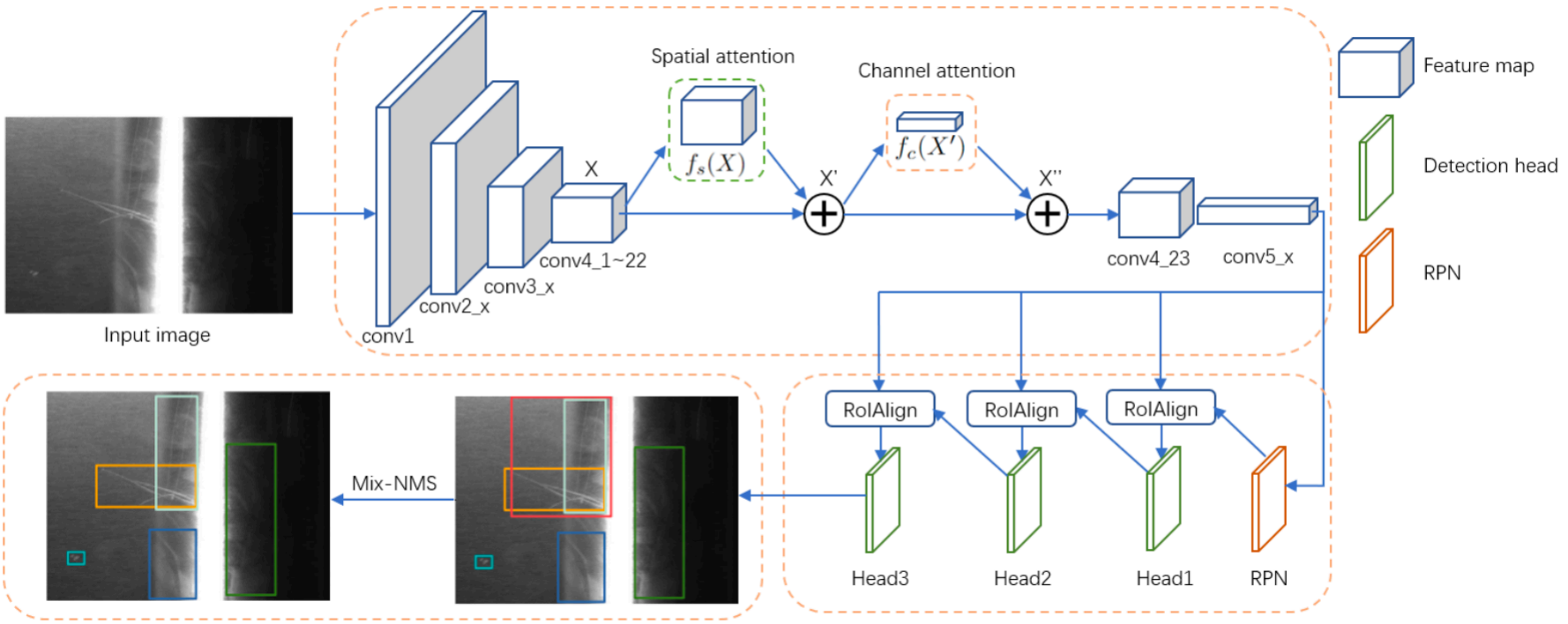
ACRM: Attention Cascade R-CNN with Mix-NMS for Metallic Surface Defect Detection

Junting Fang
Xiaoyang Tan

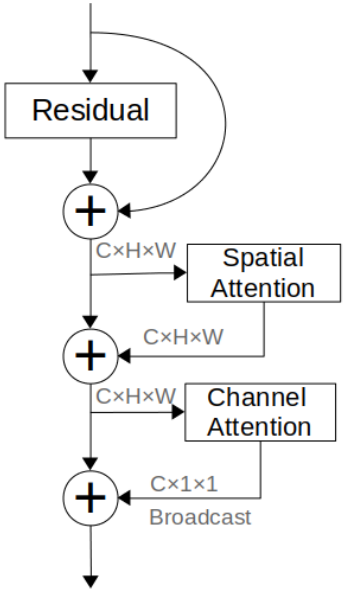
College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics

