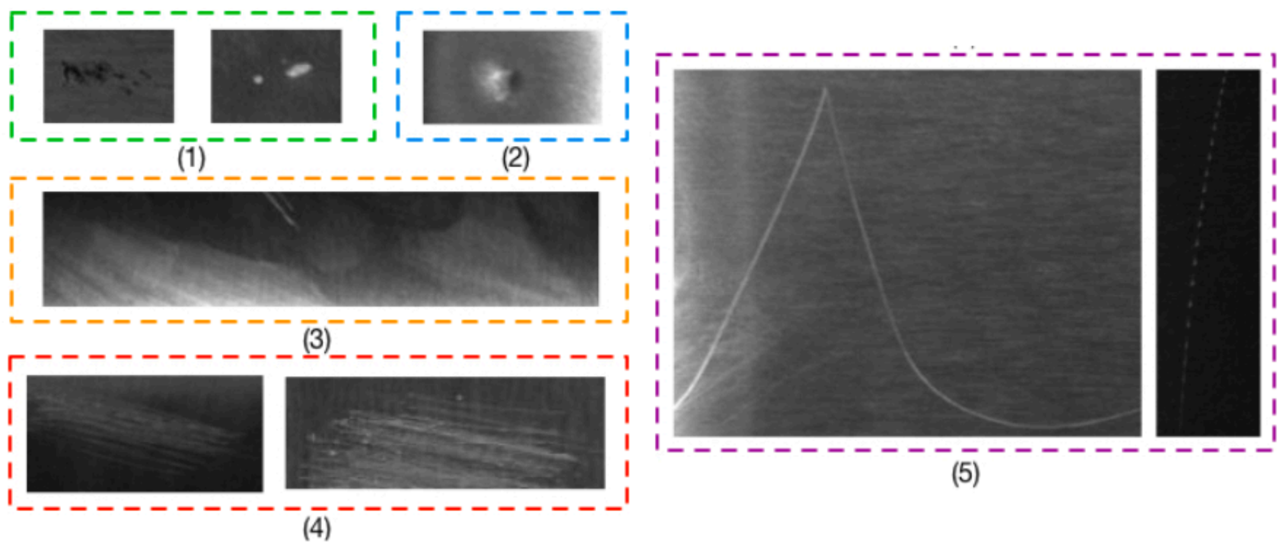
Junting Fang, Xiaoyang Tan

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics

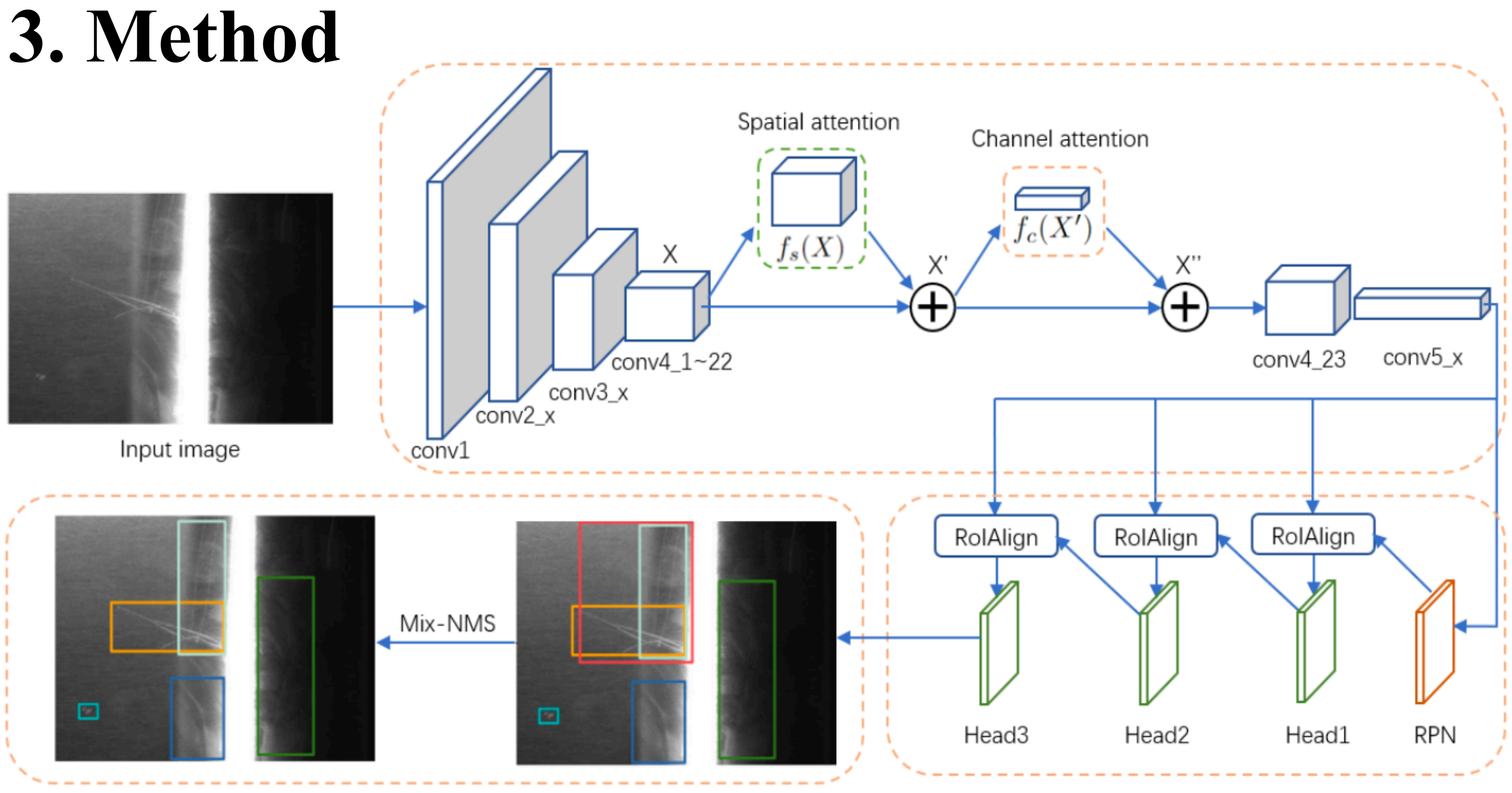
1. Introduction

In this paper, we propose a new defect detection method named Attention-based Cascade R-CNN with Mixed-NMS (ACRM). Firstly, to improve the feature representation of large reflections and defects, we introduce a new lightweight attention block so as to enable CNNs to weigh features according to their importance to the detection task underlying. Secondly, to meet the high detection accuracy requirements in the industry, we use Cascade R-CNN as the baseline, which enables high-quality detection by utilizing cascade head networks for gradually increasing the quality of region proposals. Last but not least, to refine the detected results and reduce the false positive rate, we propose a fast and effective post-processing method, called Mix Non-Maximum Suppression (Mix-NMS), to suppress special redundant detection bounding boxes.

