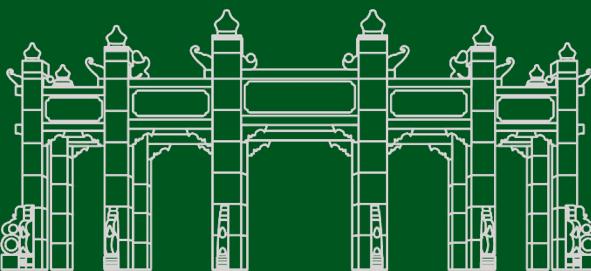


RWSC-Fusion: Region-Wise Style-Controlled Fusion Network for the Prohibited X-ray Security Image Synthesis



The X-ray penetrates through objects to form X-ray image. The image intensity is related to the X-ray energy, object material, object thickness [21], simply expressed as:

$$I = I_0 e^{-\mu h} \quad (1)$$

where I_0 is the X-ray beam intensity, μ is the absorption coefficient and h is the thickness of the objects.

Accordingly, the prohibited item X-ray image I_f , and background image I_b , e.g. a baggage, can be expressed as:

$$I_f = I_0 e^{-\mu_f h_f}, \quad I_b = I_0 e^{-\mu_b h_b} \quad (2)$$

where μ_f and μ_b are the respective absorption coefficients, h_f and h_b are the respective thickness.

When superimposing the prohibited items I_f onto the baggage image I_b , the fused X-ray image I_{fb} could be:

$$I_{fb} = I_0 e^{-\mu_f h_f - \mu_b h_b} = I_0 e^{-\mu_f h_f} e^{-\mu_b h_b} \quad (3)$$

Thus, we can get:

$$I_{fb} = \frac{I_f \cdot I_b}{I_0} \quad (4)$$

2.2. Color X-ray Security Images Rendering

如上所述，真实的X射线安全图像是根据某些规则用伪颜色渲染的[5]。最常见的规则基于HSV颜色空间中的三个组成部分：色调 (H)、饱和度 (S) 和值 (V)，这近似于人类感知和解释颜色的方式。首先，根据物体的材料类别来分配伪色的H分量，这取决于双能量响应。例如，有机材料通常显示为红色，无机材料显示为蓝色，混合物显示为绿色。接下来，三者的V分量是归一化的非线性映射，而S分量大多与单能量或双能量响应线性相关。因此，X射线安全图像中物体的颜色（见补充材料中的示例）根据材料、厚度、视点和背景而不同。

As mentioned above, real X-ray security images are rendered with pseudo-colors according to certain rules [5]. The most common rule is based on the three components: Hue (H), Saturation (S) and Value (V) in the HSV color space, which approximates the way human perceives and interprets color. First, the H component of pseudo-color is assigned according to the material category of the object, which depends on the dual energy responses. e.g., organic materials are usually shown in red, inorganic materials are in blue, and mixtures are in green. Next, the V component of the three is a normalized nonlinear mapping, while the S component is mostly correlated linearly with the single energy or dual energy responses. As a consequence, the colors of objects in X-ray security image (see examples in the supplementary material), vary differently according to the materials, thickness, viewpoint and background.

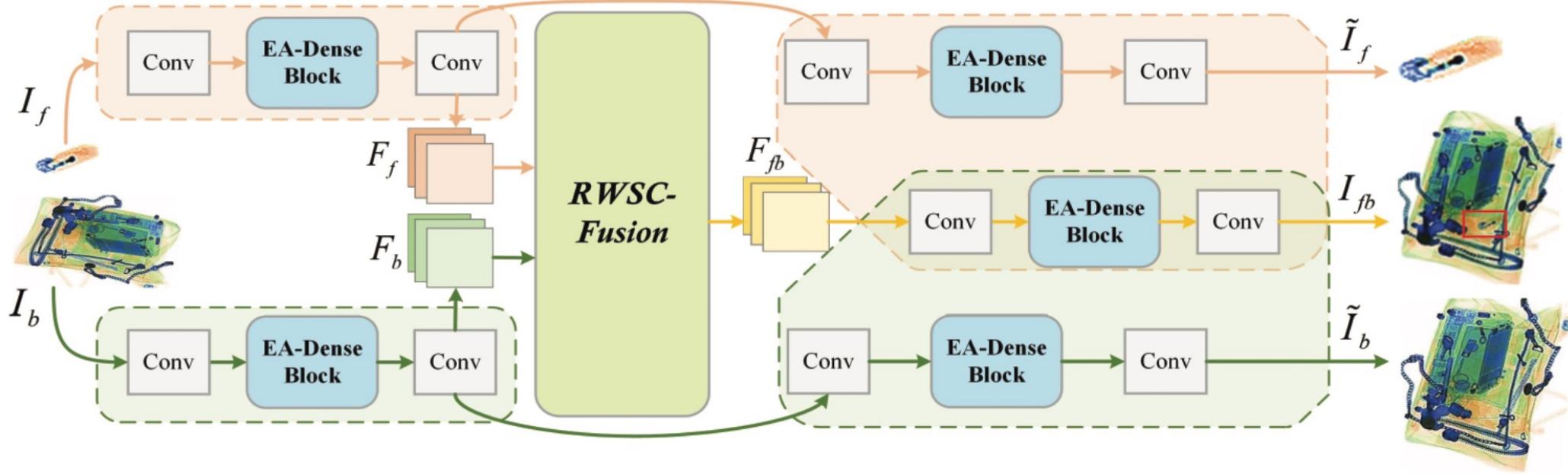
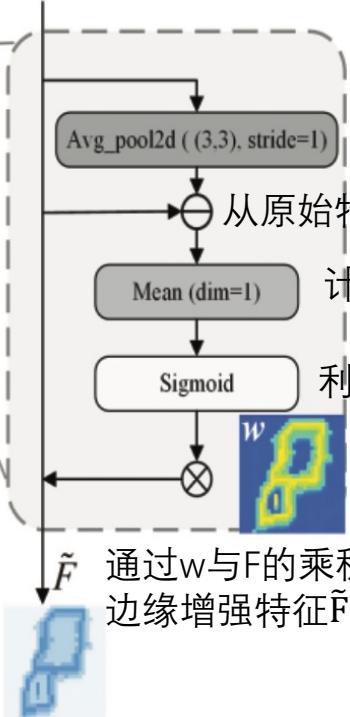
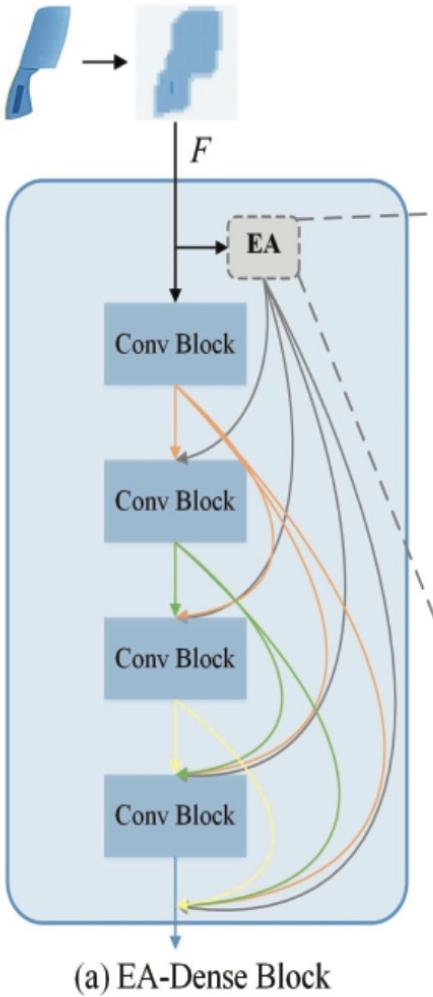


