Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IEF-YOLO: Infrared simulation image Fidelity Evaluation algorithm based on YOLO detection model

Ruitao Lu[1]*, Zhanhong Zhuo[1], Guanchen Yue[2], Yiran Gong[1]
[1]*Department of Automation, Rocket Force University of Engineering, Xi'an, 710025, China*
[2] *College of Business Administration, Northeastern University, Shenyang, 110819, China*
lrt19880220@163.com

ARTICLE INFO

Received: 29 Dec 2024

Revised: 12 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Infrared scene simulation generation technology is an important component of infrared imaging guidance hardware-in-the-loop simulation. However, there are significant differences in the infrared simulation images generated by different simulation platforms, and how to effectively analyze and evaluate the realism of infrared visual simulation generated images is a research difficulty. This paper proposes an infrared simulation image fidelity evaluation algorithm based on object detection tasks, IFE-YOLO. The model integrates Swin Transformer, YOLOX, and attention mechanism to achieve fidelity evaluation at the target task level. Firstly, we propose that STCNet backbone network extracts target feature information, which not only has Swin Transformer's excellent ability to model based on global information, but also uses attention mechanism to capture location information and channel relationship to enhance the ability to process feature information. Secondly, based on the improved PANet, a feature pyramid network is constructed for deep fusion of high-level and low-level features to effectively use semantic information and location information. Thirdly, decoupling detection head and improved position loss function are used in the target detection part to improve the model performance. Finally, the infrared fidelity evaluation process for detection task is designed, and the validity of the model is verified on the constructed infrared simulation image dataset.

Keywords: Fidelity Evaluation, Infrared simulation image, YOLO detection model.

INTRODUCTION

With the increasing importance of scene simulation technology in the simulation field, the research on infrared scene simulation generation technology is also more in-depth. The fidelity of infrared image simulation is the key factor of visual simulation generation system. The fidelity of target and background simulation will directly affect the accuracy of performance evaluation of complex photoelectric equipment, the effectiveness of data set preparation, and the reliability of intelligent cognitive algorithm. Therefore, in-depth research on infrared simulation image analysis and evaluation technology [1] is of great significance for improving the performance of simulation system.

Current simulation data fidelity evaluation algorithms mainly include traditional evaluation methods and evaluation methods based on deep learning. Traditional simulation data fidelity analysis methods appeared in the 1990s, Mason [2] evaluated the verisimilitude of the DIRSIG model from the perspective of radiant energy by designing experiments and analyzing the contrast between the simulation image and the actual image. Zhou [3] can capture the characteristics of image structure features through HVS, and proposed an image quality evaluation method based on structural similarity (SSIM). Tang [4] established a fidelity evaluation system for the combat visual simulation system based on the fuzzy analytic hierarchy process theory. Li [5] realized a simulation system evaluation based on fidelity, providing a method for quantitative evaluation of simulation systems. Wang [6] studied the factor decomposition of fidelity analysis and the index aggregation method of index standardization. Du [7] proposed a comprehensive evaluation framework of visual simulation fidelity combining subjective and objective from the perspective of human visual system's perception and cognition of information. However, the traditional algorithm is only image level evaluation, and the evaluation effect is not good under complex tasks. Using deep learning algorithm to evaluate simulation degree is the focus of future research.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Kang [8] extracted statistical information of natural scenes by modifying the input form of convolutional neural network, and applied multitask convolutional network to evaluate image quality and distortion. Kim [9] used convolutional neural network to learn features, and divided the network into training stage and prediction stage to improve its learning ability for distorted images. Lv [10] proposed an image quality evaluation algorithm based on depth learning by studying the relationship between the attention mechanism of convolutional network and image quality. Hou [11] proposed an integrated intelligent flexible simulation evaluation method combining big data and deep confidence network. Dong [12] proposed a simulation image evaluation method based on the generation countermeasure network by using the idea of image generation and discrimination in the generation countermeasure network. Quan [13] proposed a fidelity evaluation algorithm based on random forest, and realized the evaluation results consistent with human vision. Cao [14] summarized the image quality evaluation methods for in-depth learning in recent years. However, the current algorithms based on deep learning are mostly based on feature level fidelity evaluation, and there is still no research on the fidelity in complex task environments containing targets.

