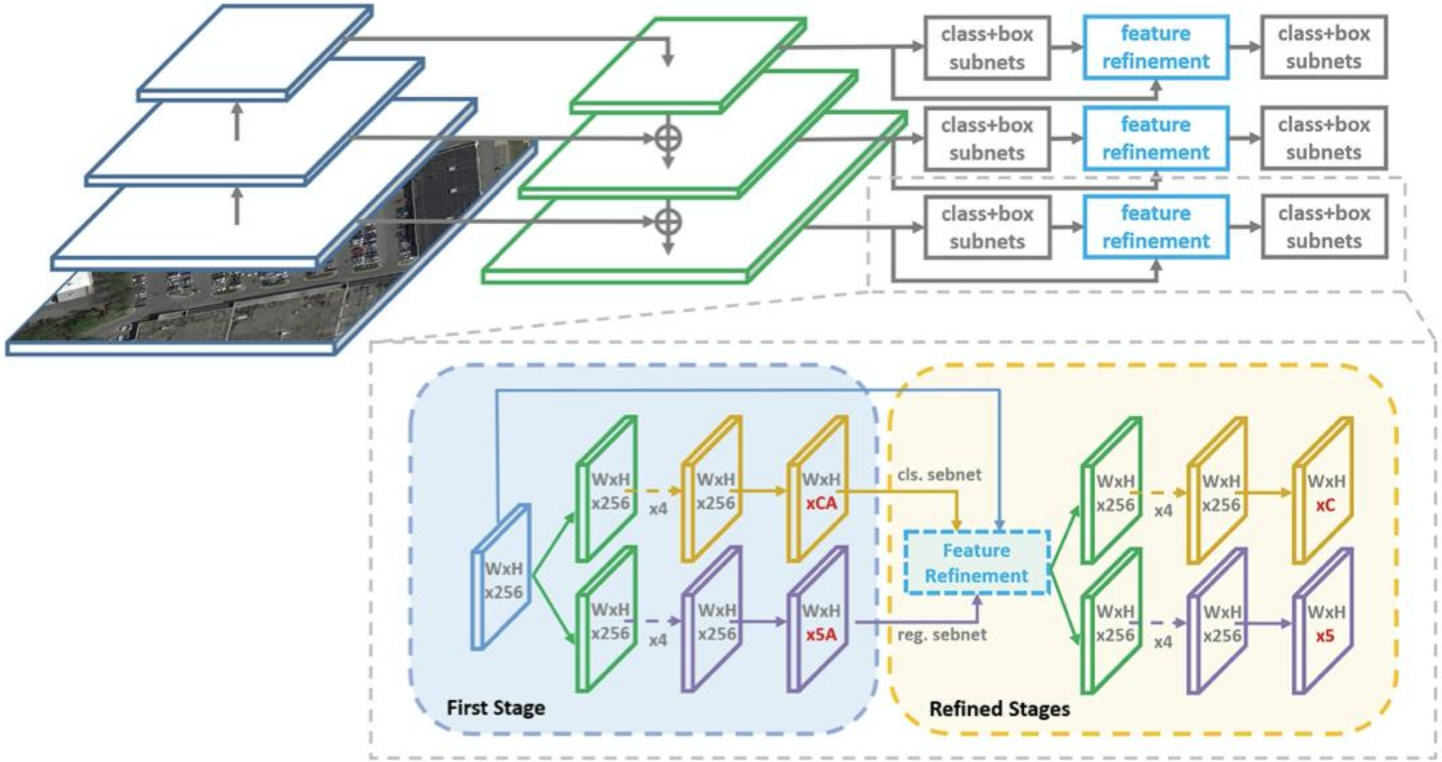