Figure 2. Overview of the proposed RWSC-Fusion.



3×3 平均池化层，步长为1

从原始特征图中减去局部平均特征图

计算沿通道维度的平均响应，以将通道数减少到1

利用sigmoid函数获得边缘感知激活特征图w

\tilde{F} 通过w与F的乘积，可以获得
边缘增强特征 \tilde{F}

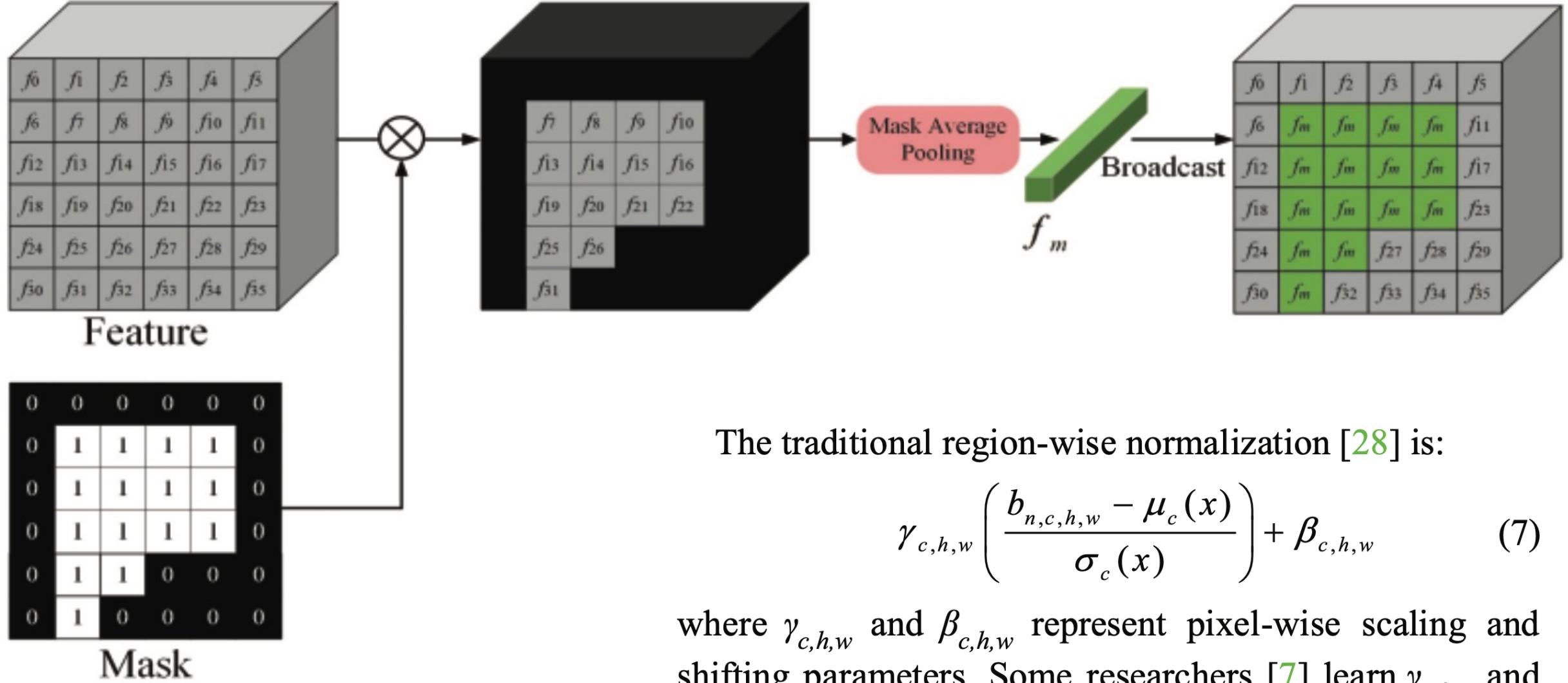
whole process of EA module can be expressed as:

$$w = \sigma(M(|AP2D(F) - F|)) \quad (5)$$

$$\tilde{F} = w \odot F \quad (6)$$

where $AP2D$ denotes 2D average pooling, M denotes the mean value along the channel dimension, σ denotes the sigmoid activation, \odot denotes pixel-wise multiplication.