Therefore, this paper proposes an infrared image simulation fidelity evaluation algorithm based on target detection task, IFE-YOLO. The model combines Swin Transformer's dynamic attention and global modeling capabilities, YOLOX's strong feature fusion performance, and the target features of attention mechanism. The backbone adds the Patch Embedding module after the input layer, and then build feature maps of different sizes through multiple Swin Transformer Blocks and CA Patch Merging layers. Use the PANet path aggregation network to deeply fuse the high-level feature information, and realize the effective use of semantic information and location information. The classification and regression tasks of target detection are handled separately, and EIOU is used as the location loss function to accelerate convergence, solve the sample imbalance problem, and enhance the model generalization ability. Finally, the infrared fidelity evaluation process for detection task is designed.

METHODOLOGY

The IFE-YOLO network structure is shown in Figure 1. In the backbone network, we propose a backbone network called STCNet, which integrates the advantages of Swin Transformer and CA attention mechanism. Compared with the traditional backbone feature extraction network based on CNN, STCNet has the ability of dynamic attention and global modeling which considers the remote dependency. STCNet network adopts a hierarchical architecture, which is composed of three parts: Patch Embedding layer, Swin Transformer Block and CA Patch Merging layer. In the neck network, we still use PANet to build a feature pyramid for deep feature fusion. In addition, we also introduced SE and CBMA attention mechanisms in the neck to enhance the attention to target information and further improve the model performance. IFE-YOLO uses more advanced EIOU Loss to accelerate convergence and improve model performance.

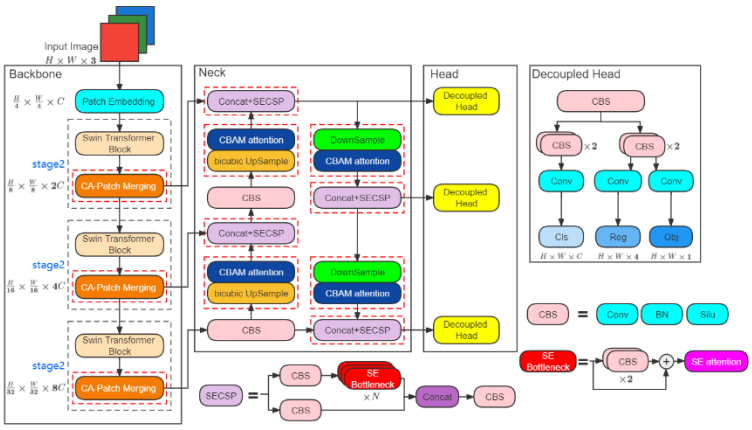


Figure. 1 IFE-YOLO network structure.

2.1 Backbone

2.1.1 Patch Embedding Layer

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The patch embedding module first chunks the image at the front end of the feature extraction network, dividing the image into 4×4 non-overlapping blocks so that the feature dimension of each block is 4×4×3. Then, the original 2D image is converted into a series of 1D embedding vectors by projecting the feature dimensions to arbitrary dimensions through linear transformation, and the transformed embedding vectors are input to three-stage feature extraction layers to generate a hierarchical feature representation.

2.1.2 Swin Transformer Block

Swin Transformer block [17,18] calculates the attention between pixels in the form of a moving window, which helps to connect the front window, reduce the complexity of the original attention calculation, overcome the lack of global effects, and significantly enhance the modeling effect. The multiheaded self-attention (MSA) mechanism in the Swin Transformer Blocks is constructed based on the shift window. There are two consecutive Swin Transformer Blocks. Each Swin Transformer Block consists of a LayerNorm (LN) layer, an MSA module, a residual connection, and a multilayer perceptron (MLP) that contains two fully connected layers using the GELU nonlinear activation function. The two consecutive Swin Transformer Blocks adopt the window multi-head self-attention (W-MSA) module and the shifted window multi-head self-attention (SW-MSA) module, respectively, which enables different windows to exchange information while reducing computational effort. Based on this window division mechanism, the continuous Swin Transformer Blocks are calculated as follows:

$$\hat{z}^{i} = \mathrm{W-MSA}\left(\mathrm{LN}\left(z^{i-1}\right)\right) + z^{i-1} \tag{1}$$

$$z^{i} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{z}^{i}\right)\right) + \hat{z}^{i} \tag{2}$$

$$\hat{z}^{i+1} = \mathrm{SW-MSA}\left(\mathrm{LN}\left(z^{i}\right)\right) + z^{i} \tag{3}$$

$$z^{i+1} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{z}^{i+1}\right)\right) + \hat{z}^{i+1} \tag{4}$$

where $\hat{z}^{i}$ denotes the output of the (S)W-MSA module and z^{i} denotes the output of the MLP module of the i^{th} Block.

