




Intelligent Home Scene Recognition Based on Image Processing and Internet of Things

Zhongpeng Liu^{1*}, Lijuan Liu², Lei An³

¹ Academic Affairs Office, Baoding University, Baoding 071000, China

² College of Information Science and Technology, Hebei Agricultural University, Baoding 071000, China

³ College of Artificial Intelligence, Baoding University, Baoding 071000, China

Corresponding Author Email: liuzhongpeng@bdu.edu.cn



<https://doi.org/10.18280/ts.400333>

ABSTRACT

Received: 15 February 2023

Accepted: 10 May 2023

Keywords:

image processing, internet of things technology, intelligent home, scene recognition

Intelligent home systems interconnect various devices within the home using Internet of Things (IoT) technology. In order to achieve the objectives of remote control, automated management, and intelligent services, these systems require robust scene recognition capabilities. However, the accuracy and real-time performance of current image processing algorithms in complex environments and diverse scenarios remain to be improved. Additionally, the interoperability and security issues among intelligent home devices are challenging to address. Therefore, this study delves into the scene recognition technology of intelligent homes based on image processing and IoT. A GLN network is constructed to process multi-view images of intelligent home scenes, enabling the determination of sub-region positions within the scenes. A model aggregation algorithm based on distributed learning is proposed, selecting intelligent home edge devices as the intelligent nodes of the IoT. By processing data and training models on these intelligent nodes, distributed intelligent home scene recognition is achieved. A dual-channel deep neural network-based intelligent home scene recognition model is constructed, and experimental results verify the effectiveness of the proposed model.

1. INTRODUCTION

With the rapid development of information technology, intelligent homes have gradually become an essential component of people's lives. Intelligent homes refer to the interconnection of various devices within the household via Internet of Things (IoT) technology, achieving remote control, automated management, and intelligent services, thereby enhancing the comfort, convenience, and security of home life [1-3]. To accomplish this goal, intelligent home systems necessitate robust scene recognition capabilities to automatically adjust the working status of devices based on the environment and user needs. Image processing technology, as an effective means of scene recognition, has received widespread research and application in recent years [4-7].

Image processing technology analyzes captured images, extracts useful information, and performs recognition, classification, and other processes to comprehend target scenes [8]. In the intelligent home domain, image processing technology can be applied to various scenarios, such as facial recognition, behavior analysis, and posture recognition [9-11]. As an emerging information technology, IoT connects various devices to networks, enabling real-time transmission and sharing of information, providing an essential foundation for intelligent home scene recognition [12-14].

In recent years, many researchers have dedicated themselves to applying image processing and IoT technology in intelligent home scene recognition. For instance, some studies have employed deep learning technology to develop image processing-based facial recognition systems, achieving automatic identification of family members and supporting

personalized intelligent home services [15-17]. On the other hand, other researchers have analyzed family members' activity trajectories and behavior patterns, designing intelligent surveillance systems capable of recognizing abnormal behaviors to enhance home security [18]. Moreover, image processing technology can be applied to intelligent home energy management. For example, by monitoring and analyzing indoor lighting conditions in real-time, systems can automatically adjust curtains and lighting to achieve more efficient energy utilization [19]. Simultaneously, IoT technology advancements have facilitated collaborative work among intelligent home devices, such as air conditioning, lighting, and security systems, which can share information and work together to create a more intelligent home environment [20, 21].

Despite the achievements of image processing and IoT technology in intelligent home scene recognition, current research still presents challenges and issues. For example, the recognition accuracy and real-time performance of current image processing algorithms in complex environments and diverse scenarios need improvement [22]. Additionally, the interoperability and security of intelligent home devices are pressing problems to be resolved [23-26].

To address these challenges, this study conducts in-depth research into intelligent home scene recognition technology based on image processing and IoT. Firstly, in section 2, a GLN network is constructed to process multi-view images of intelligent home scenes, enabling the determination of sub-region positions within these scenes. Secondly, in section 3, a model aggregation algorithm based on distributed learning is proposed, selecting intelligent home edge devices as the

intelligent nodes of the IoT. By processing data and training models on these intelligent nodes, distributed intelligent home scene recognition is achieved. Finally, in section 4, a dual-channel deep neural network-based intelligent home scene recognition model is constructed, consisting of a sequence generation network, a global self-attention encoding network, a feature fusion network, and a classification network. Experimental results verify the effectiveness of the proposed model.

2. INTELLIGENT HOME SCENE REGION LOCALIZATION

Images of IoT-based intelligent home scenes from different perspectives at the same location contain rich background information about the position. To utilize this information, a coarse sub-region division of the intelligent home scene areas is performed in this study. Subsequently, a GLN network is constructed to process multi-view images of intelligent home scenes, achieving the determination of sub-region positions within these scenes. Figure 1 presents the network framework.

