JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

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

Luwen Duan, Min Wu, Lijian Mao, Jun Yin, Jianping Xiong, Xi Li

ZHEJIANGDAHUATECHNOLOGYCO.,LTD. China
Zhejiang University

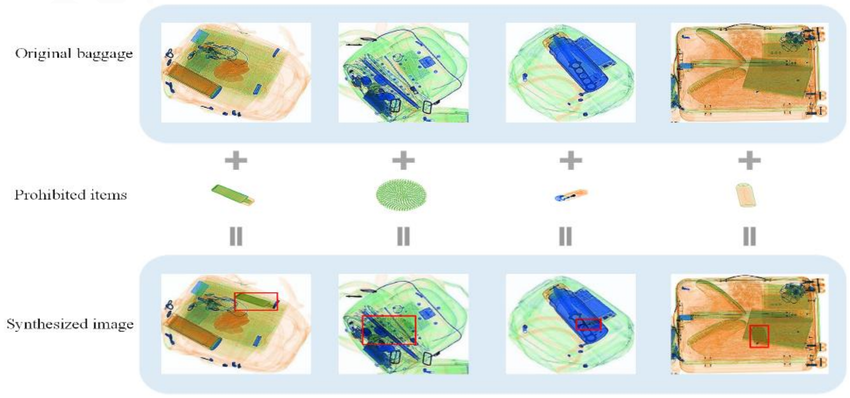
Poster Location: West Building Exhibit Halls ABC 171

- 1. Abstract
- 2. Introduction and Related work
- 3. Method
- 4. Experiment and Results
- 5. Conclusion

Abstract

■ Background:

The abundance and diversity of the X-ray security images with prohibited item, are essential for training the automatic prohibited item detection model. Existing datasets have not been up to the standard of model training.



■ Goal :

To solve the data insufficiency, we propose a RWSC-Fusion network, which superimposes the prohibited items onto normal X-ray security images, to synthesize the prohibited X-ray security images.

■ Method :

- **RWSC-Fusion** module: Region-Wise Style-Controlled fusion of overlapping region with novel modulation parameters. **Edge-Attention** module: improve the sharpness of the synthetic images effectively.
- Luminance loss in Logarithmic form (LL) and Correlation loss of Saturation Difference (CSD): optimize the fused X-ray security images in terms of luminance and saturation.

■ Conclusion:

We evaluate the authenticity and the training effect of the synthetic images on SIXray and OPIXray dataset, confirming that our synthetic images are reliable enough to augment the prohibited X-ray security images.

Introduction and Related work

Background

∵ Existing public datasets have not been up to the standard of real-world scenario for prohibited item detection model training.

Dataset	Color	Categories	Number of training samples
GDXray	Gray	guns, shurikens, razor blades	8850
OPIXray	RGB	cutter	8885
Sixray	RGB	guns, knives, wrenches, pliers, scissors, hammers	8929

∵ The traditional image augmentation methods are still unable to improve the diversity and complexity for the inter-occlusion between prohibited items.

∵ Existing X-ray security images synthesis models are supervised, and lack automation and versatility.

∴ We propose an unsupervised color X-ray security image fusion model, to synthesize prohibited X-ray security images and obtain annotation automatically.

Introduction and Related work

X-ray image (Threat image projection: TIP)

The X-ray penetrates through objects to form X-ray image:

$$I = I_0 e^{-\mu h} \tag{1}$$

(I_0 : X-ray beam intensity, μ : absorption coefficient, h : thickness of object)

The prohibited item X-ray image I_f, and a baggage image I_b, expressed as:

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

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

$$I_{fb} = I_b e^{-\mu_f h_f} = I_0 e^{-\mu_b h_b} e^{-\mu_f h_f} = I_0 e^{-\mu_b h_b - \mu_f h_f} \tag{3}$$