2.1.3 CA-Patch Merging

The Patch Merging layer [19,20] plays a pooling role in the backbone network, which can reduce the resolution of the feature map and adjust the number of channels, thus forming a layered design. At the same time, it can also save a certain amount of computation. Its working process is shown in Figure 2.

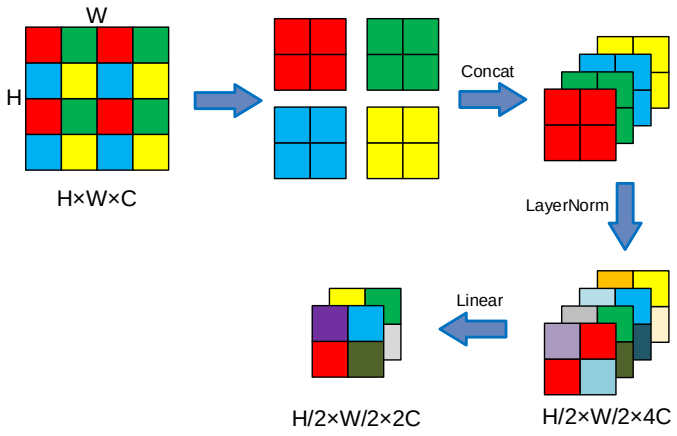


Figure. 2 Schematic diagram of the Patch Merging layer.

Considering the limited context encoding capability of Swin Transformer, we added CA attention mechanism after the Patch Merging layer. CA attention decomposes the working process of channel attention into two one-dimensional feature coding processes, and then aggregates features along two directions in the space。

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\begin{cases} z_c^h(h) = \dfrac{1}{W} \sum\limits_{0 \le i < W} x_c(h,i) \\ z_c^h(w) = \dfrac{1}{H} \sum\limits_{0 \le j < H} x_c(j,w) \end{cases} \tag{5}$$

The above transformation obtains the feature map in both width and height directions in the space, and then CA pays attention to stitching the feature map first, and then carries out F_1 transformation to obtain the feature map f :

$$f = \delta(F_1([z^h, z^w])) \tag{6}$$

where F_1 is 1×1 convolution transformation function, [,] indicates splicing operation, δ is the nonlinear activation function.

Then the feature map f is convolved in the original height width direction, activated by the sigmoid activation function, and the attention weights g^h and g^w of the feature map are obtained.

$$\begin{cases} g^h = \sigma(F_h(f^h)) \\ g^w = \sigma(F_w(f^w)) \end{cases} \tag{7}$$

where σ is sigmoid activation function.

Finally, the CA attention mechanism obtains the feature map output with attention weight through multiplication weighted calculation

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{8}$$

2.2 Neck

CSPLayer [21] is mainly divided into two parts. The main part consists of convolution branch and residual branch, and the rest is directly connected to the CSPLayer output part through a 1×1 convolution layer。The Bottleneck structure of the residual module is an important part of CSPnet [22]. Its stack consists of a 1×1 convolution and a 3×3 convolution. At the end, a shortcut connection is used to add the initial input directly to the output of the convolution layer. The use of CSPLayer makes the model over consider the surrounding context information, thus causing local information interference. To solve this problem, we introduce SE attention mechanism in Bottleneck module to selectively emphasize feature information, weaken interference information, and further enhance the attention to target features. At the same time, the 3×3 convolution layer in Bottleneck needs to process a large number of parameter operations, which results in a large number of parameter redundancy. SE compresses the feature map along the spatial dimension, extrudes the global spatial information into the channel description, and the output feature map z of channel c after compression is

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{9}$$

where x_c is input, and H and W represent the height and width of the space respectively. This process greatly reduces the redundant parameters in the network.

