# Ultra Low Power Multimodal Remote Sensing Night Emergency Search and Rescue Model for Resource Constrained Unmanned Platforms


Kefan Li
*Wuhan University*
430072 Wuhan, China
likefan@whu.edu.cn

Yuting Wan
*Wuhan University*
430072 Wuhan, China
wanyuting@whu.edu.cn

Haoyu Yao
*Wuhan University*
430072 Wuhan, China
2022302131127@whu.edu.cn

Shiyu Liu
*Wuhan University*
430072 Wuhan, China
3218201562@qq.com

Ailong Ma
*Wuhan University*
430072 Wuhan, China
maailong007@whu.edu.cn

Yanfei Zhong
*Wuhan University*
430072 Wuhan, China
zhongyanfei @whu.edu.cn



***Abstract*—In nighttime emergency rescue scenarios involving unmanned aerial vehicles, traditional artificial neural network (ANN)-based object detection algorithms are challenging to efficiently deploy on resource-constrained edge computing platforms due to their high computational complexity and energy consumption. In contrast, spiking neural networks (SNN), which emulate the brain's information processing through sparse event-driven computations, significantly improve energy efficiency. Low-light and infrared images, as critical sources of information for nighttime rescue missions, can provide clear feature representations under low-light and complex environmental conditions. The efficient computational characteristics of SNN are highly compatible with the properties of these modalities, further enhancing their potential for nighttime emergency rescue tasks. For this emergency rescue scenario, this paper proposes a low-power object detection method for nighttime emergency rescue tasks, leveraging the fused characteristics of low-light and infrared imagery and the energy-efficient computational advantages of SNN. The method is based on Spiking-YOLO and was tested on self-constructed emergency rescue tasks. Experimental results demonstrate that, compared to the ANN-based Tiny YOLO, the proposed method achieves theoretical power consumption reduced to 1/2000 of the original network, with performance metrics degradation not exceeding 2%.**



***Keywords—nighttime emergency rescue, SNN, low-power, Low-light and infrared imagery, object detection***


## I. INTRODUCTION

Emergency rescue operations play a critical role in saving lives and mitigating losses during disasters and accidents. These operations often rely on advanced technologies to efficiently locate victims and assess hazardous environments. In recent years, unmanned aerial vehicles (UAVs) have emerged as indispensable tools in emergency rescue scenarios due to their mobility, adaptability, and ability to access hard-to-reach areas[1]. By equipping UAVs with object detection systems, rescue teams can rapidly identify survivors, obstacles, or key features in disaster-stricken areas, greatly enhancing operational efficiency.


This work was supported by the National Key Research and Development Program of China under Grant 2022YFB3903404, in part by the National Natural Science Foundation of China under Grant No. 42301418, the China National Postdoctoral Program for Innovative Talents under Grant BX20230275, the Open Fund of Key Laboratory of China-ASEAN Satellite Remote Sensing Applications, Ministry of Natural Resources of the People' s Republic of China (Grant No. ZDMY202305). (Corresponding authors: Yuting Wan).


Nighttime rescue missions represent a vital yet particularly challenging aspect of emergency response operations. Limited visibility and complex environmental conditions significantly reduce the effectiveness of traditional visible-light imaging. Low-light and infrared imaging[2], as key technologies to address these limitations, can capture clear features in dark environments. While these imaging modalities provide the hardware foundation for nighttime object detection, they demand more advanced computational methods to fully leverage their capabilities. Traditional ANN-based object detection algorithms, such as Tiny YOLO[3], although effective in daytime scenarios, struggle to meet the requirements of nighttime missions. Their high computational complexity and energy consumption make them unsuitable for resource-constrained edge computing platforms, which are commonly used in UAV-based systems. This tradeoff between computational demand and energy efficiency limits the deployment of ANN models in nighttime rescue scenarios, particularly when balancing system endurance and real-time performance.