The problem of determining sub-region positions in multi-view intelligent home scenes is described as follows. It is assumed that the images of the front, back, left, and right directions of the sub-region position to be located are represented by $z = \{z_1, z_2, z_3, z_4\}$, $z \in Z$. The set of all sub-region position images is denoted by Z . The predicted position of z is represented by $t \in T$, and the set of predicted positions of all J sub-regions to be located is represented by $T = \{t_1, t_2, t_3, t_4\}$. Therefore, the goal of solving this problem is to learn the mapping function $d: Z \rightarrow T$, realizing the mapping of the image z to the position target class t .

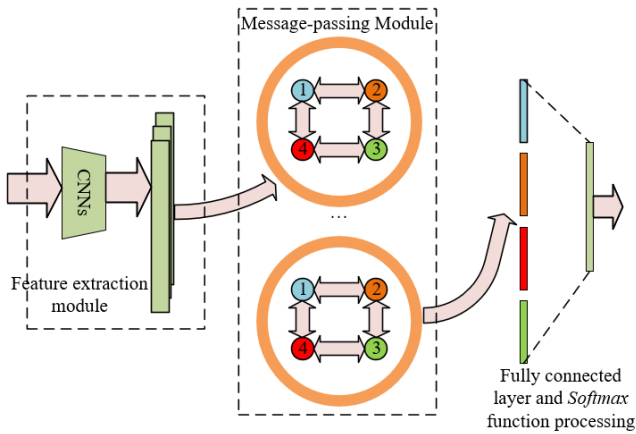


Figure 1. Network framework

Given a set of $z = \{z_1, z_2, z_3, z_4\}$, z is used as the input to the model to extract high-dimensional features of intelligent home scene images, $e = \rho(z) = \{e_1, e_2, e_3, e_4\}$.

In the localization model, a message-passing module is established. It is assumed that the undirected edge between nodes c_u and c_k is represented by r_{ij} , and quadrilateral graph $H = (C, R)$ is defined using $C = \{c_1, c_2, c_3, c_4\}$ and $R = \{r_{12}, r_{23}, r_{34}, r_{41}\}$. The hidden states of node u at layers m and $m+1$ are represented by $g_u^m \in E^D$ and $g_u^{m+1} \in E^D$, respectively. The set of neighboring nodes of node c_u is denoted by $B(u)$, the non-linear activation function is represented by δ , the normalization constant is represented by $s_{uk} = (|B(u)B(k)|)^{1/2}$, the weight matrix of the feature transformation at layer m is

denoted by Q^m , and the update formula for the hidden state g_u of node c_u is given as follows:

$$g_u^m = \begin{cases} e; m = 1 \\ \delta \left(\sum_{k \in B(u)} \frac{1}{s_{uk}} Q^{m-1} g_k^{m-1} \right) \end{cases} \quad (1)$$

It is assumed that the shared attention mechanism, s , is used for calculating the importance of features of node k for u . The weight update formula is as follows:

$$s_{uk}^m = \frac{\exp(\delta(s[Q^m G_u^m, Q^m g_k^m]))}{\sum_{j \in B(u)} \exp(\delta(s[Q^m g_u^m, Q^m g_j^m]))} \quad (2)$$

In the localization model, a position prediction module is established. The hidden state g undergoes M -layer information passing, eventually forming a single robust feature represented by z^e , as follows:

$$z^e = [g_1^M, g_2^M, g_3^M, g_4^M] \quad (3)$$

Finally, z^e serves as the input for the fully connected layer Ψ , and the probability distribution $o' = d(\Gamma(\Psi(z^e)))$ of all position target classes is output after processing by the *Softmax* function $\Gamma(\cdot)$. The real position label of position u is represented by o_u , and the objective function is set according to the *Softmax* loss as follows:

$$M = -\sum_u o_u \log(o'_u) \quad (4)$$

3. SCENE RECOGNITION DISTRIBUTED LEARNING ALGORITHM

In IoT-based intelligent homes, the main reason for low scene recognition efficiency is that cloud-based devices serve as the primary means of data processing and model training. This paper proposes a model aggregation algorithm based on distributed learning, which selects intelligent edge devices in the IoT network as intelligent nodes. By utilizing these intelligent nodes for data processing and model training, distributed intelligent home scene recognition is achieved.

First, a global model is initialized on the intelligent home cloud server. The parameters of the global model are then broadcast to all edge devices participating in distributed learning. Each edge device trains the model locally using the data it has collected and updates the model weights. Once the training is completed, each edge device uploads its local model weights to the cloud server. Figure 2 presents the main functions of IoT devices in intelligent homes.