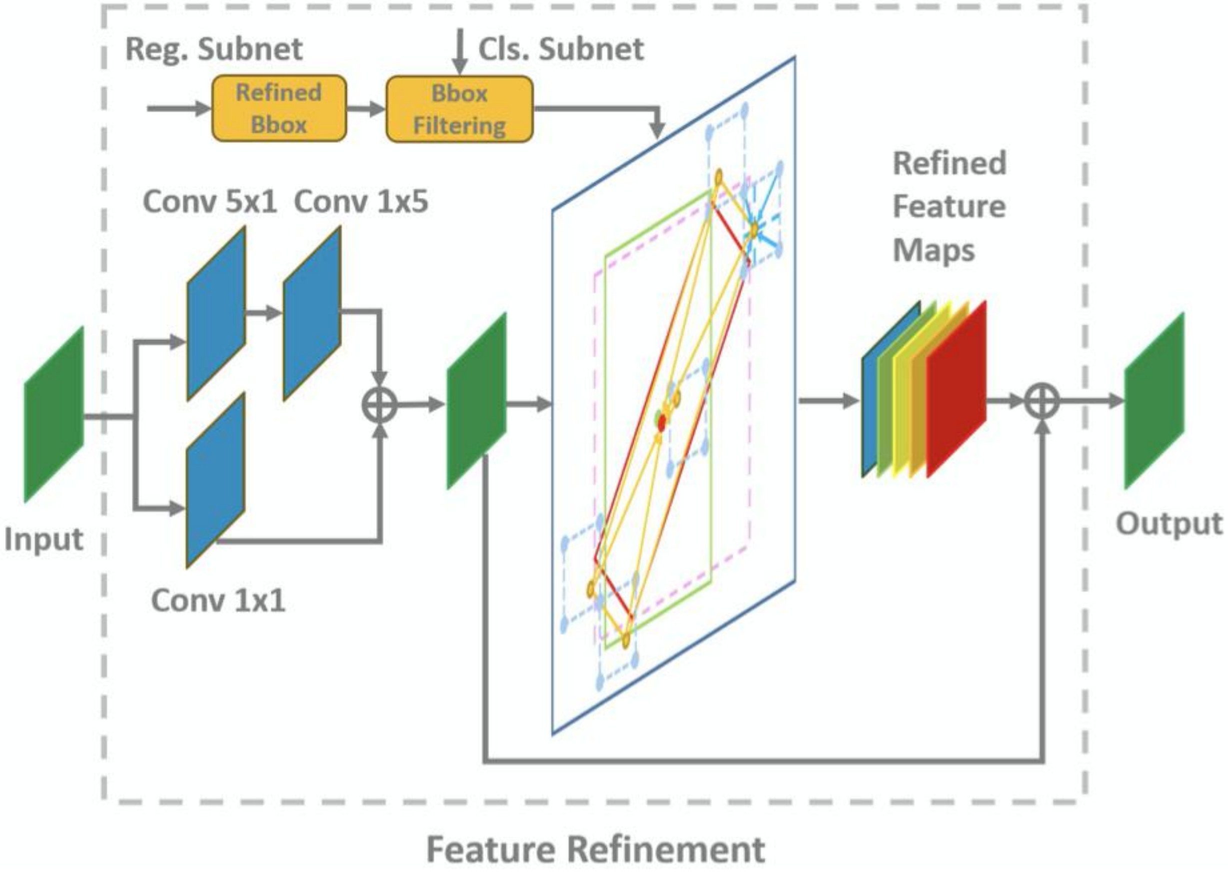
论文总结和改进方向