Figure 3. EA-Dense Block with Edge-Attention module.



The traditional region-wise normalization [28] is:

$$\gamma_{c,h,w} \left(\frac{b_{n,c,h,w} - \mu_c(x)}{\sigma_c(x)} \right) + \beta_{c,h,w} \quad (7)$$

where $\gamma_{c,h,w}$ and $\beta_{c,h,w}$ represent pixel-wise scaling and shifting parameters. Some researchers [7] learn $\gamma_{c,h,w}$ and $\beta_{c,h,w}$ from the style maps by using mask average pooling, which is shown in Figure 4. However, such parameters have less potential for realizing rich and fine style-control.

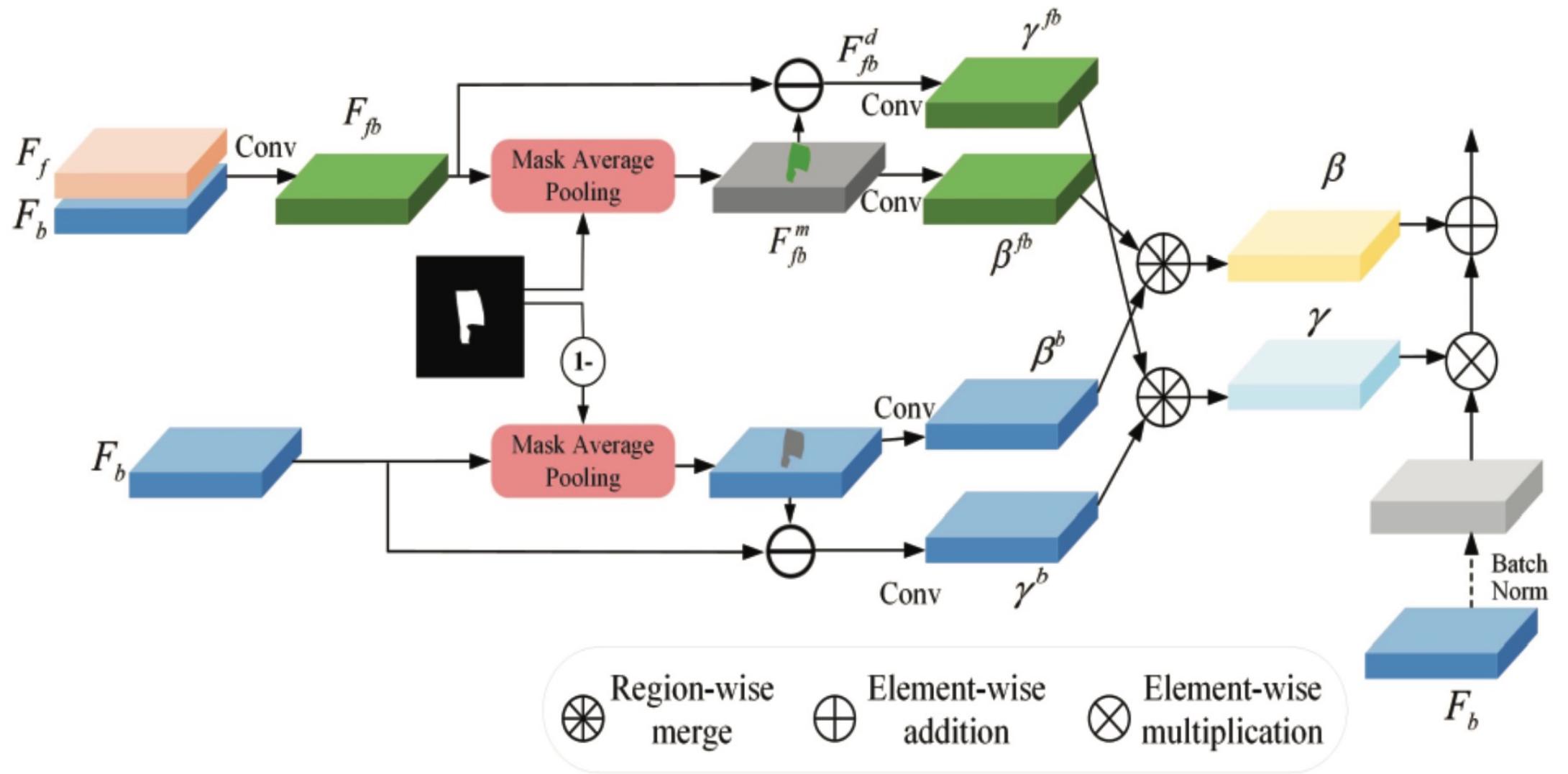
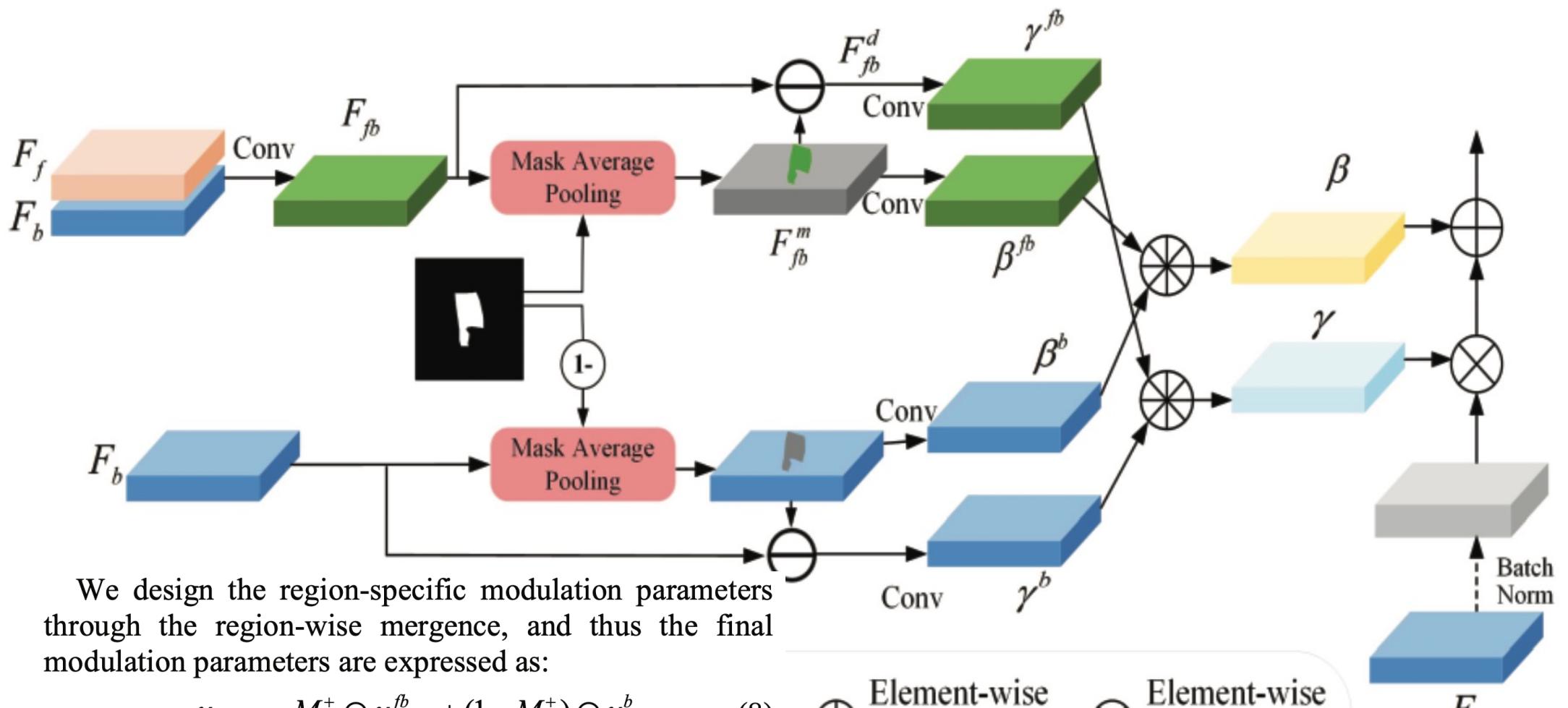