Assuming that edge device models are assigned weights based on the number of edge devices, the weights are represented by $\{q_{y,n}^b\}_{b=1}^B$. Let the bias first moment estimate from the n th batch of the b th edge device at time y be $l_{y,n}^b$, and the corresponding second moment estimate be $c_{y,n}^b$. The bias-corrected first moment estimate is denoted by $\hat{l}_{y,n}^b$, and the corresponding second moment estimate is denoted by $\hat{c}_{y,n}^b$. The model weight gradient of the b th edge device in the n th batch at time y is represented by $h_{y,n}^b$. Decay factors are denoted by α_1 and α_2 , and their corresponding decay factors are

represented by $\alpha^b_{1,y,n}$ and $\alpha^b_{2,y,n}$, respectively. The learning rate at time y is denoted by λ^y , and the constant to prevent zero denominators is represented by γ . The Adam algorithm expressions for updating local model weights are as follows:

$$l^b_{y,n+1} = \alpha_1 l^b_{y,n} + (1 - \alpha_1) \cdot h^b_{y,n} \quad (5)$$

$$c^b_{y,n+1} = \alpha_2 c^b_{y,n} + (1 - \alpha_2) (h^b_{y,n})^2 \quad (6)$$

$$\hat{l}^b_{y,n+1} = \frac{l^b_{y,n+1}}{1 - \alpha^b_{1,y,n}} \quad (7)$$

$$\hat{c}^b_{y,n+1} = \frac{c^b_{y,n+1}}{1 - \alpha^b_{2,y,n}} \quad (8)$$

$$\mu^b_{y,n+1} = \mu^b_{y,n} - \lambda_y \frac{\hat{l}^b_{y,n+1}}{\sqrt{\hat{c}^b_{y,n+1} + \gamma}} \quad (9)$$

There is no direct connection between the edge devices

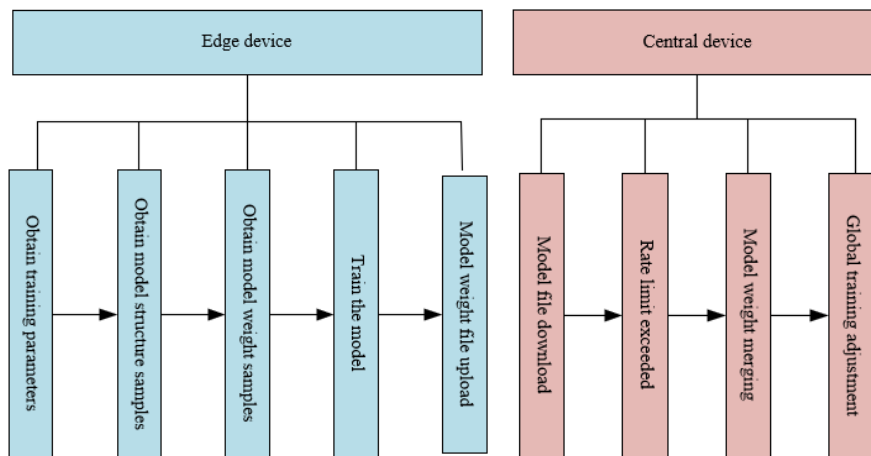


Figure 2. Main functions of IoT devices in intelligent homes

4. INTELLIGENT HOME SCENE RECOGNITION MODEL CONSTRUCTION

This study constructs an intelligent home scene recognition model based on a dual-channel deep neural network, comprising a sequence generation network, a global self-attention encoding network, a feature fusion network, and a classification network. The sequence generation network serves as the basis of the dual-channel structure and effectively extracts local features from the input image data. These local features aid in capturing the objects within the scene and their relationships, providing valuable information for the subsequent global self-attention encoding network. The global self-attention encoding network employs a self-attention mechanism to capture long-range dependencies within the image on a global scale. This encoding approach helps the model understand the relationships between different regions in the scene, further enhancing scene recognition accuracy. The feature fusion network is responsible for fusing the features generated by the two networks in the dual-channel

throughout the entire model training process. During IoT communication, all model weights $\{q^b_{y,n}\}_{b=1}^B$ are uploaded to the central IoT device.

The cloud server collects the model weights uploaded by all edge devices. A model aggregation algorithm is applied to merge these weights into a single set of global model weights. As the core of the decentralized system, this paper employs model averaging as the aggregation algorithm to summarize and average all local model parameters. Let the global model weights at time y be represented by Q_y , with the algorithm expression as follows:

$$W_{t+1} = \frac{1}{N} \sum_{i=1}^N w_{t,B}^n \quad (10)$$

Finally, the updated global model weights are broadcast to all edge devices. The edge devices update their local models based on the received global model weights. The updated models are then tested using local data to verify the recognition performance in actual scenarios. If the model performance is unsatisfactory, multiple training and updating rounds can be performed until the desired performance is achieved.

structure. This design effectively leverages the strengths of both networks, combining local and global information to generate more expressive feature representations. The classification network is responsible for classifying the fused features and outputting the final scene recognition results. Figure 3 presents a schematic diagram of the model structure.

This structure effectively combines local feature extraction and global attention mechanisms, allowing the model to better understand and recognize intelligent home scenes. Simultaneously, in the context of the Internet of Things, the model can process vast amounts of data from various devices, adapt to different scenarios, and provide a robust solution for intelligent home scene recognition.