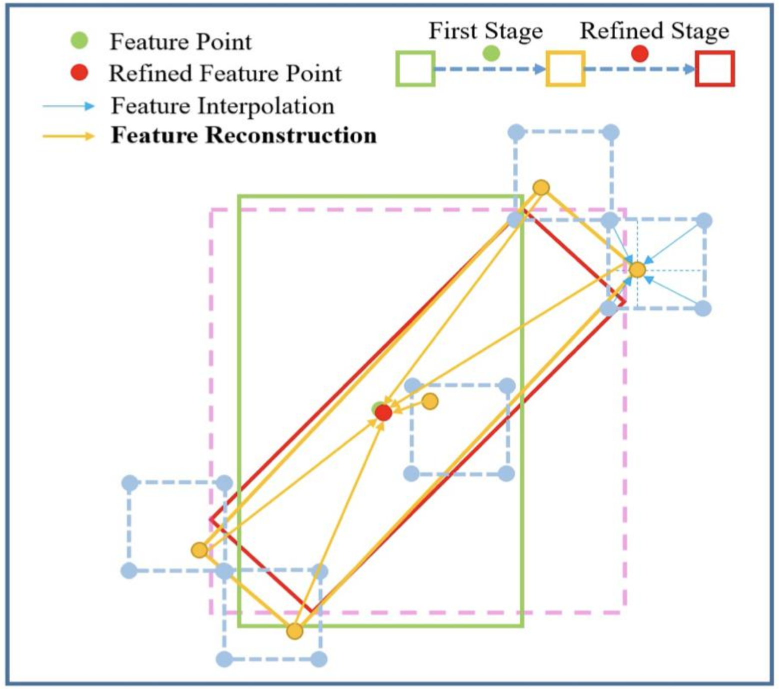
- 特征不对齐问题
- 旋转不变性

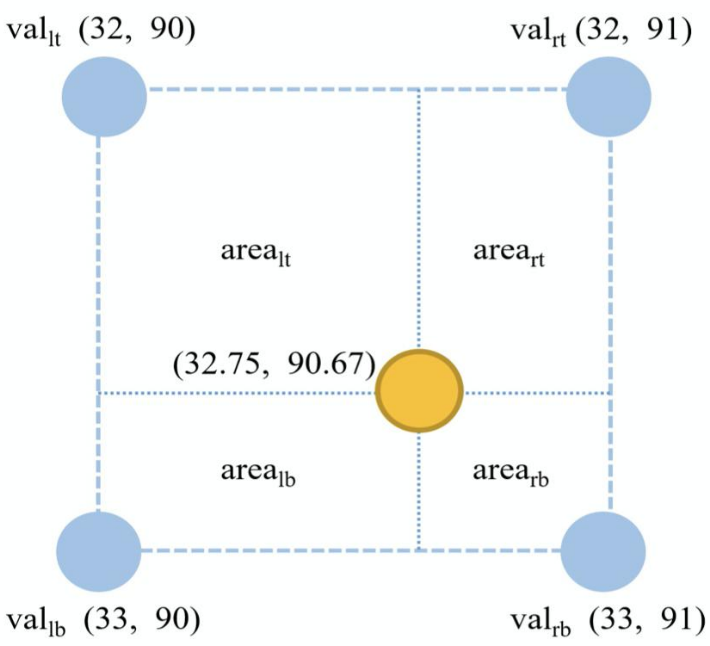
R 3 Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object



Considering the shortcoming of feature misalignment in existing refined single-stage detector, we design a feature refinement module to improve detection performance by getting more accurate features.




Feature Point
Refined Feature Point
Feature Interpolation
Feature Reconstruction
First Stage
Refined Stage


val_lt (32, 90)
val_rt (32, 91)
area_lt
area_rt
(32.75, 90.67)
area_lb
area_rb
val_lb (33, 90)
val_rb (33, 91)

Algorithm 1 Feature Refinement Module

Input: original feature map F, the bounding box (B) and confidence (S) of the previous stage

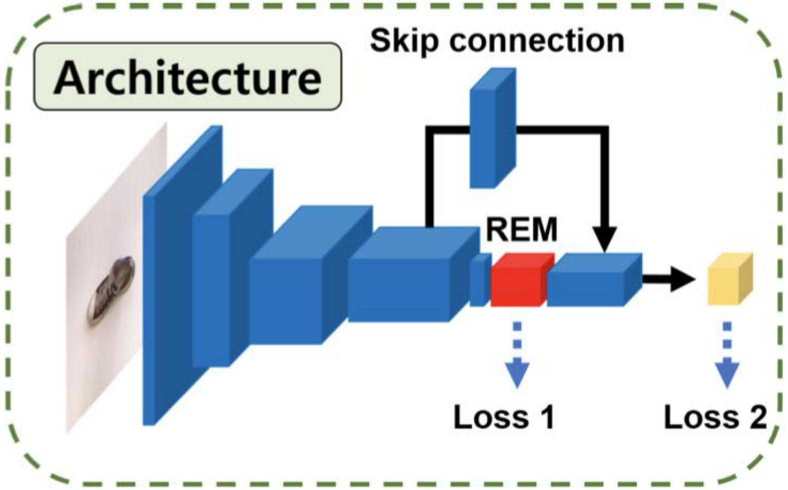
Output: reconstructed feature map F'

