WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945
DOI: 10.24425/ace.2024.151910
ARCHIVES OF CIVIL ENGINEERING
Vol. LXX ISSUE 4 2024
pp. 563–578

Research paper

Energy consumption reduction method for green buildings based on human thermal discomfort posture recognition algorithm

Shijiu Song[1], Li Zhou[2]

Abstract: The energy consumption of air conditioning systems accounts for more than 50% of building energy consumption. The supercooled and overheated environment provided by intelligent buildings can bring a large amount of energy loss. How to create comfortable spaces with energy-saving goals is currently the focus of research. The aim of this study is to improve the accuracy of human thermal discomfort pose recognition algorithms. This study first extracts human key points on the ground of bone key points, then normalizes the data, and finally constructs a human thermal uncomfortable posture recognition algorithm on the ground of deep learning technology. The experiment showcases that in the training set, when the iteration number is 1500, the accuracy reaches its maximum value, which is 99.98%. In the test set, the accuracy reached its maximum value of 89.85% when the iteration number was 400. After classifying the dataset, the accuracy of the first type dataset reached 99.51%. The accuracy rate of the second type dataset is 98.56%, and the accuracy rate of the third type dataset is 98.95%. In the comparison of the four algorithms, the accuracy of the research algorithm is significantly higher than the other three algorithms, indicating that the research algorithm can accurately recognize the thermal uncomfortable posture of the human body. This research algorithm can timely and effectively identify the uncomfortable posture of the human body, thereby automatically adjusting indoor temperature and achieving energy conservation and emission reduction.

Keywords: air conditioning system, architecture, attitude recognition algorithm, energy consumption, human thermal discomfort

[1]Ph.D., Shandong Jianzhu University, School of Management Engineering, 250014 Jinan, China, e-mail: bj-tusaad2020@163.com, ORCID: 0000-0002-8928-1851

[2]A.P., Jinan Vocational College, School of Finance and Economics, 250103 Jinan, China, e-mail: mail2, ORCID: 0009-0002-2873-4885

1. Introduction

With the improvement of people's living standards, building energy consumption has increasingly become a focus of attention. To reduce building energy consumption, various energy-saving technologies are constantly emerging, among which green building technology has received widespread attention due to its dual advantages of environmental protection and energy conservation [1]. However, the thermal comfort of people indoors directly affects the usage of building energy consumption [2]. Therefore, how to reduce the energy consumption of green buildings through human thermal uncomfortable posture recognition algorithms has become a current research hotspot. In deep learning algorithms, fully connected networks are a common deep learning model composed of multiple fully connected layers. The fully connected network has a simple structure, is easy to train and debug, and is suitable for various regression and classification problems [3–5]. In view of this, this study innovatively constructed three pose recognition algorithms after extracting key skeletal points: fully connected network, 1D convolution, and 1D convolution+LSTM. The research aims to achieve non-contact human thermal discomfort detection based on visual perception, solving the problem of low accuracy of existing visual detection methods due to personality differences. Research on human thermal discomfort from the perspective of human posture detection, based on the recognition algorithm of human thermal discomfort posture, to reduce the energy consumption of green buildings. Providing new research directions, as an emerging interdisciplinary research direction, human thermal discomfort pose recognition algorithms provide new research directions and ideas for fields such as green buildings and smart homes. The research content is mainly divided into four parts. The first part is a brief introduction to the research related to pose recognition, bone key point extraction algorithms, and green building energy consumption reduction methods. In the second part, the extraction of bone key points was first introduced, followed by information normalization processing methods. Finally, to achieve recognition of human thermal uncomfortable postures, deep learning algorithms were introduced. The third part is to validate the performance of the research and construction of human thermal uncomfortable posture recognition algorithms, and conduct performance testing and comparative analysis experiments. The fourth part is a summary and outlook of the research content.