2. Common defects on metallic surface



3. Method



The architecture of Attention Cascade R-CNN with Mix-NMS(ACRM)

The overall attention operation can be formulated as:

$$X' = f_s(X) \oplus X$$
$$X'' = f_c(X') \oplus X'$$

The spatial attention is computed as:

$$f_s(X) = \left\{ W_z \sum_{\forall j}^{N} \frac{f(x_i, x_j)}{C(X)} x_j \right\}_{i=1}^{N} = \left\{ W_z \sum_{\forall j}^{N} \frac{e^{\theta(x_i)^T \phi(x_j)}}{C(X)} x_j \right\}_{i=1}^{N} = W_z(softmax(X^T W_\theta^T W_\phi X)X)$$

The channel attention is computed as:

$$f_c(X') = W_\alpha(f_\delta(X'_{avg}) + f_\delta(X'_{max}))$$

The loss function is:

$$L = \sum_{t=1}^{T} \alpha_t (L_{cls}(c_t, \hat{c}_t) + \beta L_{reg}(r_t, \hat{r}_t))$$

The detailed implementation pipeline of our Mix-NMS algorithm：

Algorithm 1 Mix-NMS

Input: Initial detection boxes set $\boldsymbol{B} = \{b_1, \dots, b_n\}$, corresponding detection scores set $\boldsymbol{S} = \{s_1, \dots, s_n\}$, thresholds $\omega_1, \omega_2, \omega_3$