Real-time processing is a non-negotiable core requirement for emergency rescue missions. Whether it involves post-disaster search and rescue or fire monitoring, object detection systems must provide fast and accurate inference to ensure timely decision-making. At the same time, energy consumption poses a significant challenge in nighttime rescue scenarios. Object detection systems on UAVs often rely on high-performance processors, but the enhanced computational capability comes at the cost of high energy consumption, directly impacting the mission duration of UAVs. For example, traditional ANN models rely on intensive floating-point computations and resource-demanding operations, making it difficult to operate on compact, low-power platforms for extended periods[4]. Addressing the dual challenges of real-time performance and energy efficiency is therefore critical to advancing nighttime rescue technologies.

To tackle these challenges, this paper proposes a novel low-power object detection method specifically designed for nighttime emergency rescue missions. The proposed method is based on spiking neural networks (SNN), which simulate the brain's information processing through sparse, event-driven computations, achieving significantly higher energy efficiency compared to the dense computations of traditional ANNs. Our main contributions are:

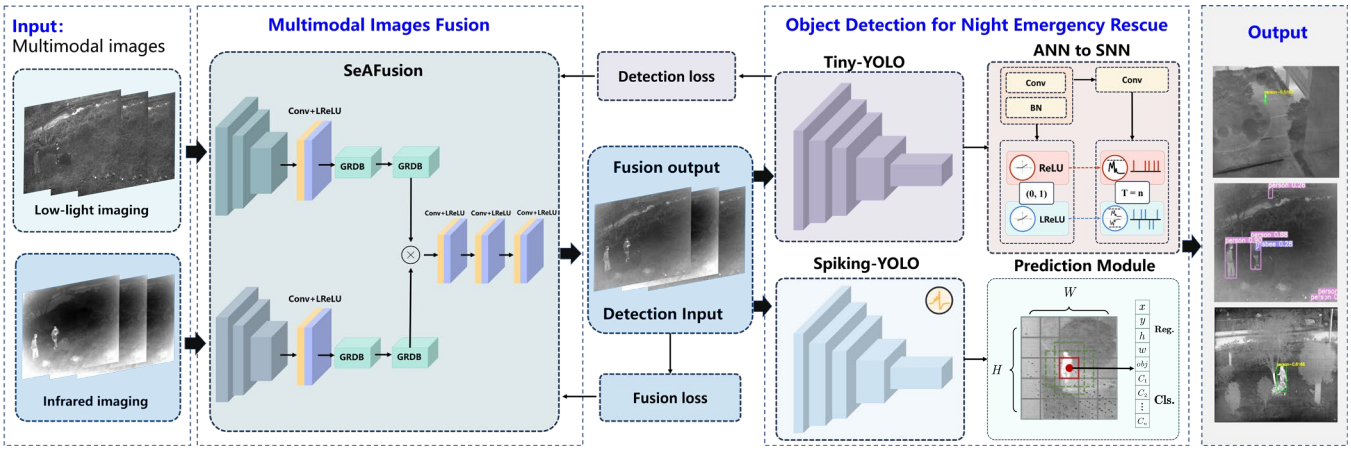



Fig. 1. Overall framework of multimodal remote sensing night emergency search and rescue model.

*1) A Low-Power Object Detection Framework:* A Spiking-YOLO-based framework is developed for nighttime emergency rescue scenarios, leveraging the event-driven computational advantages of Spiking Neural Networks (SNN). This framework significantly reduces energy consumption while maintaining high detection accuracy.

*2) Fusion of Low-Light and Infrared Imaging Modalities:* To enhance feature representation and improve object detection performance under nighttime and complex environmental conditions, a fusion strategy is employed to combine the complementary characteristics of low-light and infrared imagery.

*3) Extensive Evaluation on Emergency Rescue Tasks:* The proposed method is extensively evaluated on simulated emergency rescue scenario. Experimental results demonstrate a theoretical power consumption reduction to 1/2000 of the ANN-based Tiny YOLO, with performance degradation not exceeding 2%, validating the method's efficacy in real-world applications.