2.3 Fidelity calculation based on target detection task

The target detection network designed above is trained based on the self-made infrared simulation training data set, and the model weight is retained. Then in the simulation image similarity evaluation phase, the similarity evaluation of any two input images can be completed by loading the trained model weights. The specific assessment process is as follows:

First, use the trained target detection model to detect the two images to be evaluated containing the detected target in turn, and obtain their respective detection results, including detection category, confidence and bbox detection frame coordinates.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Then, compare the detection results of the two images in turn, and screen out the target results that have the same detection category and the coordinates of the bbox detection box meet the set cross merge ratio threshold.

Finally, the evaluation similarity of the test results that do not meet the screening conditions is 0. The corresponding evaluation similarity of the test results that meet the screening conditions is obtained by calculating the confidence difference of the screening results. Finally, the evaluation similarity of all the test results is traversed and the average value is taken as the final similarity of the two images.

The target detection results of the two images are $(O_{A1},\ O_{A2},\cdots,\ O_{An})$ and $(O_{B1},\ O_{B2},\cdots,\ O_{Bm})$. The detection confidence levels corresponding to the two images are $(P_{A1},\ P_{A2},\cdots,\ P_{An})$ and $(P_{B1},\ P_{B2},\cdots,\ P_{Bm})$, where $n \geq m$. The fidelity can be calculated by

$$s(A,B)=\begin{cases}0 & IoU(O_A,O_B)<\theta\\ 1-\left|P_A-P_B\right| & IoU(O_A,O_B)\geq\theta\end{cases} \tag{10}$$

$$S=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}s(O_{Ai},O_{Bj})}{n} \tag{11}$$

where θ is the set intersection to union ratio threshold, S is the similarity of the two images based on the target detection evaluation.

RESULTS

3.1 Dataset and Experimental Environment

The data set of this paper is generated based on infrared simulation software, including 2000 infrared simulation images, including aircraft, tanks, transport vehicles and other targets. The training set and test set were randomly divided in an 8:2 ratio, and 20% of the training set was randomly selected as the validation set. In order to increase the diversity of data and ensure that the model has better training results, two data enhancement methods, Mosaic and Mixup, are used to perform data enhancement operations on the data sets.

The experiment environment of this article uses the Ubuntu18.04 operating system, carries the Intel Core i9 10980 XE CPU, NVIDIA RTX 2080TI graphics card, and 11 GB of memory. The Deep Learning Framework employs Python, accelerating training with CUDA10.1 and CUDNN7.6. The main settings of the experiment parameters in this paper are as follows: the training period is set to 300 iterations, the maximum learning rate of the model is 0.01, the minimum learning rate is 0.0001, the "SGD" random gradient descent is selected for the optimization, the weight decay is 0.0005, the learning rate descent mode is "COS" cosine descent, and multi-threaded data reading is adopted to speed up data reading. The predicted probability threshold at the time of testing was 0.5 with a confidence of 0.001 and a NMS threshold of 0.65. The input image sizes are all 512×512.。

3.2 Analysis of fidelity results based on brightness adjustment

A contrast group is generated using a brightness adjustment algorithm based on HSV space, in which the brightness of the entire image can be adjusted within a certain range by adjusting the value of brightness channel V.

Decomposing image I into HSV space results in three channels $(I_H\ \ I_S\ \ I_V)$ with the value of brightness channel I_V adjusted by the adjustable parameter v to achieve brightness adjustment of the image, calculated as follows:

$$a=\begin{cases}1/(1-v), 0\leq v\leq 1\\ 1+v, -1\leq v<0\end{cases} \tag{12}$$

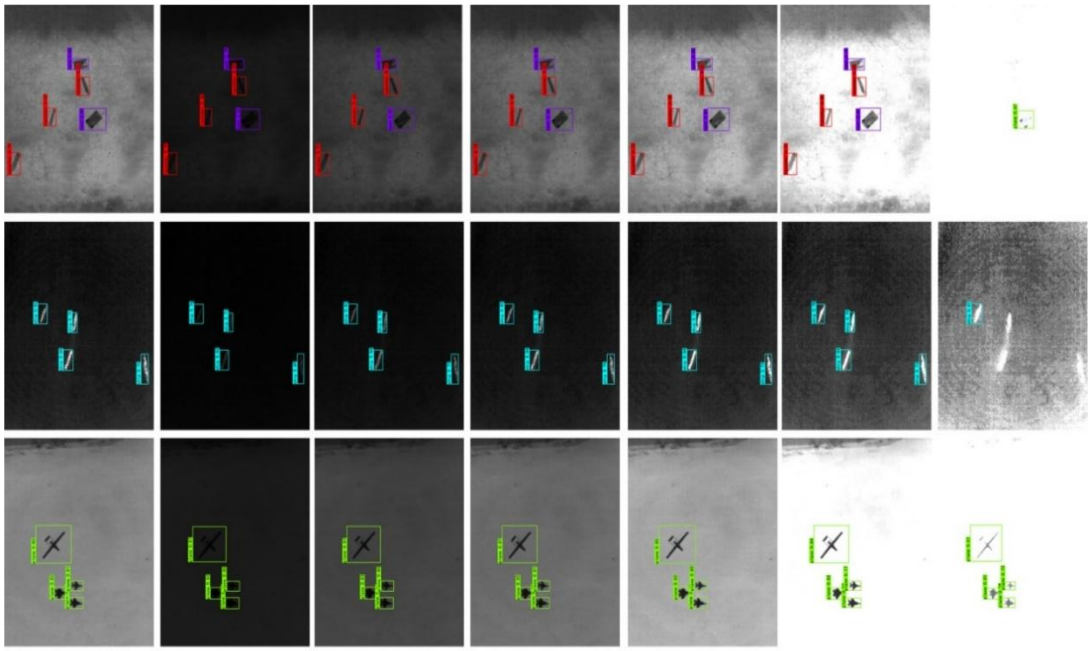
$$I_V'=I_V\alpha \tag{13}$$

Based on the above formula, the image brightness can be adjusted by adjusting the control parameter v, where when v is less than 0, the image brightness decreases, and vice versa. Setting the v values sequentially to $[-0.8,-0.5,-0.2,0.2,0.5,0.8]$ gives the brightness adjusted. As shown in figure 3. Table 1 gives the results of the fidelity evaluation of the brightness adjustment. From the results, we can see that the IFE-YOLO evaluation algorithm

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

proposed in this paper can achieve effective evaluation of infrared simulation image quality at the target detection mission level. After adjustment of brightness, the evaluation indicators are effective and discernible.



(a)Origin (b)v=-0.8 (c)v=-0.5 (d)v=-0.2 (e)v=0.2 (f)v=0.5 (g)v=0.8

Figure. 3 Comparison of target detection results for brightness adjustment

Table 1 Fidelity Assessment of Brightness Adjustment

	v =-0.8	v=-0.5	v=-0.2	v=0	v=0.2	v=0.5	v =0.8
Image1	0.974	0.992	0.994	1	0.988	0.974	0.0
Image2	0.955	0.9925	1.0	1	0.995	0.985	0.1975
Image3	0.995	0.990	0.9975	1	0.995	0.985	0.8775

3.3 Analysis of fidelity results based on contrast adjustment

The contrast adjustment algorithm mainly adjusts the image's RGB space. The adjustment of image contrast can be achieved by setting a suitable threshold and calculating a suitable adjustment parameter based on this threshold. First compare the size of the current pixel 3D color value with the threshold value and obtain its difference, then use the adjusted coefficient index to amplify the difference when the contrast needs to be increased; When a reduction is needed, the reduced difference is linear using the adjustment coefficient.

Image $I_{RGB}=(I_R,I_G,I_B)$ can adjust the value of each channel through parameter c to achieve image contrast adjustment, the calculation process is as follows:

$$\beta=\begin{cases}1/(1-c)-1, 0\le c\le 1\\ c, -1\le c<0\end{cases} \tag{14}$$

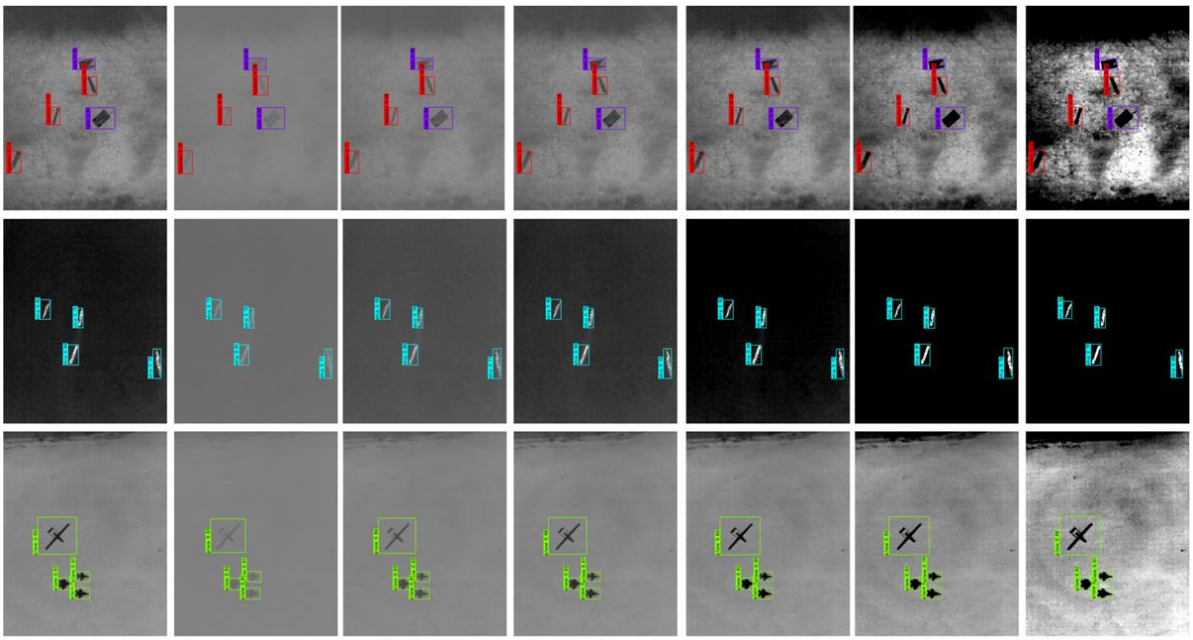
$$I'_{RGB}=I_{RGB}+(I_{RGB}-\theta\times 255)\times\beta \tag{15}$$

Setting the θ threshold to 0.5, the brightness of the image can be adjusted by adjusting the control parameters, where c is less than 0, the image contrast decreases, and vice versa. Setting the c values sequentially to $[-0.8,-0.5,-0.2,0.2,0.5,0.8]$ gives the contrast adjusted. As shown in Figure 4. Table 2 gives the results of the fidelity evaluation of the contrast adjustment. Contrast adjustment can easily affect the accuracy of analysis for common evaluation indicators. From the results, we can see that the fidelity evaluation algorithm proposed in this paper still

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

maintains good evaluation performance on the test images with drastic contrast changes, can distinguish the similarity between the tested images, and has strong robustness.



(a)Origin (b)c=-0.8 (c)c=-0.5 (d)c=-0.2 (e)c=0.2 (f)c=0.5 (g)c=0.8

Figure. 4 Comparison of target detection results for contrast adjustment

Table 2 Fidelity Assessment of Contrast Adjustment

	c=-0.8	c=-0.5	c=-0.2	c=0	c=0.2	c=0.5	c=0.8
Image1	0.98	0.99	0.994	1	0.998	0.98	0.952
Image2	0.9625	0.99	0.9975	1	0.99	0.9625	0.915
Image3	0.9875	0.9925	0.99	1	0.9975	0.9925	0.995