Thus, we can get:

$$I_{fb} = \frac{I_f \cdot I_b}{I_0} \tag{4}$$

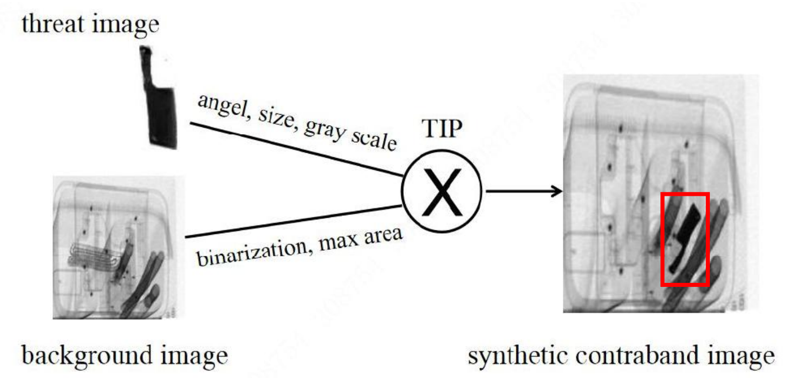


Figure 1. TIP method

TIP method is almost exclusively applicable to grayscale X-ray images.

Method

Model

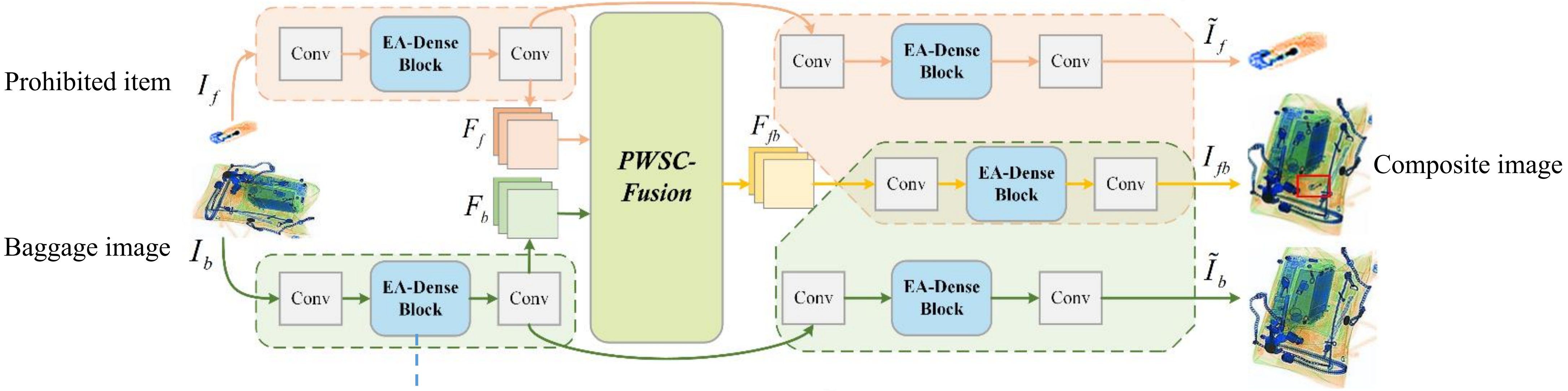
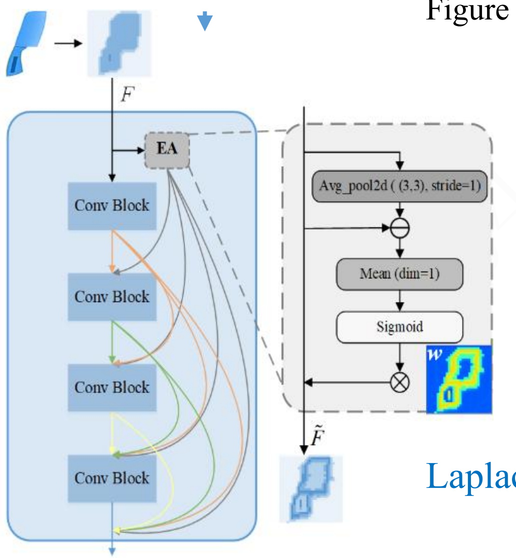


Figure 2. Overview of the proposed RWSC-Fusion.