## II. Proposed Method

Under nighttime rescue conditions, low-light images effectively enhance visual information in low-illumination environments[5], presenting contours and texture details of target objects, and providing clear visual data in scenarios where human vision is limited. Thermal infrared images, by capturing the thermal radiation emitted by objects rather than relying on visible light, exhibit strong detection capabilities, particularly for targets with significant temperature differences. By leveraging the complementary characteristics of these two imaging modalities, we employ the SeAFusion model[6] to perform frame-by-frame fusion of video data, using the fused images for subsequent target detection tasks. To minimize power consumption on resource-constrained platforms and ensure sustained hardware operation, we introduce the Spiking-YOLO model[7] for nighttime multi-scenario emergency rescue conditions. This approach achieves efficient target detection on low-light and thermal infrared fused images while maintaining energy efficiency.

### A. Fusion of low-light infrared images

The images produced by the fusion of low-light and infrared images not only contain salient objects and rich texture details, but also are conducive to high-level vision tasks. Although image fusion algorithms based on deep learning can produce satisfactory fused images in recent years, existing fusion algorithms tend to pursue better visual quality and higher evaluation indicators, and rarely systematically consider the fused image Whether it can meet the requirements of high-level visual tasks.

SeAFusion employed in this study incorporates a target detection network to predict the detection results of fused images, utilizing semantic loss through backpropagation to guide the training of the fusion network, thereby ensuring that the fused images contain richer semantic information. To meet the real-time requirements of high-level vision tasks, a lightweight network based on Gradient Residual Dense Block (GRDB) is designed[6], which enhances detail extraction by leveraging gradient-based features.

To address the challenge of the inability to directly pretrain a target detection model for guiding the fusion network, the method adopts a joint adaptive training strategy inspired by generative adversarial networks. This approach effectively balances the performance of upstream fusion tasks and downstream target detection tasks, achieving both performance equilibrium and efficient fusion.

In the design of the loss function, the method defines the total loss as a weighted sum of image fusion loss and target detection loss. The image fusion loss comprises intensity loss and texture loss, where the intensity loss constrains the overall appearance intensity of the fused result, and the texture loss ensures that the fused image retains as much texture detail as possible.

### B. Object Detection with Fusion Images

In the target detection module, this project employs the Spiking-YOLO algorithm for object detection. Spiking-YOLO constructs an SNN-based object detection model by utilizing two key techniques: channel normalization and threshold imbalance in signed neurons[7]. This approach enables the conversion of a YOLO model from a deep neural network (DNN) to a spiking neural network (SNN), facilitating energy-efficient and biologically inspired object detection.

*1) SNN Neurons:* Unlike traditional neural networks, spiking neural networks (SNN) transmit information between neurons using spike trains composed of discrete spikes. Integrate-and-fire neurons accumulate input $z$ into the membrane potential $V_m$, as described below.

$$V^{l}_{mem,j}(t) = V^{l}_{mem,j}(t-1) + z^{l}_{j}(t) - V_{th}\Theta^{l}_{j}(t), \tag{1}$$

Here, $\theta_j^l(t)$ represents the spike, $z_j^l(t)$ denotes the input to the $j$-th neuron in the $l$-th layer, and $V_{th}$ is the threshold voltage. The input $z_j^l(t)$ can be expressed as follows:

$$z_j^l(t) = \sum_i \omega_{i,j}^l \Theta_i^{l-1}(t) + b_j^l \tag{2}$$

Here, $\omega$ and $b$ represent the weight and bias, respectively. When the membrane potential $V_{mem}$ exceeds the threshold voltage $V_{th}$, a spike $\theta_j^l$ is generated, as described below:

$$\Theta_i^l(t) = U\left(V_{mem,i}^l(t) - V_{th}\right) \tag{3}$$

Here, $U(x)$ denotes the unit step function, which outputs 1 if the condition is satisfied and 0 otherwise. As shown in the ANN-to-SNN module in Fig 1, during the standard ANN-to-SNN conversion, batch normalization layers are absorbed into convolutional layers, followed by weight normalization. The neuron's membrane potential accumulation directly controls spike firing, encoding floating-point values over multiple timesteps.