3.4 Fidelity result evaluation based on noise adjustment

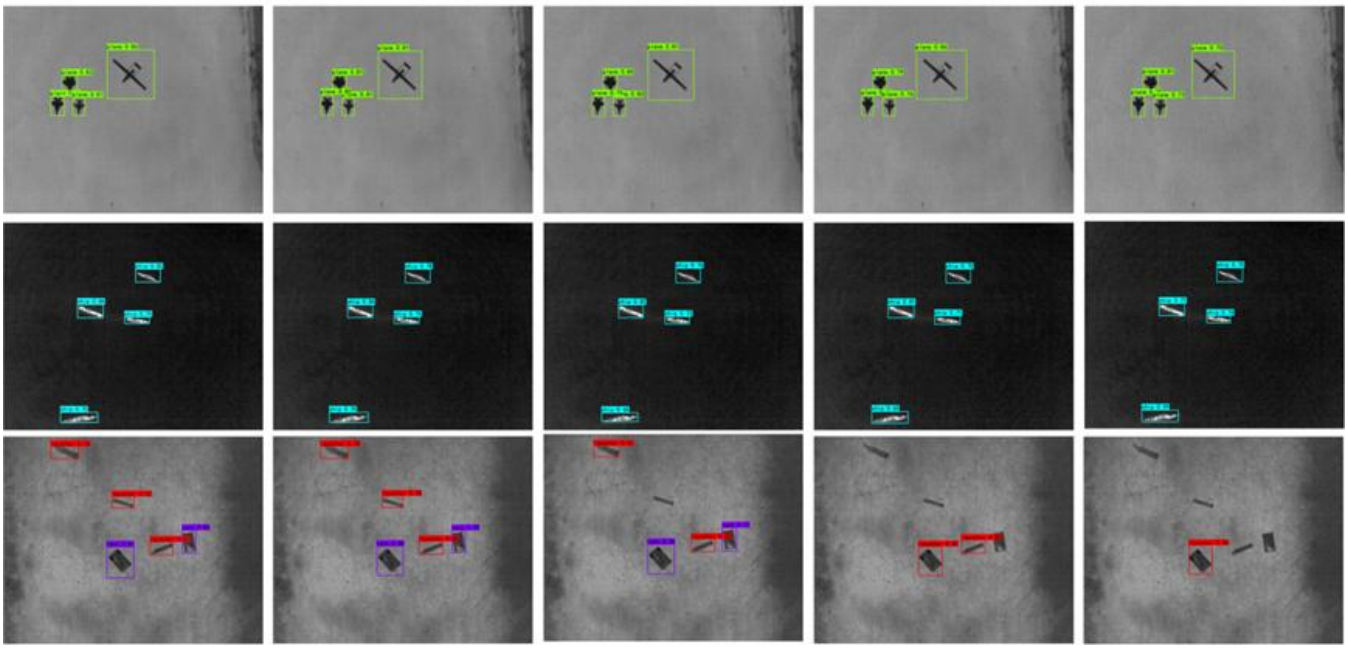
A control group was constructed by adding a gauss noise adjusted image, and the ambiguity of the noise was adjusted by adjusting the variance σ of the noise.:

$$I = I + \mathrm{N}(0, \sigma) \tag{16}$$

Setting the σ values sequentially to $[2,5,8,10]$ gives the noise adjusted. As shown in Figure 5. Table 3 gives the results of a fidelity assessment incorporating the effects of noise. Adding noise can affect the value of local pixels, making fidelity evaluation difficult. Traditional fidelity evaluation algorithms are affected by noise, resulting in inaccurate evaluation results. From the results, we can see that the IFE-YOLO fidelity assessment algorithm proposed in this paper can still assess the tested images well under the influence of noise, which shows a good environmental adaptability.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



(a)Origin (b) σ=2 (c) σ=5 (d) σ=8 (e) σ=10

Figure. 5 Comparison of target detection results for noise adjustment

Table 3 Fidelity Assessment of Noise Adjustment

	σ= 0	σ= 2	σ=5	σ= 8	σ= 10
Image1	1	0.98	0.9625	0.9725	0.94
Image2	1	0.99	0.95	0.9275	0.91
Image3	1	0.982	0.74	0.182	0.0

DISCUSSION

A new infrared simulation image fidelity evaluation algorithm IFE-YOLO based on target detection mission is proposed in this paper. The STCNet backbone network is proposed to extract target feature information, and a feature pyramid network is constructed based on the improved PANet for deep fusion of low-level features. Sorting of target detection was handled separately from the regression mission, using EIOU as a function of location loss. Finally, the fidelity analysis flow of the detection mission level is constructed. The simulation results show that the IFE-YOLO proposed in this paper has good discrimination and robustness, and can provide theoretical support for infrared image fidelity analysis. In the future, the paper will combine deep semantic information, domain adaptation information, content information and so on to further build a deep learning fidelity analysis framework to improve the effectiveness of analysis in complex scenarios.