(a) EA-Dense Block (b) EA

Laplacian sharpening

$$w = \sigma(M(|AP2D(F) - F|))$$
$$\tilde{F} = w \odot F \qquad (5)$$

$AP2D$: 2D average pooling,
M : the mean value along the channel dimension
σ : sigmoid activation
$\odot$: pixel-wise multiplication
$\tilde{F}$: edge-enhanced feature

- The weight w can activate the edge and texture information, the enhanced feature $\tilde{F}$ take advantage of both global and local texture.

Method

Region-Wise Style-Controlled Fusion Module

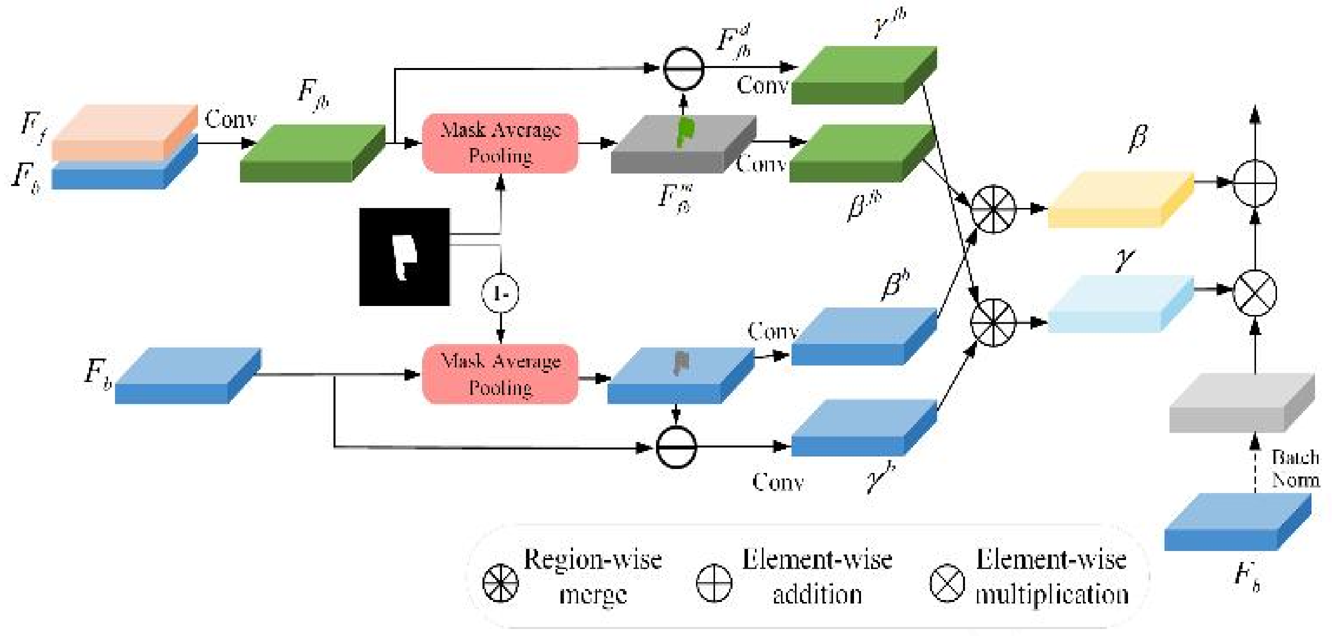


Figure 3. Details of RWSC-Fusion module.

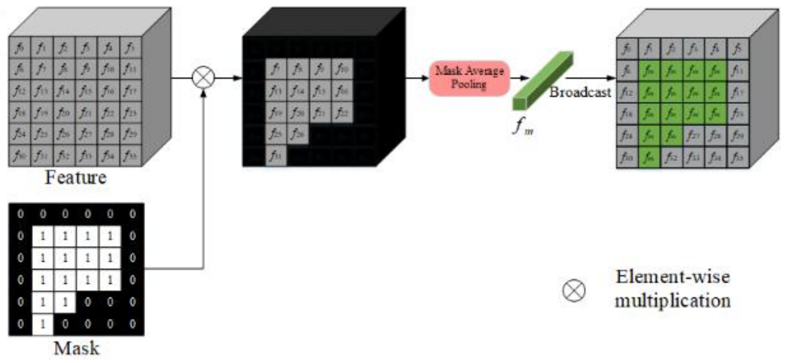


Figure 4. Details of mask average pooling layer.