*2) Channel Normalization:* In SNN, generating spike trains based on input amplitude for lossless content transmission is critical. However, over-activation or under-activation of neurons within a fixed time frame can lead to content loss, which is influenced by the threshold voltage $V_{th}$. A higher $V_{th}$ requires prolonged voltage accumulation for spike generation, while a lower $V_{th}$ results in excessive spiking. To address this, Spiking-YOLO introduces channel-wise normalization, which normalizes weights across channel dimensions using the maximum activation value. This approach prevents excessively low activation values, maintaining a higher spiking rate and enabling neurons to transmit information accurately within shorter durations. The structure of channel-wise normalization is defined as follows:

$$\tilde{w}^l \frac{\lambda^{l-1}}{\lambda^l} \text{ and } \tilde{b}^l = \frac{b^l}{\lambda^l}. \tag{4}$$

Let $i$ and $j$ denote dimension indices, and $w^l$ represent the weights of layer $l$. Weights are normalized in each channel using the maximum activation value $\lambda_j^l$, precomputed from the training set. For non-input layers, normalized activations must be scaled by $\lambda_j^{l-1}$ to revert the inputs to their pre-normalized values from the preceding layer before applying the current layer's normalization. Failing to do so would progressively diminish the transmitted information.

*3) Signed neuron with unbalanced threshold:* Current DNN-to-SNN conversion methods primarily focus on mapping integrate-and-fire (IF) neurons to ReLU activation functions, often neglecting the negative values in activation functions. To address this limitation, Spiking-YOLO employs signed neurons with imbalanced thresholds, using distinct threshold voltages for the positive and negative regions. This enables neurons to transmit both positive and negative activation values effectively. The imbalanced-threshold neurons resolve the issue of information loss associated with the negative part of traditional ReLU activation functions, making them more adaptable to diverse data distributions. When the neuron's input exceeds the positive voltage threshold, it generates a positive spike; when the input falls below the negative voltage threshold, it generates a negative spike. This approach preserves the sign information of the input signal, enhancing the performance and energy efficiency of SNN.

As shown in Fig 1, the overall process of the multimodal remote sensing nighttime emergency rescue model is primarily divided into two steps: fusion and detection. The input data consists of low-light-infrared fused images generated using the SeAFusion method. During the training process, this method integrates high-level vision task training, resulting in fused images with high accuracy metrics. The fused images retain substantial information on the position and attributes of the target of interest, while minimizing background overexposure, making them suitable for the subsequent object detection stage. During the detection phase, an ANN-to-SNN converter is employed to transform the ANN model into a SNN, including the introduction of neurons and precise mapping of weights. The converted SNN model, leveraging event-driven computational characteristics, significantly reduces energy consumption while maintaining high detection accuracy. The entire process design fully demonstrates the advantages of multimodal fusion and efficient detection in complex nighttime environments.

## III. EXPRIMENTAL RESULT

The experiment was conducted primarily in a simulated nighttime rescue scenario. To validate the effectiveness of the employed fusion method, we compared SeAFusion with alternative approaches, including wavelet transform[8], DenseFuse, and RFN-Nest. Furthermore, to demonstrate the superiority of the detection algorithm in terms of power efficiency and accuracy retention, we compared Spiking-YOLO at different timesteps with Tiny-YOLO, analyzing various accuracy metrics and theoretical power consumption.

### *A. Dataset and experimental setting*

The experiment utilized low-light-infrared fusion imaging equipment to collect the dataset. The shooting scenarios were constructed based on practical requirements, simulating nighttime conditions with low visible light illuminance and various rescue scenarios involving multi-scale targets due to differences in scenes and target characteristics. Over 9,000 frames of low-light and infrared images were captured, and corresponding image registration was performed to ensure alignment and accuracy.