In the constructed intelligent home scene recognition model, the input for the sequence generation network is a 2D intelligent home scene image with a width and height of G and Q , respectively, and a channel number of V , represented by $Z \in E^{G \times Q \times V}$. These images need to be transformed into 1D sequences, first dividing the image into $O \times O$ sub-blocks $Z_o \in E^{B \times (o^2 \cdot V)}$, with the input sequence length represented by B .

To retain the positional information of the sub-blocks, the 2D matrix is fused with the position encoding RBM, generating 2D matrices Z_{RG} and Z_{DE} . Using Z_{RG} as an example, the corresponding expression for the above steps is given by the following equation:

$$X_{RG} = [Z_{CL}; Z_o^1 R; Z_o^2 R; \dots; Z_o^B R] + R_{PO} \quad (11)$$

In the global self-attention encoding network, the concepts of *Query*, *Key*, and *Value* for the attention mechanism are established. For B *Query* sequences, the attention output can be calculated using the following equation:

$$ATT(Q, K, V) = \Omega \left(\frac{QK^Y}{\sqrt{f_w}} \right) V \quad (12)$$

Figure 4 presents a schematic diagram of the global attention encoding module. If *Query*, *Key*, and *Value* are all obtained from a sequence $J \in E^{B \times F}$ containing b vectors through linear transformation, the input sequence is processed using g self-attention mechanisms, ultimately dividing the sequence into g sequences of size $B \times f$, satisfying $F = gf$. The final data is a feature matrix generated by concatenating the outputs of the g self-attention mechanisms, which constitutes the final output. Assuming there are L neurons in a given layer with input

$\{x^m_1, x^m_2, \dots, x^m_L\}$, the following equations apply:

$$\omega = \frac{1}{L} \sum_{l=1}^L x_l^m \quad (13)$$

$$\delta^2 = \frac{1}{L} \sum_{l=1}^L (x_l^m - \omega)^2 \quad (14)$$

In the feature fusion network, a learnable category vector is added to the generated sequence 2D matrix to help the model recognize different scene elements. Moreover, position encoding is introduced to facilitate the model's understanding of the relative positions of elements within the scene, resulting in X_{RG} and X_{DE} . Subsequently, X_{RG} and X_{DE} are fed into the global self-attention encoding network, which allows the model to capture long-range dependencies on a global scale, yielding updated X_{RG} and X_{DE} . The feature fusion network combines the two, generating matrix D . Before being input into the classification network, the true fused features must be extracted from D , which contains learnable category vector information, by taking the first data of each dimension in D to generate the fused feature D_0 .

The classification network takes the fused feature D_0 as input, with the structure consisting of a simple fully connected layer.