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

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

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

M^{+} : the mask of prohibited item

$1-M^{+}$: the complementary mask

- It learns the shifting and scaling parameters separately based on the mask average pooling and the deviation maps to modulate the mean and standard deviation;
- It normalizes the local region adaptively with pixel-wise modulation parameters, to control the appearance of the region of interest.

Method

Loss function

$$\mathcal{L}_{total} = \mathcal{L}_{LL} + \mathcal{L}_{CSD} + \mathcal{L}_{recon} \quad (11)$$

Existing fusion loss: $\mathcal{L} = 1 - [w_1 \bullet \ell(x_1, y) + w_2 \bullet \ell(x_2, y)]$ (6)

- They produce global fusion, including the non-overlapping region, which should remain intact;
- The addition strategy does not conform to the attenuation character of X-ray imagery.

∵ According to TIP, When fusing the prohibited items I_f and baggage image I_b with weights w_1 and w_2, the fused X-ray image I_{fb} could be:

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

$$I_{fb} = \frac{I_f^{w_1} \bullet I_b^{w_2}}{I_0^{w_1 + w_2 - 1}} \quad (7)$$

∴ The logarithmic form for multiplication relationship.

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

$$\mathcal{L}_{LL} = 1 - [\log \ell(I_f, I_{fb}, w_1) + \log \ell(I_b, I_{fb}, w_2)] \quad (8)$$

$$\ell(x, y, w) = \frac{2\mu_x \mu_y + \varepsilon}{\mu_x^{4w} + \mu_y^{2} + \varepsilon}$$

So, the local mean of fused image μ_{fb} is expected to convere to $\mu_f^{w_1} \cdot \mu_b^{w_2}$

- Correlation loss of Saturation Difference (CSD) $\mathcal{L}_{CSD}$:

$$D_b = (1 - S_{fb}) / (1 - S_f)$$

$$D_f = (1 - S_{fb}) / (1 - S_b)$$

$$\mathcal{L}_{CSD} = 1 - [CC(D_f, 1 - S_f) + CC(D_b, 1 - S_b)] / 2 \quad (9)$$

- The reconstruction loss $\mathcal{L}_{recon}$ of two source images $\tilde{I}_f$ and $\tilde{I}_b$

$$\mathcal{L}_{recon} = \| I_f, \tilde{I}_f \| + \| I_b, \tilde{I}_b \| \quad (10)$$

Experiment and Results

■ Ablation Study —— Edge-Attention module

Table 1. Image quality evaluations of the models with and without the EA modules.

Model	without EA module	with EA module
MAE	2.091	1.938
PSNR	38.402	39.084

- The EA modules can activate and enhance texture and edge information by using the attention map. Thus it can alleviate the shortage of texture information and offer clearer appearance.
- This indicates the introduction of EA modules can effectively preserve the source information more completely in the resulting non-overlapping regions and improve the definition and sharpness of synthesized X-ray security images.