### *B. Image fusion results comparison*

The experimental results of image fusion are presented in Table I. The SeAFusion-based method outperforms the other three approaches in terms of metrics such as entropy (EN), spatial frequency (SF), mutual information (MI), the sum of the correlation of differences (SCD), and gradient-based fusion quality (Qabf), with RFN-Nest ranking second. Although the wavelet transform achieves faster fusion speeds, traditional fusion methods that extract features in the spatial or transform domain typically require manually designed transformation rules and fusion strategies. These algorithms are relatively simple compared to deep learning-based fusion methods and offer limited fusion depth. In contrast, deep learning-based fusion methods require a large number of well-annotated samples and a considerable amount of time for learning and training. However, with the improvement of learning networks, the integration of residual dense blocks and

densely connected layers, advancements in fusion rules, and the modification of loss functions, deep learning-based fusion algorithms have achieved significant improvements in multiple metrics, delivering superior fusion performance.

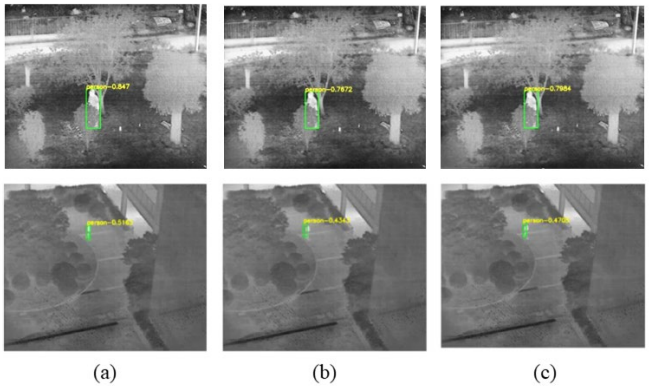
TABLE I. IMAGE FUSION INDEX COMPARISON

| Fusion Indicators | Wavelet transform | DenseFuse | RFN-Nest | SeAFusion |
|---|---|---|---|---|
| EN | 6.52 | 6.17 | 6.33 | 7.12 |
| SF | 0.032 | 0.019 | 0.035 | 0.047 |
| SD | 8.82 | 8.81 | 9.38 | 9.09 |
| MI | 2.31 | 2.45 | 2.51 | 2.69 |
| SCD | 1.51 | 1.62 | 1.62 | 1.47 |
| Qabf | 0.27 | 0.35 | 0.28 | 0.42 |

### *C. Comparison of target detection results*

After converting the trained Tiny-YOLO model into Spiking-YOLO using the DNN-to-SNN transformation, the performance metrics for target detection tasks at time steps 1000 and 2000 are presented in Table II and Fig. 2. The DNN-to-SNN conversion process incurs minimal accuracy loss, making it highly suitable for lightweight deployment on neuromorphic computing platforms. The converted Spiking-YOLO demonstrates fast detection speeds and low energy consumption, which are advantageous for real-time feedback and extended operational endurance in nighttime drone rescue missions. Moreover, as the time step increases, the performance of Spiking-YOLO improves across various metrics, and the accuracy loss from DNN-to-SNN conversion gradually diminishes. However, the increase in time steps also leads to longer processing times for target detection. Additionally, due to the model's limited parameter count, its detection accuracy for small objects is relatively low, and it is prone to false detections of anomalous small targets.

To investigate the energy efficiency of Spiking-YOLO, the experiment conducted a comparative analysis of AC (Additive Calculation) and MAC (Multiplicative Accumulation) operations, as well as energy consumption, between Spiking-YOLO and Tiny-YOLO through simulation in digital signal processing. DNN relies on energy-intensive multiply-accumulate (MAC) operations, while SNN uses binary spikes for low-energy accumulation (AC) operations (4.6 pJ vs. 0.9 pJ). The total energy consumption of Tiny-YOLO and Spiking-YOLO is computed by multiplying FLOPs with the respective energy per operation.