$B' \leftarrow Filter(B, S)$;
$h, w \leftarrow Shape(F)$, $F' \leftarrow ZerosLike(F)$;
$F \leftarrow Conv_{1\times1}(F) + Conv_{1\times5}(Conv_{5\times1}(F))$
for $i \leftarrow 0$ **to** $h - 1$ **do**
    for $j \leftarrow 0$ **to** $w - 1$ **do**
        $P \leftarrow GetFivePoints(B'(i, j))$;
        for $p \in P$ **do**
            $p_x \leftarrow Min(p_x, w - 1)$, $p_x \leftarrow Max(p_x, 0)$;
            $p_y \leftarrow Min(p_y, h - 1)$, $p_y \leftarrow Max(p_y, 0)$;
            $F'(i, j) \leftarrow F'(i, j) + BilinearInte(F, p)$;
        end for
    end for
end for
$F' \leftarrow F' + F$;
return F'

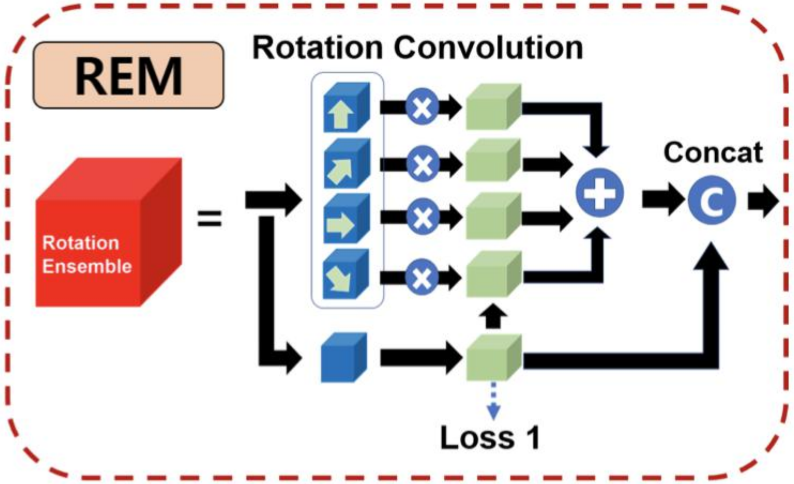
将回归框的5个点的双线性插值特征累加到瞄点（i, j)

Real-Time, Highly Accurate Robotic Grasp Detection using Fully Convolutional Neural Network with Rotatior

旋转不变性



(a)



(b)

Consider a typical scenario of convolution with input feature maps $f \in \mathbb{R}^{H\times W\times C}$ where $N = H \times W$ is the number of pixels and C is the number of channels. Let us denote $g_l \in \mathbb{R}^{K\times K\times C}$, $l = 1, \ldots, n_f$ a convolution kernel where $K \times K$ is the spatial dimension of the kernel and there are n_f number of kernels in each channel. Similar to the group convolutions [4], we propose n_r rotations of the weights to obtain $n_f \cdot n_r$ rotated weights for each channel. Bilinear interpolations of four adjacent pixel values were used for generating rotated kernels. A rotation matrix is

$$R(r) = \begin{bmatrix} \cos(r\pi/4) & -\sin(r\pi/4) & 0 \\ \sin(r\pi/4) & \cos(r\pi/4) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