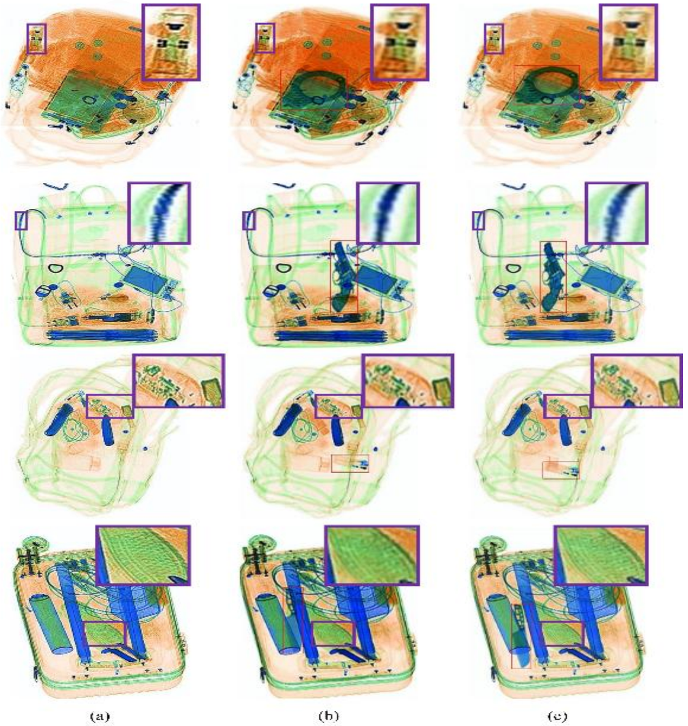


Figure 5. The images from the models with and without the EA modules. (a) baggage image; (b) without EA modules; (c) with EA modules.

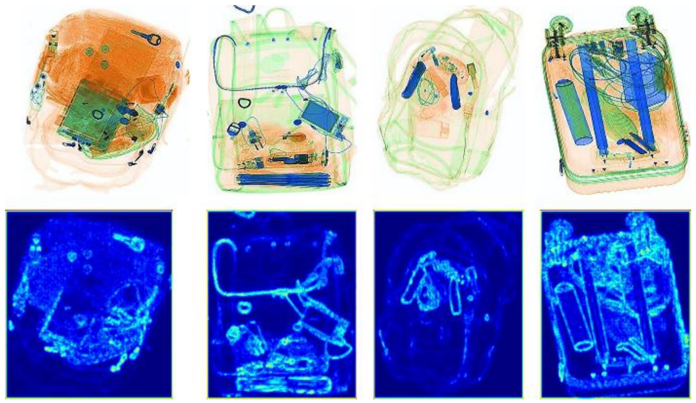


Figure 6. The edge attention maps extracted by the EA module.

Experiment and Results

Ablation Study —— Region-wise normalization

◆ The prohibited items synthesized by the model with only mask average pooling layer are fuzzy due to a lack of texture details, while the prohibited items from our model are more clear. It illustrates that our region-adaptive normalization mechanism allows for fine control to generate rich stylization by considering the local information from the deviation maps.