Figure 5. Details of RWSC-Fusion module.



$$\gamma_{c,h,w} = M^+ \odot \gamma_{c,h,w}^{fb} + (1 - M^+) \odot \gamma_{c,h,w}^b \quad (8)$$

$$\beta_{c,h,w} = M^+ \odot \beta_{c,h,w}^{fb} + (1 - M^+) \odot \beta_{c,h,w}^b \quad (9)$$

where M^+ denotes the mask of prohibited item, while $1 - M^+$ denotes the complementary mask. The appearance of the overlapping region within the mask is modulated jointly by the prohibited item and baggage image, while the non-overlapping region is modulated by the single baggage image, avoiding the interference of prohibited item image.

掩码内重叠区域的外观由违禁物品和行李图像共同调制，而非重叠区域由单个行李图像调制，避免了违禁物品图像的干扰

Existing unsupervised image fusion models usually maximize the similarity of fused image with source images instead of ground-truth, by using loss function as follows:

$$\mathcal{L} = 1 - [w_1 \bullet \ell(x_1, y) + w_2 \bullet \ell(x_2, y)] \quad (10)$$

As mentioned in Sec.2.1, when fusing the prohibited items I_f with the baggage image I_b by using the weights w_1 and w_2 , the fused X-ray image I_{fb} could be:

$$I_{fb} = I_0 e^{-w_1 \mu_f h_f - w_2 \mu_b h_b} = I_0 e^{-w_1 \mu_f h_f} e^{-w_2 \mu_b h_b} \quad (11)$$

$$I_{fb} = \frac{I_f^{w_1} \cdot I_b^{w_2}}{I_0^{w_1 + w_2 - 1}} \quad (12)$$

In this case, we adjust Eq. (10) by using the logarithmic form, giving access to multiplication relationship in linear form. Finally, we develop a Luminance loss in Logarithmic form (*LL*) \mathcal{L}_{LL} as follows:

$$\mathcal{L}_{LL} = 1 - [w_1 \bullet \log \ell(I_f, I_{fb}) + w_2 \bullet \log \ell(I_b, I_{fb})] \quad (13)$$

We put w_1 and w_2 inside the metric formula to develop a new similarity metric $\ell(x, y, w)$, and thus \mathcal{L}_{LL} turns into:

$$\mathcal{L}_{LL} = 1 - [\log \ell(I_f, I_{fb}, w_1) + \log \ell(I_b, I_{fb}, w_2)] \quad (14)$$

where $\ell(x, y, w) = \frac{2\mu_x\mu_y + \varepsilon}{\mu_x^{4w} + \mu_y^2 + \varepsilon}$, and thus the local mean of fused image μ_{fb} is expected to converge to $\mu_f^{w_1} \cdot \mu_b^{w_2}$. The details and deduce are in the supplementary materials.