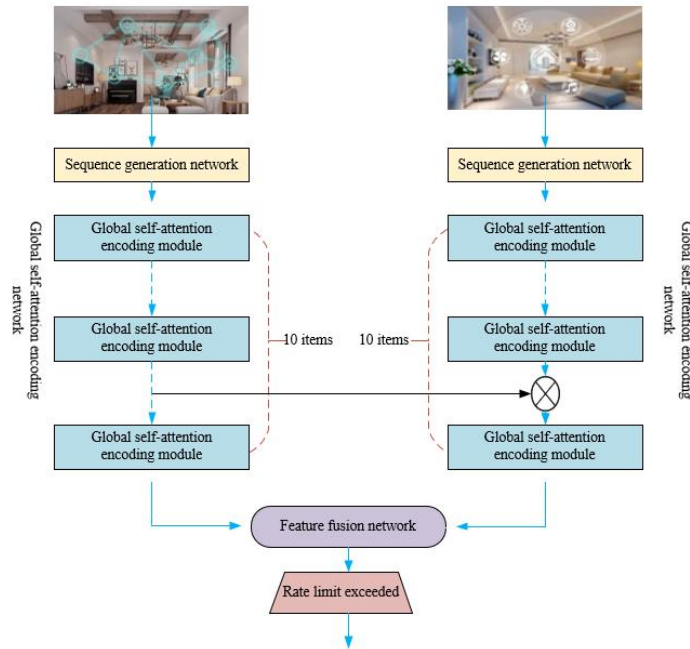


Figure 3. Model structure schematic diagram

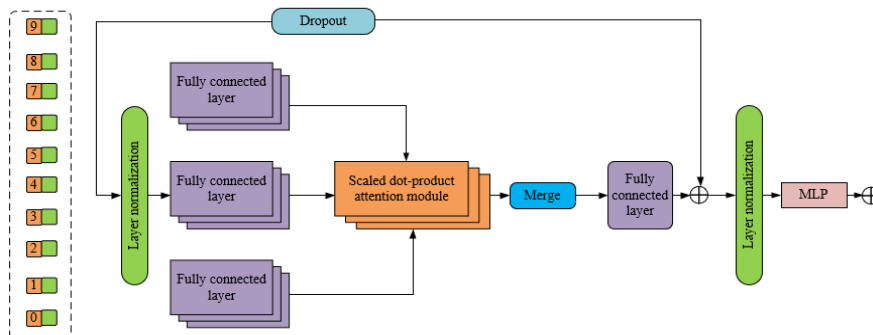


Figure 4. Global attention encoding module schematic diagram

5. EXPERIMENTAL RESULTS AND ANALYSIS

Based on the experimental results from Figure 5, an analysis of the *Top-n* localization prediction accuracy for smart home scene recognition, founded on image processing and the Internet of Things (IoT), is offered. As observed in the figure, under *Top-1* conditions, the probability of the model's highest scoring scene matching the actual scene is 0.78, thus indicating a 78% prediction accuracy. In a *Top-2* scenario, the highest scoring two scenes predicted by the model reveal a probability of 0.91 for the accurate scene to appear, which translates to a 91% prediction accuracy. Under *Top-3* conditions, the three highest scoring scenes predicted by the model have a 0.93 probability of the correct scene appearing, equating to a 93% prediction accuracy. Under *Top-5* conditions, the five highest scoring scenes predicted by the model show a 0.96 probability of the correct scene appearing, corresponding to a 96% prediction accuracy. Finally, in a *Top-10* scenario, the ten highest scoring scenes predicted by the model demonstrate a 1.0 probability of the correct scene appearing, marking a 100% prediction accuracy. From the aforementioned data, it is evident that as the value of *Top-n* increases, the scene recognition prediction accuracy also increases correspondingly. When considering *Top-10*, the accuracy reaches 100%, but in practical application, a balance between accuracy and computational resources must be maintained. In most cases, the prediction accuracy of *Top-3* or *Top-5* is already quite high (93% and 96% respectively), enough to meet the demands of smart home scene recognition. Therefore, an appropriate *Top-n* value can be chosen based on the real-world application scenario and resource constraints, to achieve a higher prediction accuracy.

From Figure 6, the Cumulative Distribution Function (*CDF*) of the *Top-1* scene localization error distribution for smart home scene recognition based on image processing and IoT can be analyzed. As illustrated in the figure, when the localization error is 1, the *CDF* is 0.941, which means there is a 94.1% probability that the prediction error will be less than or equal to 1. When the localization error is 2, the *CDF* is 0.972, which suggests a 97.2% probability that the prediction error will be less than or equal to 2. When the localization error is 3, the *CDF* is 0.978, signifying a 97.8% probability that the prediction error will be less than or equal to 3. When the localization error is 4, the *CDF* is 0.988, representing a 98.8% probability that the prediction error will be less than or equal to 4. When the localization error is 5, the *CDF* is 1, indicating a 100% probability that the prediction error will be less than or equal to 5. When the localization error is 6, the *CDF* is 0.999, denoting a 99.9% probability that the prediction error will be less than or equal to 6. From the aforementioned data, it is clear that the *Top-1* scene localization error distribution of the smart home scene recognition, based on image processing and IoT, performs well. When the localization error is 1, there is already a 94.1% probability that the prediction error will be less than or equal to 1. As the localization error increases, the *CDF* value also increases correspondingly. When the localization error is 5, the prediction accuracy reaches 100%. This indicates that the model proposed in this study demonstrates high accuracy and stability in the task of intelligent home scene recognition. In practical applications, an appropriate localization error threshold can be selected according to the requirements and error tolerance. This indicates that the model proposed in this study demonstrates high accuracy and stability in the task of intelligent home scene recognition. In

practical applications, an appropriate localization error threshold can be selected according to the requirements and error tolerance.