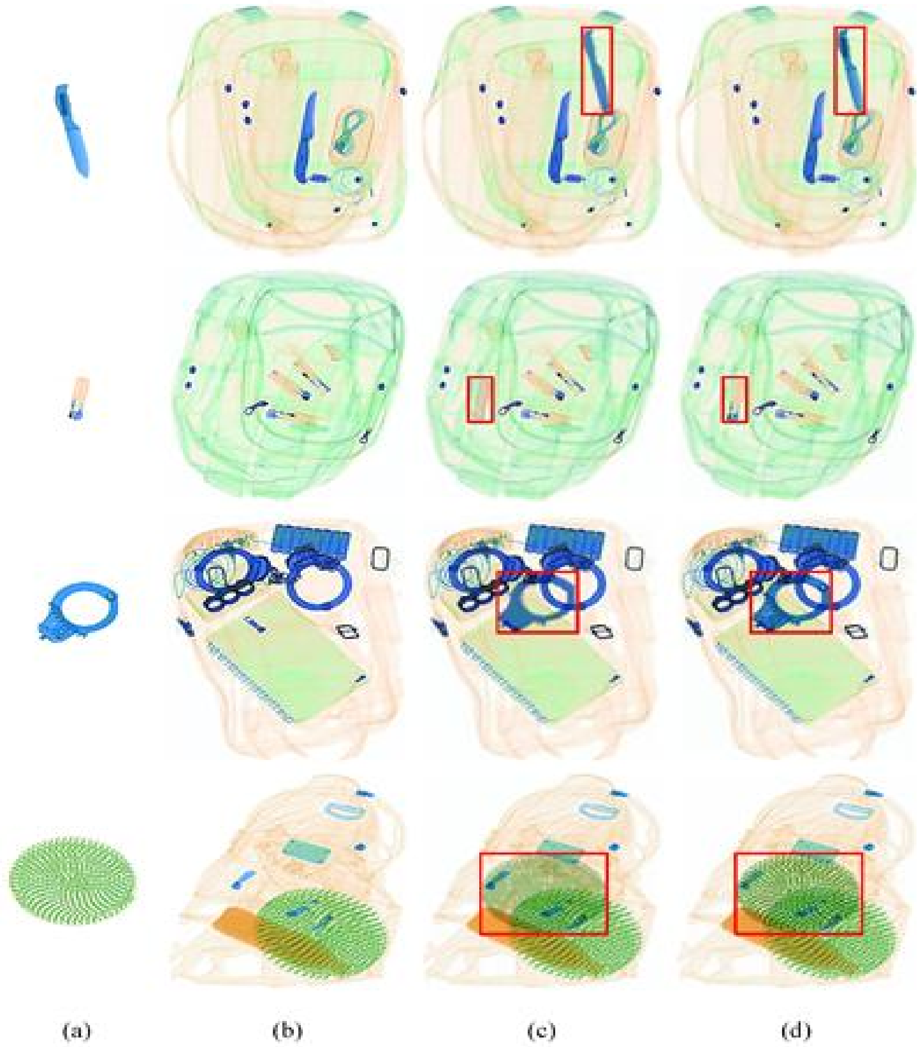


Figure 7. The images from the fusion module with only mask average pooling layer and RWSC-Fusion module. (a) prohibited items; (b) baggage image; (c) model with only mask average pooling layer; (d) our model.

Experiment and Results

Comparison with State-of-the-art Methods

U2Fusion [1], FusionGAN [2], DeepFuse [3] and MEFNet [4]