The architecture of ACRM



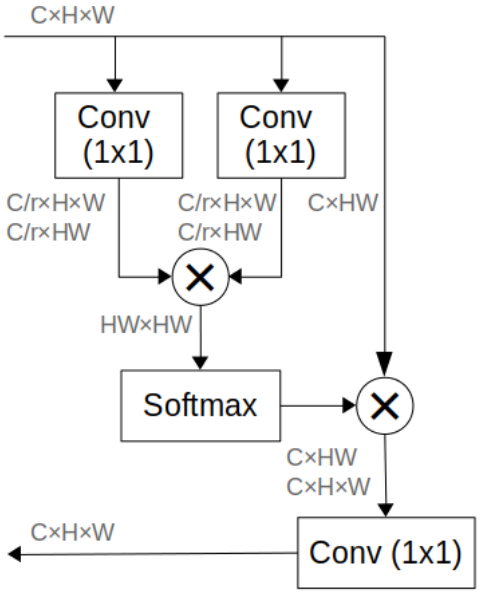
Three steps: a backbone network with an attention module;
multi-IOU cascade R-CNN detectors;
a Mix-NMS post-processing

Attention module



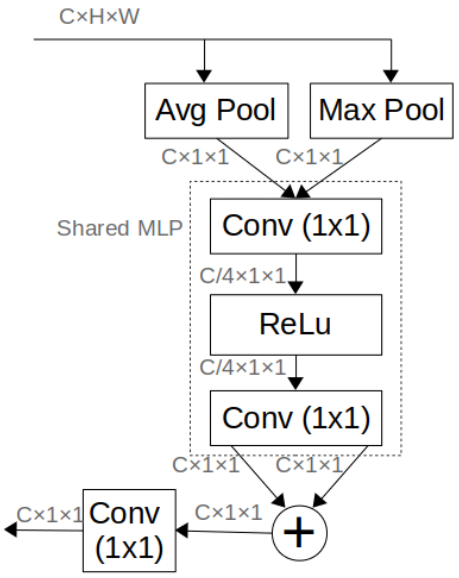
$$X' = f_s(X) \oplus X$$
$$X'' = f_c(X') \oplus X'$$

Spatial attention



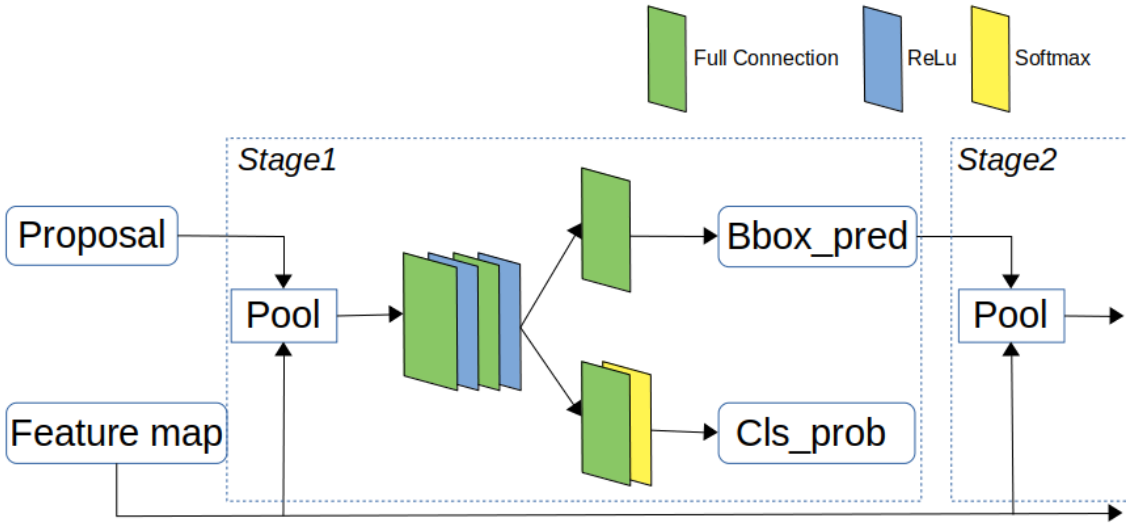
$$f_s(X) = \left\{ W_z \sum_{\forall j}^{N} \frac{f(x_i, x_j)}{C(X)} x_j \right\}_{i=1}^{N}$$
$$= \left\{ W_z \sum_{\forall j}^{N} \frac{e^{\theta(x_i)^T \phi(x_j)}}{C(X)} x_j \right\}_{i=1}^{N}$$
$$= W_z(softmax(X^T W_\theta^T W_\phi X)X)$$