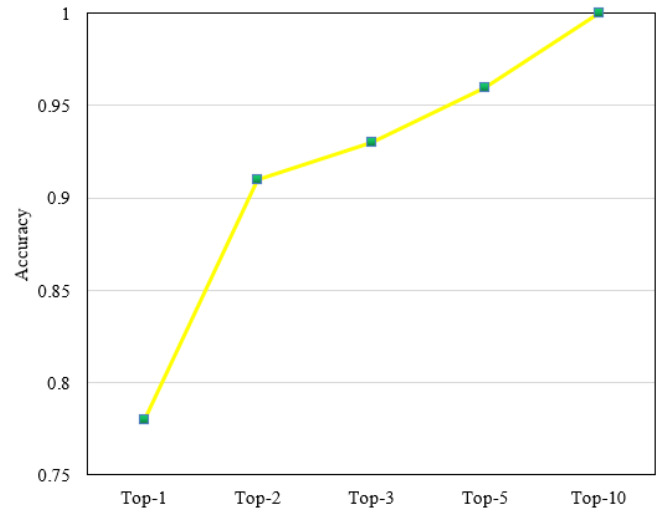


Figure 5. Top-n localization prediction accuracy

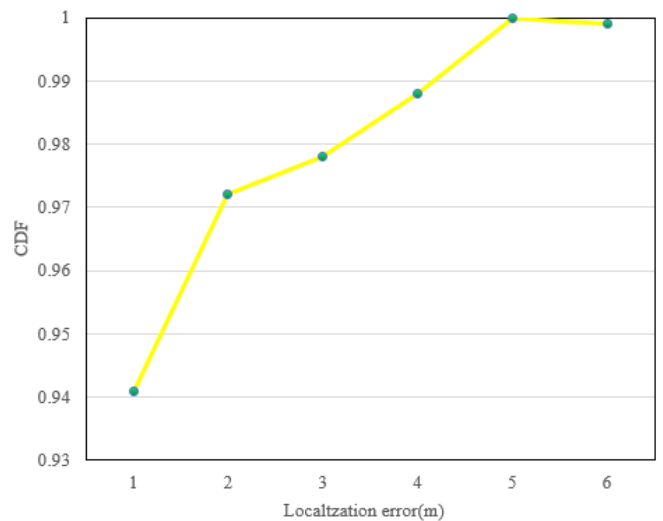


Figure 6. CDF Curve of *Top-1* scene localization error distribution

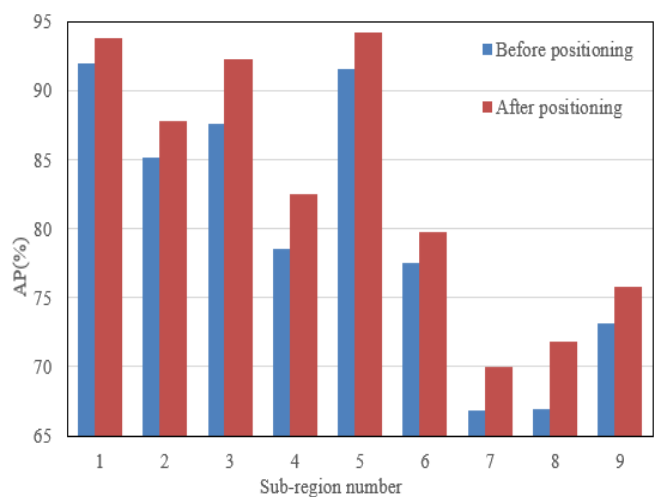


Figure 7. Detection results in various scene sub-area categories before and after introducing the localization model

Based on Figure 7, an analysis of the detection results in various scene sub-area categories for intelligent home scene recognition based on image processing and the Internet of Things can be conducted. The figure shows that the scene recognition accuracy in all sub-areas improves after introducing the localization model. The accuracy in sub-area 1 increases from 92% to 93.8%, in sub-area 2 from 85.2% to 87.8%, in sub-area 3 from 87.6% to 92.3%, in sub-area 4 from 78.5% to 82.5%, in sub-area 5 from 91.6% to 94.2%, in sub-area 6 from 77.5% to 79.8%, in sub-area 7 from 66.8% to 70%, in sub-area 8 from 66.9% to 71.8%, and in sub-area 9 from 73.1% to 75.8%. It can be observed that the detection results in all scene sub-area categories improve after introducing the localization model, indicating that the localization model has a positive impact on overall scene recognition accuracy.

Figure 8 presents the loss convergence curves under different numbers of global self-attention encoding network modules. The figure shows that without using global self-attention encoding network modules, the loss value gradually decreases with the increase in the number of iterations, but remains slightly above 0.34 when the number of iterations approaches 100. With 5 global self-attention encoding network modules, the loss value also decreases with the increase in the number of iterations, eventually reaching slightly below 0.24. With 10 global self-attention encoding network modules, the loss value also decreases with the increase in the number of iterations, eventually reaching slightly below 0.14. It can be observed that using more global self-attention encoding network modules (10) achieves a lower loss value. This indicates that increasing the number of global self-attention encoding network modules helps improve model performance.

Based on Table 1, the performance comparison of intelligent home scene recognition based on image processing and the Internet of Things under different detection models can be analyzed. The table shows that the *FV-CNN* model has a weight size of 321 MB, an *mAP* of 78.9%, and a processing speed of 22.2 frames/second. The *MOP-CNN* model has a weight size of 111 MB, an *mAP* of 79.8%, and a processing speed of 41.1 frames/second. The *MFAFVNet* model has a weight size of 109 MB, an *mAP* of 74.5%, and a processing speed of 61.2 frames/second. The proposed model has a weight size of 67 MB, an *mAP* of 82.3%, and a processing speed of 71.4 frames/second. It can be observed that the proposed model outperforms the other models in terms of weight size, *mAP*, and processing speed. Compared to other models, the proposed model has a smaller weight size, higher scene recognition accuracy, and faster processing speed. This demonstrates that the proposed model exhibits better performance in the task of intelligent home scene recognition. In practical applications, the proposed model can be considered to improve the performance of intelligent home scene recognition.