(a) (b) (c)

Fig. 2. Comparison of target detection results. (a) represents the detection result of Tiny-YOLO, (b) represents the detection result of Spiking-YOLO at 1000 steps, and (c) represents the detection result of Spiking-YOLO at 2000 steps.

The results indicate that regardless of using traditional layer-wise normalization or the proposed channel-wise normalization, the AC energy consumption of Spiking-YOLO is significantly lower than that of Tiny-YOLO, approximately 1/2000.

TABLE II. COMPARISON OF TARGET DETECTION INDICATORS

| Detection indicators | Tiny-YOLO | Spiking-YOLO (steps=1000) | Spiking-YOLO (steps=2000) |
|---|---|---|---|
| Accuracy | 0.979 | 0.946 | 0.971 |
| Recall | 0.956 | 0.921 | 0.948 |
| mAP 0.5 | 0.973 | 0.934 | 0.967 |
| mAP 0.5:0.95 | 0.806 | 0.767 | 0.794 |

## IV. CONCLUSION

In conclusion, this study demonstrates the potential of leveraging SNN for energy-efficient object detection in nighttime emergency rescue scenarios. By utilizing the fused characteristics of low-light and infrared imagery, combined with the computational advantages of SNN, the proposed Spiking-YOLO-based method achieves significant reductions in power consumption—approximately 1/2000 of the ANN-based Tiny-YOLO—while maintaining a negligible performance degradation of less than 2%. These results validate the effectiveness of the approach in addressing the challenges of deploying object detection algorithms on resource-constrained edge platforms, offering a promising solution for real-time, low-power applications in complex nighttime rescue environments.

## REFERENCES


[1] M. Ahmed et al., "Advancements in RIS-Assisted UAV for Empowering Multi-Access Edge Computing: A Survey," IEEE Internet of Things Journal, pp. 1–1, 2025, doi: 10.1109/JIOT.2025.3527041.

[2] Z. Zheng et al., "Logic combination and diagnostic rule-based method for consistency assessment and its application to cross-sensor calibrated nighttime light image products," Remote Sensing of Environment, vol. 318, p. 114598, Mar. 2025, doi: 10.1016/j.rse.2025.114598.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788. Accessed: Jan. 16, 2025.

[4] J. K. Eshraghian et al., "Training Spiking Neural Networks Using Lessons From Deep Learning," Proceedings of the IEEE, vol. 111, no. 9, pp. 1016–1054, Sep. 2023, doi: 10.1109/JPROC.2023.3308088.

[5] C. Li et al., "Low-Light Image and Video Enhancement Using Deep Learning: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 9396–9416, Dec. 2022, doi: 10.1109/TPAMI.2021.3126387.

[6] L. Tang, J. Yuan, and J. Ma, "Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network," Information Fusion, vol. 82, pp. 28–42, Jun. 2022, doi: 10.1016/j.inffus.2021.12.004.

[7] S. Kim, S. Park, B. Na, and S. Yoon, "Spiking-YOLO: Spiking Neural Network for Energy-Efficient Object Detection," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, Art. no. 07, Apr. 2020, doi: 10.1609/aaai.v34i07.6787.

[8] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques — An introduction, review and comparison," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, no. 4, pp. 249–263, Sep. 2007, doi: 10.1016/j.isprsjprs.2007.05.009.

[9] H. Li and X.-J. Wu, "DenseFuse: A Fusion Approach to Infrared and Visible Images," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2614–2623, May 2019, doi: 10.1109/TIP.2018.2887342.

[10] H. Li, X.-J. Wu, and J. Kittler, "RFN-Nest: An end-to-end residual fusion network for infrared and visible images," Information Fusion, vol. 73, pp. 72–86, Sep. 2021, doi: 10.1016/j.inffus.2021.02.023.