REFRENCES

[1] Lu R, Shen T, Yang X, et al. Analytic Hierarchy Process-based Comprehensive Assessment Algorithm of Infrared Simulated Image Fidelity[J]. Journal of Rocket Force University of Engineering, 2024, 38(06):23-30.
[2] Mason J, Schott J, Rankin P. Validation analysis of the thermal and radiometric integrity of RIT's synthetic image generation model, DIRSIG[C]. Proceedings of the International Society for Optical Engineering, 1994, 2223(4):474-487.
[3] Wang Z, Bovik A, Sheikh , et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
[4] Tang K, Kang F. Evaluation Method of Visualization Simulation Fidelity Based on Fuzzy AHP. Journal of System Simulation, 2008, (22): 6049-6053+6057.
[5] Li J. Research on evaluation of simulation systems based on fidelity [D]. Harbin Institute of Technology,2008.

Research Article

[6] Wang G, Shen X, Ye L, et al. Research on Analysis Method of Target Simulation Fidelity[J]. Computer Simulation, 2015,32(02):6-10+101.
[7] Du J, Liang Q, Yao F. Evaluation Method of Visualization Fidelity in Virtual Battlefield Environment [J]. Journal of System Simulation, 2013,25(08):1891-1895.
[8] Kang L, Ye P, Li Y, et al. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks[C]. IEEE international conference on image processing(ICIP), 2015: 2791-2795.
[9] Kim J, Nguyen A, Lee S. Deep CNN-based blind image quality predictor[J]. IEEE transactionson neural networks and learning systems, 2018, 30(1): 11-24.
[10] Lv X, Qin M, Chen X, Guo W. No-Reference Image Quality Assessment Based on Statistics of Convolution Feature Map[C]. The 2nd International Conference on Advances in Materials, Machinery, Electronics, 2018:1–5.
[11] Hou L, Zeng W, Fan H, et al. Research of Integrated Intelligent Flexible Simulation Evaluation Technology [J]. Fire Control & Command Control,2019,44(11):174-179+185.
[12] Dong C. Research on Simulation Image Evaluation Based on Deep Learning[D]. Xi Dian University, 2020.
[13] Quan X, Tan Q, Chen D, et al. Quality Assessment Study of Infrared Simulation Image [J]. Flight Control & Detection,2025,1-14.
[14] Cao Y, Liu H, Jia X, et al. Review of image quality assessment methods based on deep learning [J]. Computer Engineering and Applications, 2021, 57(23): 27-36.
[15] Wang J, Shao Z, Huang X, et al. Pan-Sharpening via Deep Locally Linear Embedding Residual Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-13.
[16] Cai Q, Qian Y, Li J, et al. HIPA: Hierarchical Patch Transformer for Single Image Super Resolution[J]. IEEE Transactions on Image Processing, 2023, 32: 3226-3237.
[17] Wang Y, Tong L, Yang J. Swin Transformer Embedding MSMDFFNet for Road Extraction From Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 1-5.
[18] Chen X, Zhao C, Liu X, et al. An Embedding Swin Transformer Model for Automatic Slow-Moving Landslide Detection Based on InSAR Products[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024,62: 1-15.
[19] Ge L, Dou L. SFP-NMS: Nonmaximum Suppression for Suppressing False Positives During Merging Patches of High-Resolution Image[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21:1-5.
[20] Falk D, Aydin K, Scheibenreif L. and Borth D. Merging Patches and Tokens: A VQA System for Remote Sensing[C]. 2024 IEEE International Geoscience and Remote Sensing Symposium, 2024,:694-698.
[21] Xia W, Li P, Huang H, et al. TTD-YOLO: A Real-Time Traffic Target Detection Algorithm Based on YOLOV5[J]. IEEE Access, 2024, 12: 66419-66431.
[22] Li F, Sun T, Dong P, et al. MSF-CSPNet: A Specially Designed Backbone Network for Faster R-CNN[J]. IEEE Access, 2024, 12: 52390-52399.
[23] Ma J, Jiang W, Tang X, et al. Multiscale Sparse Cross-Attention Network for Remote Sensing Scene Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-16.
[24] Peng C, Li X, Wang Y. TD-YOLOA: An Efficient YOLO Network With Attention Mechanism for Tire Defect Detection[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-11.