Table 1. Performance comparison of different detection models

Method	Weight Size (MB)	mAP(%)	Speed/(frame \cdot s $^{-1}$)
<i>FV-CNN</i>	321	78.9	22.2
<i>MOP-CNN</i>	111	79.8	41.1
<i>MFAFVNet</i>	109	74.5	61.2
<i>Our model</i>	67	82.3	71.4

Based on Figure 9, an analysis of the performance of intelligent home scene recognition using image processing and the Internet of Things under different detection models can be conducted. As shown in the figure, for the *FV-CNN* model, the performance gradually improves with the increase in the number of iterations, but it stabilizes after 40 iterations, reaching a maximum of 98.4%. For the *MOP-CNN* model, the performance also gradually improves with the increase in the number of iterations, but it stabilizes after 40 iterations, reaching a maximum of 99.2%. For the *MFAFVNet* model, the performance gradually improves with the increase in the number of iterations, but it stabilizes after 40 iterations, reaching a maximum of 99.0%. For the proposed model, the performance gradually improves with the increase in the number of iterations, but it stabilizes after 40 iterations, reaching a maximum of 99.4%. From the aforementioned data, it can be observed that the proposed model outperforms the other models in terms of performance. With the increase in the number of iterations, the performance of the proposed model gradually improves, eventually reaching 99.4%, which is higher than the other three models. This demonstrates that the proposed model exhibits better performance in the task of intelligent home scene recognition.

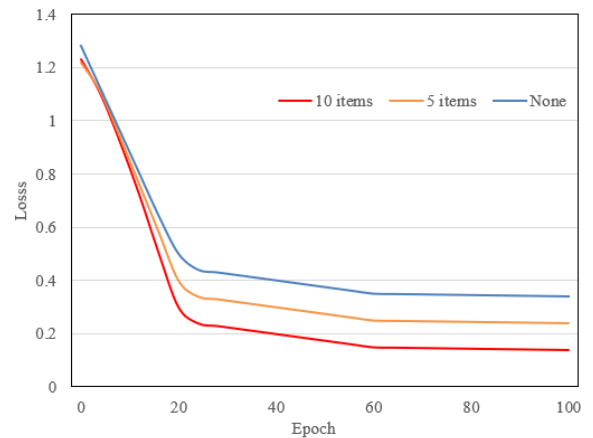


Figure 8. Loss convergence curves under different numbers of global self-attention encoding network modules

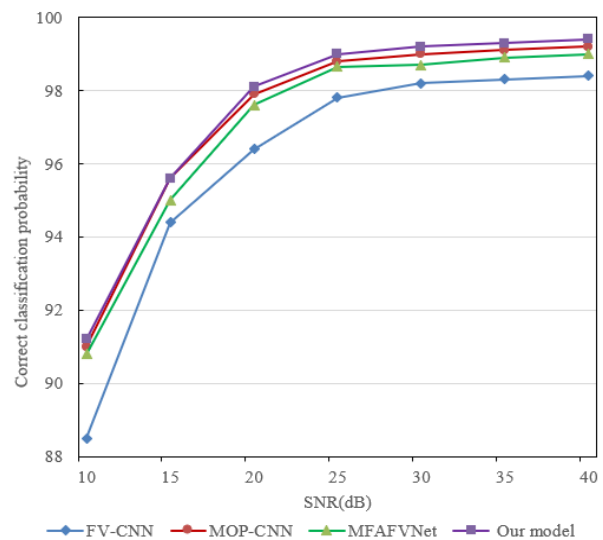


Figure 9. Performance curves of different models

6. CONCLUSION

In this study, the techniques of intelligent home scene recognition based on image processing and the Internet of Things were thoroughly investigated. Firstly, a GLN network was constructed to process multi-view images of intelligent home scenes, achieving the determination of sub-area locations in intelligent home scenes. Secondly, a model aggregation algorithm based on distributed learning was proposed, selecting intelligent home edge devices as intelligent nodes of the Internet of Things. By processing data and training models on these intelligent nodes, distributed intelligent home scene recognition was achieved. Lastly, a dual-channel deep neural network model for intelligent home scene recognition was developed. Through comprehensive experimental analysis, the following conclusions can be drawn:

(1) In the task of intelligent home scene recognition based on image processing and the Internet of Things, the proposed model outperforms the comparison models (FV-CNN, MOP-CNN, and MFAFVNet) in terms of weight size, mAP, processing speed, and performance curves. This demonstrates that the proposed model exhibits better performance in the scene recognition task and is suitable for practical applications.

(2) When analyzing the impact of the number of global self-attention encoding network modules on the loss convergence curves, it was found that using more global self-attention encoding network modules (10) achieves a lower loss value. This indicates that increasing the number of global self-attention encoding network modules helps improve model performance. However, in practical applications, the relationship between model complexity and performance should be balanced, and an appropriate number of global self-attention encoding network modules should be chosen.