On the other hand, considering that the saturation (S) component of color X-ray security image is mostly correlated linearly with the energy responses, we assume that the S component of the fused image is correlated linearly with the multiplication of the S component of the two images. Therefore, the difference between the S component of the prohibited item (S_f) and the fused image (S_{fb}) comes almost from the S component of the baggage image (S_b). Similarly, the difference between the S_b and S_{fb} comes almost from S_f . We develop an another loss term Correlation loss of Saturation Difference (CSD) \mathcal{L}_{CSD} :

$$\begin{aligned} D_b &= (1 - S_{fb}) / (1 - S_f) \\ D_f &= (1 - S_{fb}) / (1 - S_b) \end{aligned} \quad (15)$$

$$\mathcal{L}_{CSD} = 1 - [CC(D_f, 1 - S_f) + CC(D_b, 1 - S_b)] / 2 \quad (16)$$

where CC is the normalized correlation coefficient.

In addition, a third loss term is the reconstruction loss \mathcal{L}_{recon} of two source images \tilde{I}_f and \tilde{I}_b as follows:

$$\mathcal{L}_{recon} = \|I_f, \tilde{I}_f\| + \|I_b, \tilde{I}_b\| \quad (17)$$

In conclusion, the total loss function is:

$$\mathcal{L}_{total} = \mathcal{L}_{LL} + \mathcal{L}_{CSD} + \mathcal{L}_{recon} \quad (18)$$

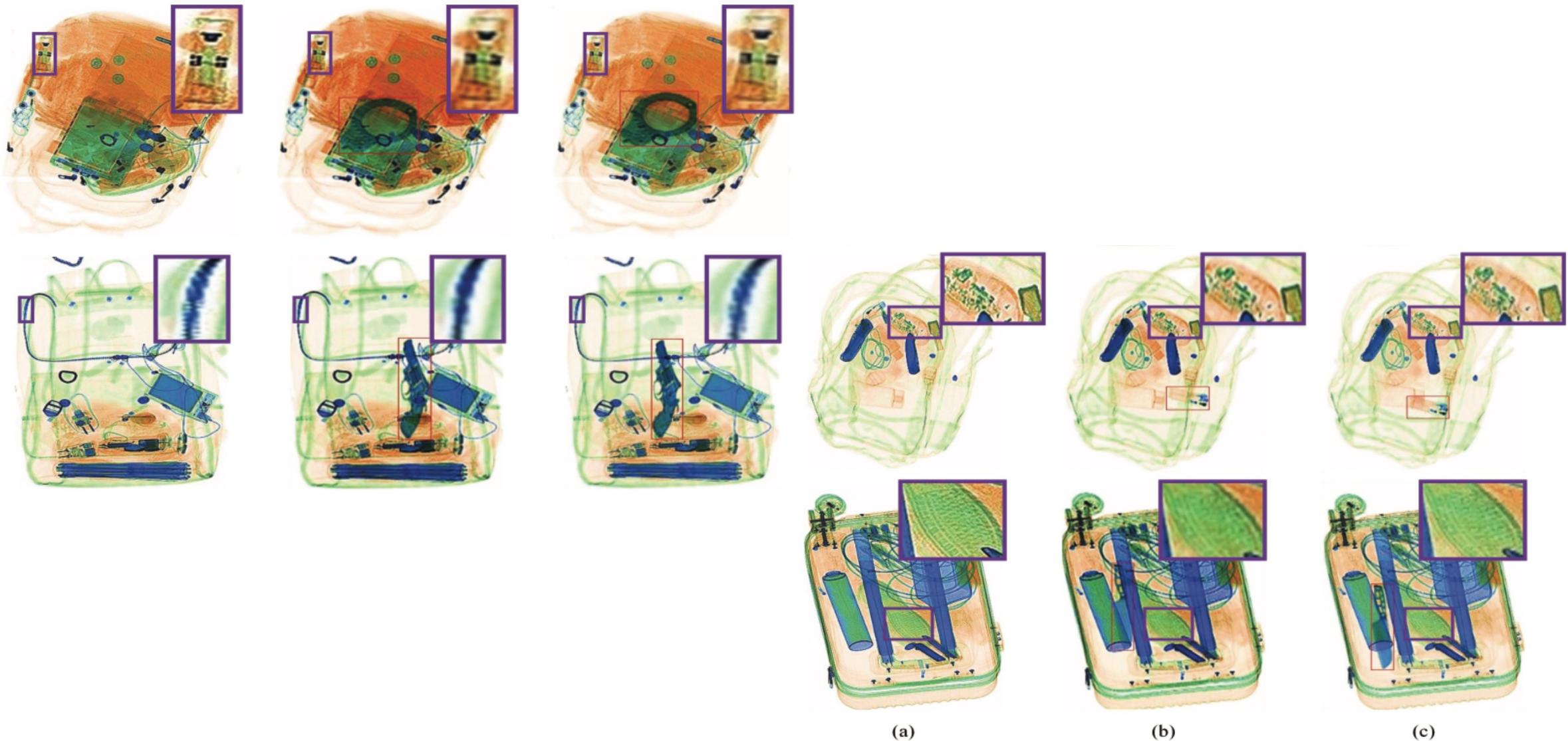


Figure 6. The generated images from the models with and without the EA modules. From left to right are: (a) original baggage image; (b) the synthesized images from the model without EA modules; (c) the synthesized images from the model with EA modules.