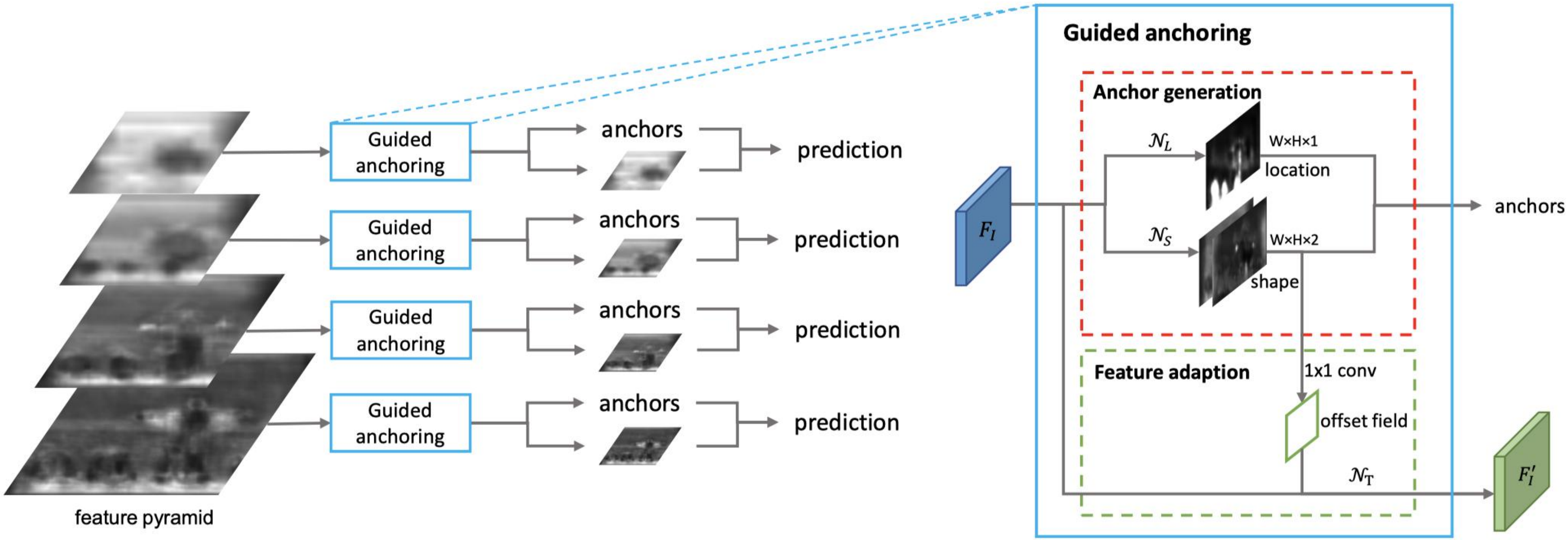
where r is an index for rotations. Then, the rotated weights (or kernels) are $g_l^i = R(i)g_l, i = 0, \ldots, 3, l = 1, \ldots, n_f$. Finally, the output of these convolutional layers with rotation operators for the input f is

$$d_l^i = g_l^i \star f, i = 0, \ldots, 3, l = 1, \ldots, n_f,$$

where $\star$ is a convolution operator. This pipeline of operations is called "rotation convolution". A typical kernel size is K=5.

Our REM contains rotation activation that aggregates all feature maps at different angles. Assume that an intermediate output for $\{t^x, t^y, \theta, t^w, t^h, t^z\}$ is available in REM, called $\{t_m^x, t_m^y, \theta_m, t_m^w, t_m^h, t_m^z\}$. Note that $\theta_m^i \in \mathbb{R}^{H\times W}$ where $i = 0, \pi/4, 2\pi/4, 3\pi/4$. For each angle, activations will be generated and all of them must be aggregated to yield one final feature map $\hat{d}_l = \sum_{i=1}^{4} d_l^i \odot \theta_m^i/4$. where $\odot$ is Hadamard product. Thus, our proposed method utilizes class probability (probability to grasp) to selectively aggregate activations along with the weight of angle classification.

Region Proposal by Guided Anchoring



改进方法