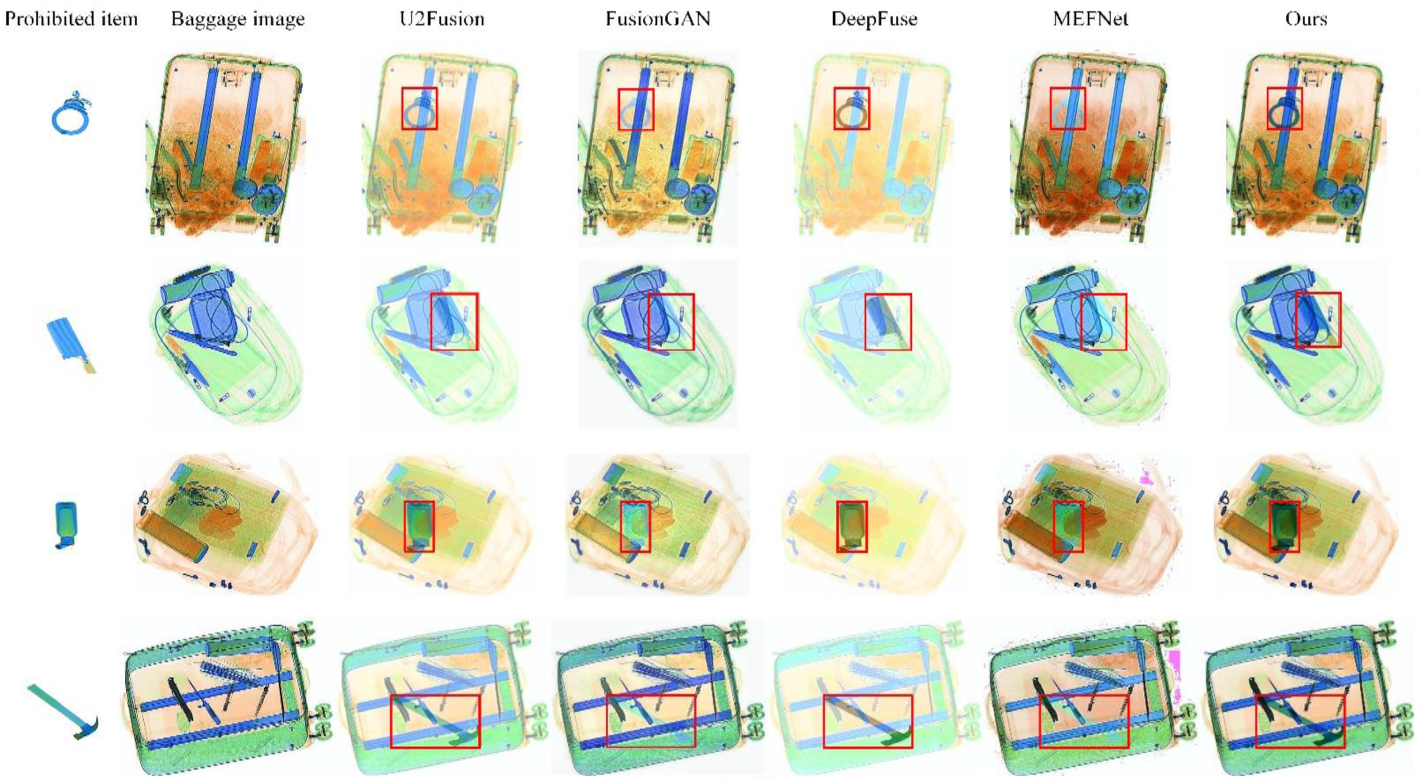


Figure 8. The X-ray baggage images fused with the prohibited item by four other fusion methods and RWSC-Fusion.

Table 2. Image quality comparison with state-of-the-art methods.

Metric	QE	Qabf	EI	PSNR
U2Fusion	0.5635	0.3467	48.8190	16.7155
FusionGAN	0.5640	0.5924	78.0934	13.8471
DeepFuse	0.6157	0.1814	41.6266	16.8173
MEFNet	0.7509	0.6090	85.0403	17.7364
Ours	0.8339	0.6472	81.3807	18.6455

- Our fusion model is better suited to the X-ray security image, and can synthesize more ideal and realistic prohibited X-ray security images.

[1] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, Haibin Ling, U2Fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell*, 2020.
[2] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. Fusiongan: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48:11-26, 2019.
[3] Prabhakar K R, Srikar V S, Babu R V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. In *Proc. IEEE Int. Conf. Comput. Vis*, 2017.
[4] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang. Deep guided learning for fast multi-exposure image fusion. *IEEE Transactions on Image Processing*. 29:2808-2819, 2020.

Experiment and Results

Supplement to SIXray and OPIXray Dataset

SIXray:

- **SIXR**: 7496 positive samples with 4322 guns, 2758 knives, 2816 wrenches, 4624 pliers and 918 scissors.
- **SIXR1/2**: We remove half the (3748) positive samples from SIXR, and supplement the other 3748 negative samples into SIXR.
- **SIXRS**: We synthesize the prohibited items into the newly-added 3748 negative samples in SIXR1/2, to supplement the positive samples, resulting in mixed 7496 true/pseudo positive samples with the same number of prohibited items as SIXR.
- **SIXRS+**: We synthesize prohibited items into other new 3748 negative samples, and replace 3748 negative samples in the SIXR, resulting in mixed 11244 true/pseudo positive samples, with 6469 guns, 4154 knives, 4195 wrenches, 6933 pliers and 1398 scissors.

Table 3. 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

Experiment and Results

Supplement to SIXray and OPIXray Dataset

OPIXray: (7109 training and 1776 testing images for cutter):

- **OPIR**: All the original 7109 images;
- **OPIRS**: 3555 images supplemented with synthesized cutter, with the same number of cutter as OPIR;
- **OPIRS+**: 7109 images supplemented with half the synthesized cutter.

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

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

Conclusion

RWSC-Fusion

Breakthrough

- We propose an unsupervised color X-ray security image fusion model with novel **LL** and **CSD** loss, extending TIP to composite color X-ray images.
- We propose a **Region-Wise Style-Controlled Fusion** module, to control the fused appearance by learning the shifting and scaling modulation parameters pertinently.
- We develop an **Edge-Attention** module, which inhibits irrelevant information and enhances texture information, to improve the sharpness of generated images.

Significance

- RWSC-Fusion outperforms the state-of-the-arts in the field of X-ray security image.
- Our synthetic images are authentic and reliable enough in promoting down-stream tasks.

THANK YOU !