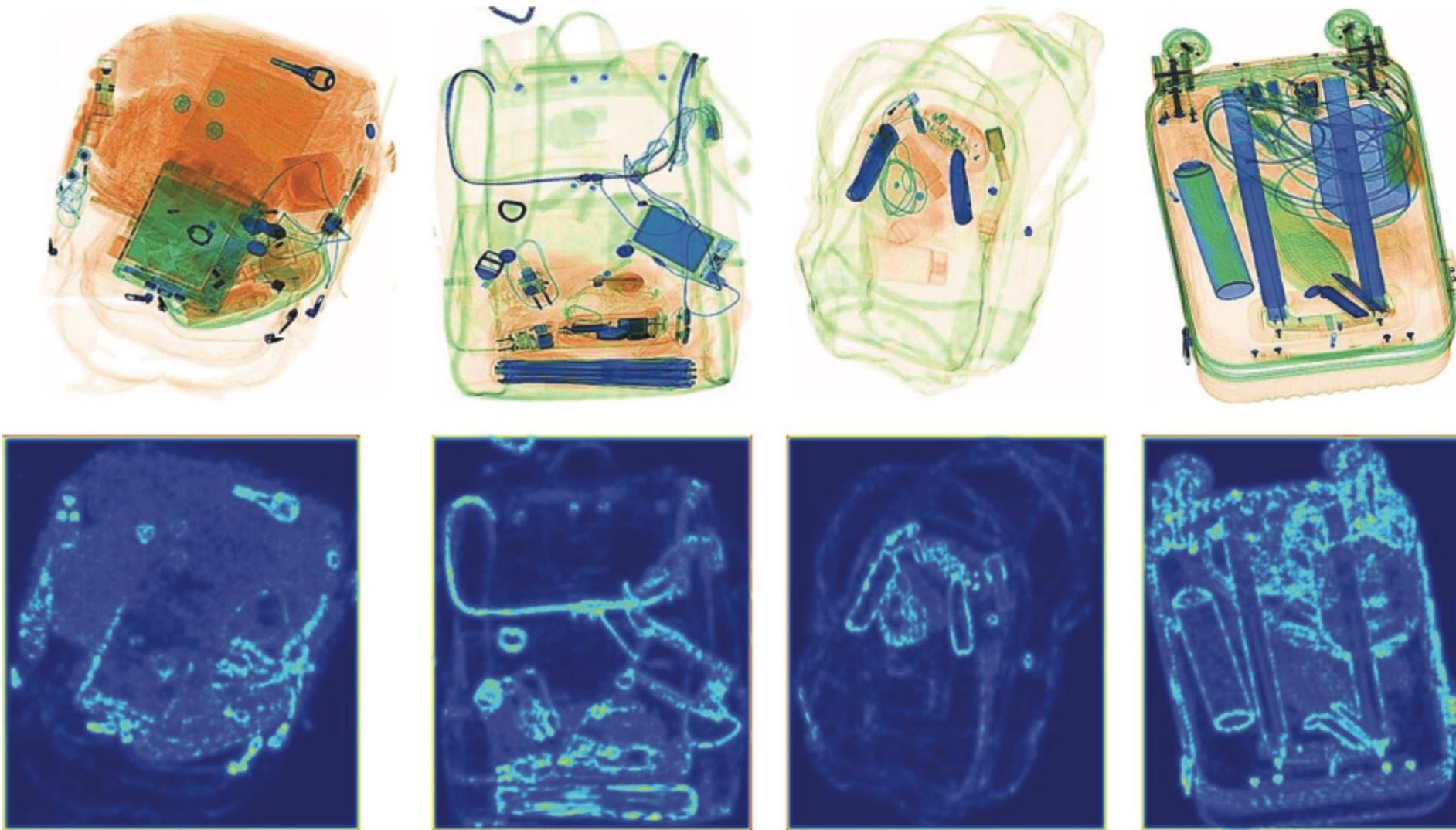


Figure 7. The edge attention maps of the X-ray security images in Figure 6. The first row shows the original images, and the second row shows the edge attention maps extracted by the EA module.