Output: Detection boxes set $\boldsymbol{D}$, corresponding detection scores set $\boldsymbol{S}$

$\boldsymbol{D} \leftarrow \{\}$
while $\boldsymbol{B} \neq \varnothing$ **do**
  $b_m \leftarrow argmax\ Area(\boldsymbol{B})$
  $\boldsymbol{B} \leftarrow \boldsymbol{B} - \{b_m\}$
  $\boldsymbol{K} \leftarrow \{\}$
  while $b_i\ in\ \boldsymbol{B}$ **do**
    if $IoS(b_i, b_m) \geq \omega_1\ and\ s_i \geq \omega_2$ **then**
      $\boldsymbol{K} \leftarrow \boldsymbol{K} \cup \{b_i\}$
    end if
  end while
  if $len(\boldsymbol{K}) >= 2\ and\ IoU(mbr(\boldsymbol{K}), b_m) > \omega_3$ **then**
    $\boldsymbol{S} \leftarrow \boldsymbol{S} - \{s_m\}$
  else
    $\boldsymbol{D} \leftarrow \boldsymbol{D} \cup \{b_m\}$
  end if
end while

The IoS is the ratio of the overlap area to its own area, while IoU is the ratio of overlap area to union area, formulated as:

$$IoS(b_1, b_2) = S(b_1 \cap b_2)/S(b_1)$$
$$IoU(b_1, b_2) = S(b_1 \cap b_2)/S(b_1 \cup b_2)$$

4. Experiments

The dataset of metallic defect images is provided from a fridge production line using a common camera.

Dataset Count

Dataset	Defect images	No defect images
Training set	1050	1050
Test set	450	450

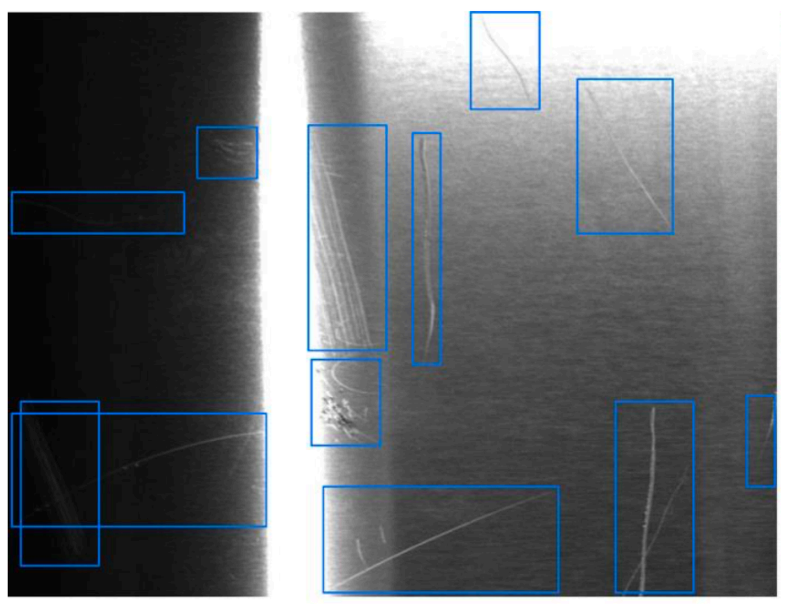


TABLE I
ABLATION STUDY

(a) Combination	AP@.50(%)	AP@.75(%)
baseline	79.2	44.7
+spatial	80.7	44.9
+spatial+channel	**81.9**	**46.8**
+channel+spatial	81.4	46.6
+spatial&channel(in parallel)	81.1	45.7

(c) Module design	AP@.50(%)	AP@.75(%)
baseline	79.2	44.7
+spatial+channel	**81.9**	**46.8**
+NL block [8]	80.6	45.1
+GC block [33]	79.6	45.8

(d) Mix-NMS	AP@.50(%)	AP@.75(%)
baseline+spatial+channel	81.9	46.8
baseline+spatial+channel+Mix-NMS	**82.3**	**46.9**