(3) When analyzing the performance curves of different models, it was observed that the performance of each model improves gradually with the increase in the number of iterations. When the number of iterations reaches around 40, the model performance tends to stabilize. Therefore, when training the model, it can be considered to stop training at this point to save computing resources.

In summary, the proposed model demonstrates good performance in the task of intelligent home scene recognition and is worth considering for practical applications. Moreover, to improve model performance, optimization methods such as adjusting the number of global self-attention encoding network modules and optimizing the number of iterations can be employed.

ACKNOWLEDGMENT

This paper was supported by Scientific Research Program of Colleges and Universities of Hebei Province (Grant No.: ZC2022104).

REFERENCES

- [1] Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7): 1645-1660. <https://doi.org/10.1016/j.future.2013.01.010>
- [2] Atzori, L., Iera, A., Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15): 2787-2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- [3] Li, S., Xu, L.D., Zhao, S. (2015). The Internet of Things: a survey. *Information Systems Frontiers*, 17: 243-259. <https://doi.org/10.1007/s10796-014-9492-7>
- [4] Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.
- [5] Gonzalez, R.C., Woods, R.E. (2007). *Digital image processing (3rd Edition)*. Pearson Education.
- [6] Xu, X., Xue, Y., Qi, L., Yuan, Y., Zhang, X., Umer, T., Wan, S. (2019). An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles. *Future Generation Computer Systems*, 96: 89-100. <https://doi.org/10.1016/j.future.2019.01.012>
- [7] Lopez, G., Quesada, L., Guerrero, L.A., Evers, L. (2020). A survey on deep learning techniques for image and video frame interpolation. *arXiv preprint arXiv:2008.08884*.
- [8] Haralick, R.M., Shapiro, L.G. (1992). *Computer and robot vision*. Reading: Addison-wesley, 1: 28-48.
- [9] Zhang, Z. (2016). Microsoft kinect sensor and its effect. *IEEE Multimedia*, 19(2): 4-10. <https://doi.org/10.1109/MMUL.2012.24>
- [10] Wang, L., Gu, T., Tao, X., Lu, J. (2010). Sensor-based human activity recognition in a multi-user scenario. *European Conference on Ambient Intelligence*, 78-87. <https://doi.org/10.1007/978-3-642-05408-2>
- [11] Thevenot, J., López, M.B., Hadid, A. (2017). A survey on computer vision for assistive medical diagnosis from faces. *IEEE Journal of Biomedical and Health Informatics*, 22(5): 1497-1511. <https://doi.org/10.1109/JBHI.2017.2754861>
- [12] Ashton, K. (2009). That 'Internet of Things' thing. *RFID Journal*, 22(7): 97-114.
- [13] Atzori, L., Iera, A., Morabito, G. (2017). Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm. *Ad Hoc Networks*, 56: 122-140. <https://doi.org/10.1016/j.adhoc.2016.12.004>
- [14] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials*, 17(4): 2347-2376. <https://doi.org/10.1109/COMST.2015.2444095>
- [15] Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708.
- [16] Schroff, F., Kalenichenko, D., Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815-823.
- [17] Wang, J., Wang, Z., Li, J., Wu, J. (2018). Multilevel wavelet decomposition network for interpretable time series analysis. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2437-2446. <https://doi.org/10.1145/3219819.3220060>
- [18] Chen, C., Jafari, R., Kehtarnavaz, N. (2015). A real-time human action recognition system using depth and inertial sensor fusion. *IEEE Sensors Journal*, 16(3): 773-781. <https://doi.org/10.1109/JSEN.2015.2487358>
- [19] Chen, W., Chen, P., Chen, J. (2013). A smart home

- energy management system using IoT and smartphone-based remote monitoring. 2013 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 36-37.
- [20] Peña-López, I., López, D. (2018). Coordination in the Internet of Things: A novel approach based on semantic information. *Ad Hoc Networks*, 79: 37-51.
- [21] Liu, Y., Wang, L., Yan, B. (2021). A survey on IoT technologies: System architecture, software and hardware, security, and application. *Journal of Network and Computer Applications*, 169: 102883.
- [22] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386>
- [23] Stankovic, J.A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1): 3-9. <https://doi.org/10.1109/JIOT.2014.2312291>
- [24] Görmüş, S., Aydın, H., Ulutaş, G. (2018). Security for the internet of things: A survey of existing mechanisms, protocols and open research issues. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 33(4): 1247-1272. <https://doi.org/10.17341/gazimmfd.416406>
- [25] Sicari, S., Rizzardi, A., Grieco, L.A., Coen-Porisini, A. (2015). Security, privacy and trust in Internet of things: The road ahead. *Computer Networks*, 76: 146-164. <https://doi.org/10.1016/j.comnet.2014.11.008>
- [26] Roman, R., Zhou, J., Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, 57(10): 2266-2279. <https://doi.org/10.1016/j.comnet.2012.12.018>