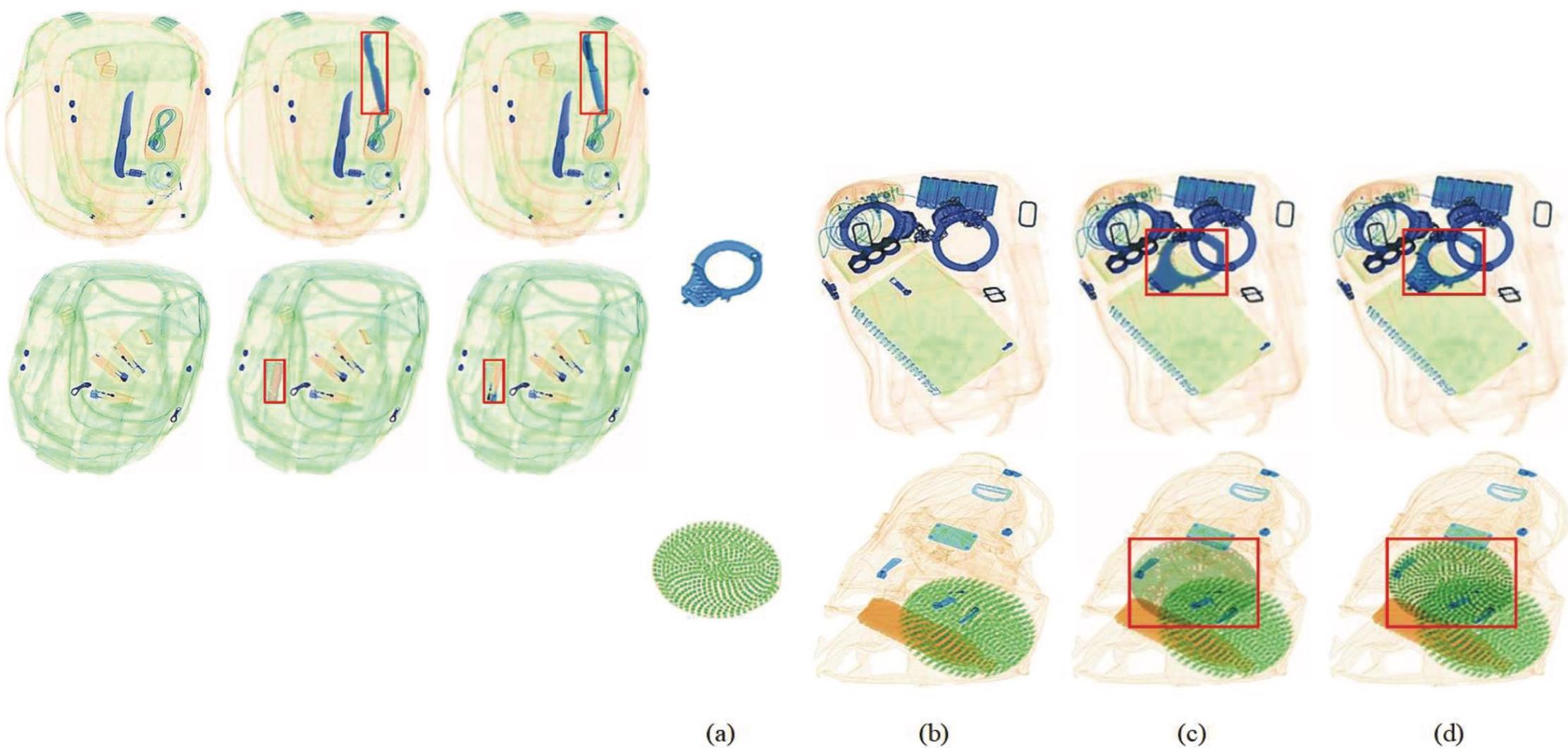


Figure 8. The generated images from the fusion module with only mask average pooling layer and our RWSC-Fusion module. From left to right are: (a) prohibited items; (b) original baggage image; (c) the synthesized images from the model with only mask average pooling layer; (d) the synthesized images from our model.

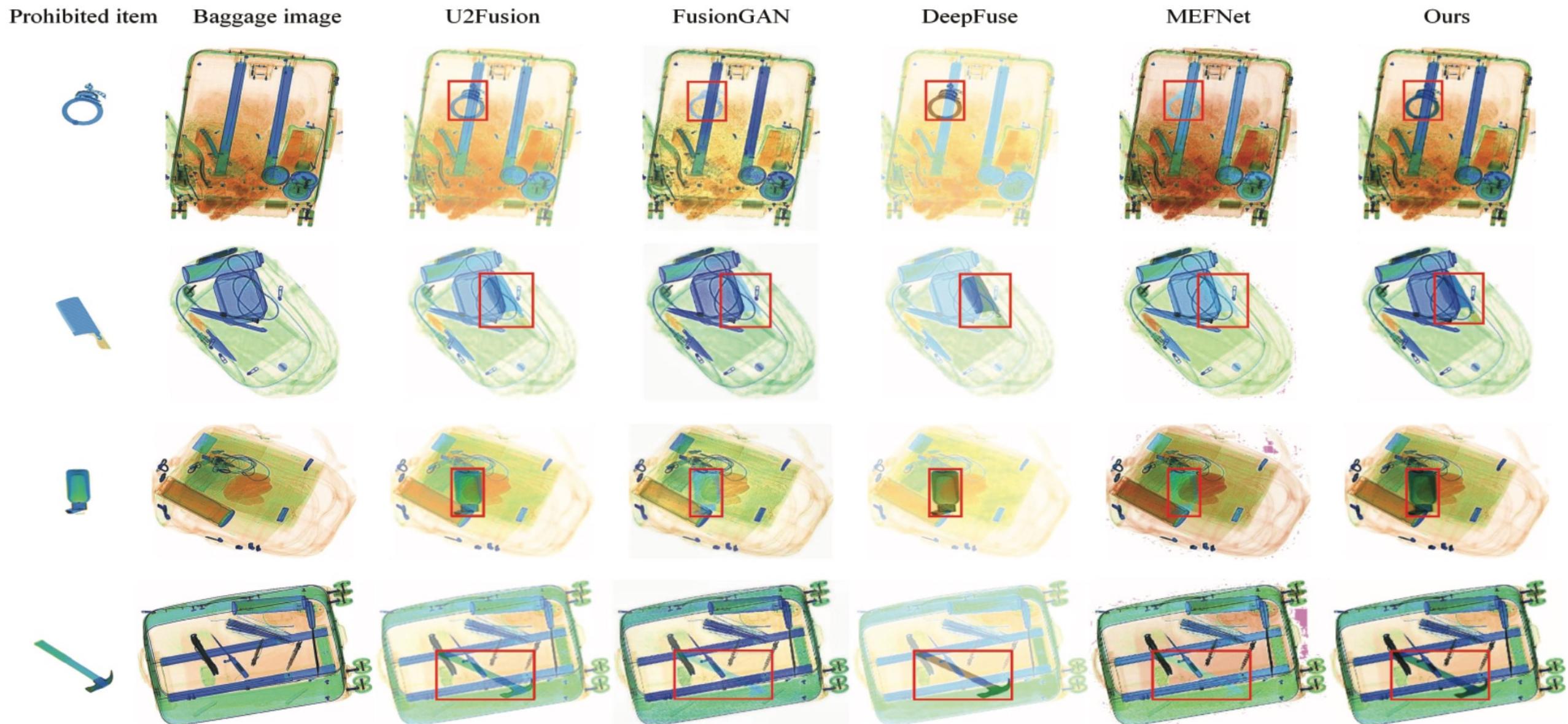


Table 3 Detection results of YOLOv4 trained with different data on private dataset (recall and precision at confidence level of 0.5).