Channel attention



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

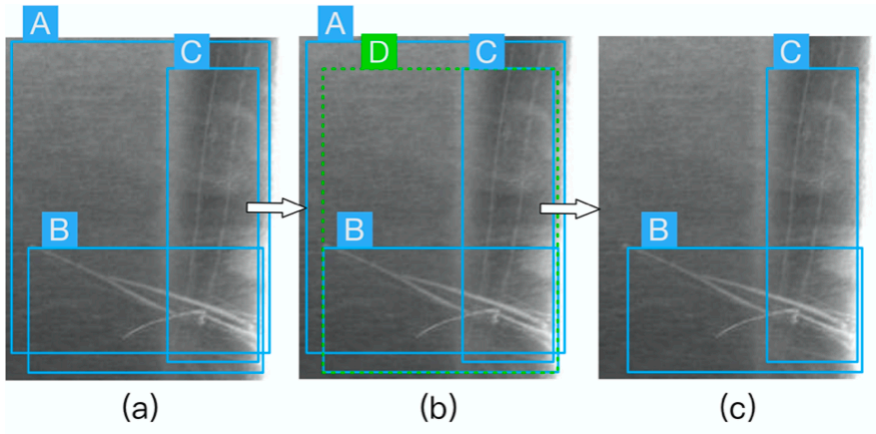
Cascade R-CNN



Loss Function :

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

Mix-NMS



Algorithm 1 Mix-NMS

Input: Initial detection boxes set $\boldsymbol{B} = \{b_1, \ldots, b_n\}$, corresponding detection scores set $\boldsymbol{S} = \{s_1, \ldots, s_n\}$, 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

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**

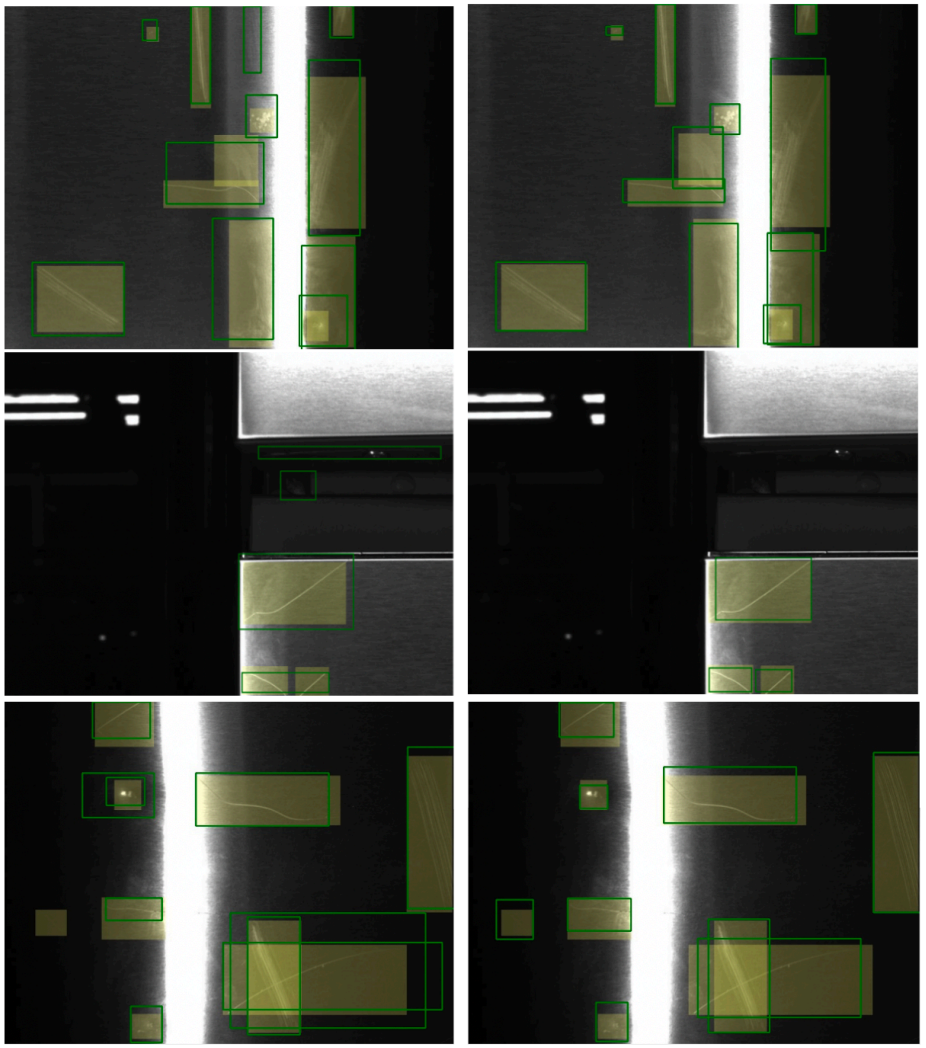
TABLE II
COMPARISON WITH TRADITIONAL METHOD

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



Faster R-CNN ACRM

THANKS