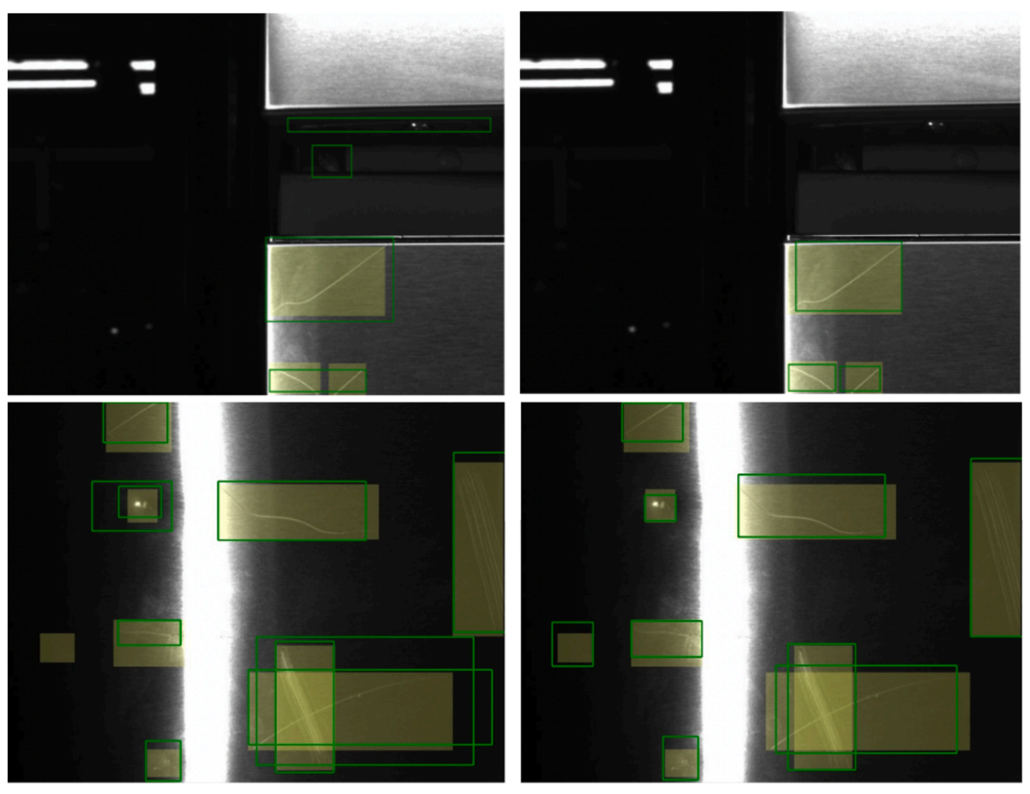
TABLE II
COMPARISON WITH TRADITIONAL METHOD

Method	Accuracy(%)
HOG+SVM [13]	73.3
GLCM+SVM [14]	76.2
ACRM(ours)	**99.7**

TABLE III
COMPARISON WITH STATE-OF-THE-ART DETECTORS

Method	Backbone	AP@.50(%)
YOLOv3 [6]	Darknet-53	67.3
CenterNet [28]	Hourglass-104	23.8
Faster R-CNN [24]	ResNet-101	73.2
TridentNet [25]	ResNet-101	71.9
Cascade R-CNN [10]	ResNet-101	74.9
Wen et al. [16]	26 layers CNN	69.1
ACRM(ours)	Attention ResNet-101	**82.3**

Defect detection results for sample images of Faster R-CNN and ACRM. Left: Faster R-CNN; Right: ACRM. The green boxes are predicted bounding boxes and the yellow boxes are ground truth annotations.



5. Conclusion

In this paper, we have proposed an attention-based Cascade R-CNN with Mix-NMS (ACRM) for high-quality metallic sur- face defect detection. By capturing long-distance dependencies and interdependence between the channels of features, the attention module is shown to succeed in challenging industrial scenarios. Furthermore, the Mix-NMS technique is proposed to remove false positive predictions. Experimental results show that our ACRM achieves satisfactory performance, and it is promising that ACRM can be used in practical industrial applications.