| Training Data | | DataR | DataR1/2 | DataRS | DataRS+ |
|----------------|-----------|-------|----------|--------|---------|
| Plastic bottle | Recall | 69.49 | 64.82 | 70.46 | 74.15 |
| | Precision | 85.36 | 83.29 | 88.14 | 85.13 |
| | AP | 69.50 | 71.16 | 71.00 | 76.97 |
| Cutter | Recall | 86.17 | 74.04 | 81.70 | 84.50 |
| | Precision | 92.62 | 87.60 | 91.81 | 93.45 |
| | AP | 86.96 | 79.98 | 83.49 | 87.70 |
| Scissor | Recall | 71.22 | 58.03 | 70.87 | 72.97 |
| | Precision | 90.60 | 89.34 | 88.37 | 88.23 |
| | AP | 74.28 | 66.81 | 72.83 | 78.29 |
| Therm os cup | Recall | 88.86 | 88.86 | 91.59 | 92.76 |
| | Precision | 93.52 | 90.13 | 91.38 | 92.54 |
| | AP | 85.52 | 86.17 | 88.98 | 90.25 |
| Glass bottle | Recall | 56.49 | 48.74 | 58.71 | 58.82 |
| | Precision | 86.79 | 80.97 | 84.02 | 88.15 |
| | AP | 56.31 | 53.76 | 58.13 | 61.87 |
| mAP | | 74.52 | 71.58 | 74.88 | 79.02 |

(1) DataR1/2：我们从DataR中删除了20000个样本，产生了20000个训练样本，包括13127个塑料瓶、17055个切割器、5117把剪刀、2518个保温杯和8710个玻璃瓶。

(2) DataRS：我们将违禁物品合成并补充到DataR1/2中，得到20000个样本，其中26048个塑料瓶、34856个刀具、10391把剪刀、4968个保温杯和17526个玻璃瓶，违禁物品数量与DataR相同。（注：由于没有额外的20000个阴性样本，不包括所有五种违禁品，我们无法实现20000个样本的纯合成，与DataR完全一致）

(3) DataRS+：我们将违禁品合成为其他20000个样本中，并将新的合成样本与DataR中的20000个样本合并，产生40000个样本，共38983个塑料瓶，51956个刀具，15597把剪刀、7522个保温杯和26034个玻璃瓶

Table 4 Detection results of YOLOv4 trained with different data on SIXray dataset.

| Training Data | | SIXR | SIXR1/2 | SIXRS | SIXRS+ |
|---------------|-----------|-------|---------|-------|--------|
| Gun | Recall | 79.73 | 73.02 | 83.23 | 83.69 |
| | Precision | 96.32 | 91.76 | 97.33 | 97.17 |
| | AP | 78.71 | 75.21 | 82.19 | 82.74 |
| Knife | Recall | 58.88 | 45.62 | 64.17 | 66.04 |
| | Precision | 92.65 | 86.39 | 92.79 | 93.39 |
| | AP | 62.84 | 47.88 | 66.64 | 69.17 |
| Wrench | Recall | 55.97 | 24.25 | 59.33 | 66.04 |
| | Precision | 87.72 | 60.75 | 83.68 | 83.49 |
| | AP | 65.93 | 26.30 | 66.82 | 68.88 |
| Plier | Recall | 66.35 | 50.07 | 71.45 | 76.14 |
| | Precision | 85.49 | 72.34 | 86.81 | 85.16 |
| | AP | 74.20 | 57.63 | 75.84 | 79.53 |
| Scissor | Recall | 63.62 | 25.70 | 52.80 | 67.76 |
| | Precision | 91.16 | 66.27 | 91.87 | 85.80 |
| | AP | 72.11 | 39.08 | 62.94 | 73.60 |
| mAP | | 70.76 | 49.22 | 70.89 | 74.78 |

(1) SIXR1/2：我们从SIXR中删除一半（3748）个阳性样本，并将其他3748个阴性样本补充到SIXR。

(2) SIXR：我们将禁用项目合成到SIXR1/2中新添加的3748个阴性样本中，以补充阳性样本，从而混合了7496个与SIXR相同数量的禁用项目的真/假阳性样本。

(3) SIXRS+：我们将违禁物品合成到其他新的3748个阴性样本中，并在SIXR中替换3748个阳性样本，结果混合了11244个真/假阳性样本，包括6469支枪、4154把刀、4195把扳手、6933把钳子和1398把剪刀

Table 5 Detection results of different training data on OPIXray.

| Training data | OPIXray dataset | | |
|---------------|-----------------|--------------|--------------|
| | Recall | Precision | AP |
| OPIR | 71.08 | 75.85 | 67.47 |
| OPIRS | 71.14 | 76.60 | 72.20 |
| OPIRS+ | 80.46 | 90.13 | 84.80 |

OPIXray数据集，它包含7109个用于刀具检测的训练图像和1776个测试图像。

- (1) OPIR：所有7109张原始图像；
- (2) OPIRS：3555张图像添加合成刀具，刀具数量与OPIR相同；
- (3) OPIRS+：7109张图像，补充了一半合成刀具。