2. Related works

In recent years, human pose recognition technology has been increasingly widely applied and researched. Zhu et al. studied a human activity recognition method on the ground of wearable sensor motion information fusion. The experimental results indicate that the system can accurately recognize the daily activities of the human body [6]. Zhu et al. proposed a method for determining target attitude through purely geometric methods on the ground of the data fusion principle of monocular cameras and LiDAR sensors. Compared with classic attitude tracking methods, the new relative navigation framework can more accurately track the target's attitude under synchronous geometric recognition [7]. Wang et al. discussed a human pose recognition method on the ground of affine invariant geometric feature sift operator and

proposed an improved sift algorithm. The experimental results show that this method can significantly improve the computational speed and attitude recognition efficiency of the sift operator [8]. Human skeleton extraction is a fundamental issue in the field of computer vision. The Gao team has proposed a simple and effective self supervised spatiotemporal contrastive learning method for action recognition on the ground of three-dimensional skeletons. To obtain an effective representation, the self supervised spatiotemporal contrastive learning method first designs some new contrastive agent tasks to provide different spatiotemporal observation scenes for the same 3D action, and aggregates them in the embedded space. The results show that the research method has achieved significant improvements in smaller model sizes and higher training efficiency [9]. Quan et al. studied gesture recognition methods in motion scenes using deep learning skeleton sequence models for human object detection, and trained the VGG16 network using motion video frame images. Then adjust the hyperparameters and network structure of the basic network on the ground of the training results, and obtain the key poses in the motion scene through the final classifier [10]. Pan et al. proposed a real-time driver behavior recognition architecture to reduce the occurrence of traffic accidents and ensure driving safety. It collects raw experimental videos through an onboard monocular camera for extracting driver upper body bone information. Then, it constructs a graph convolutional network (GCN) for single frame spatial structure feature inference, and then constructs a long and short term memory network for temporal motion feature learning within the sequence. The experiment shows that the recall rate and recognition efficiency of the research method meet the real-time requirements of engineering applications [11].

Reducing building energy consumption and environmental pollution has become an important goal of green buildings. Wen has designed a green energy-saving building energy consumption monitoring model on the ground of fuzzy neural networks to overcome the problems of large monitoring data errors and poor energy consumption control effectiveness in traditional models. The experimental results show that the monitoring error of this model is reduced, the energy consumption control effect is good, and the building energy-saving effect is high [12]. Wang et al. conducted simulation calculations on the energy consumption, energy-saving potential, and energy consumption structure of green industrial building projects. Then, it compares the simulated data of the running project with the measured data and conducts a comprehensive and systematic evaluation of it. The results indicate that the application of energy-saving technology is related to site selection, energy-saving goals, and energy-saving efficiency. The higher the specific system energy consumption ratio, the greater the energy saving potential of the system [13]. Wang et al. used Autodesk Ecotect analysis software to simulate the lighting analysis, shading performance, and annual energy consumption of electrochromic windows in specific geographical environments to verify their energy-saving performance. The results indicate that the shading performance of the target electrochromic window is significantly better than that of ordinary windows [14]. Xu analyzed the significance level of green building energy consumption using orthogonal experimental method and building information simulation software to simulate the energy consumption of green buildings, in response to the problems of large energy consumption control errors and low control efficiency in traditional energy consumption control methods. The experimental results show that the energy consumption control method has lower control error and higher control efficiency [15].

In summary, the research on pose recognition algorithms and bone key point extraction is relatively mature and has been applied in various fields, while there is little research on air conditioning energy conservation in the methods of reducing energy consumption in green buildings. In view of this, this study will reduce air conditioning energy consumption on the ground of human thermal uncomfortable posture recognition algorithm, achieving the goal of reducing energy consumption in green buildings.

3. Construction of a recognition algorithm for human thermal discomfort posture

The architecture of the research algorithm consists of two main parts: bone key point extraction and thermal uncomfortable posture recognition. By analyzing the relationship between bone key points in different movements, a thermal discomfort posture recognition algorithm was constructed, thereby achieving the goal of thermal discomfort detection.

3.1. Extracting key points of human skeleton

For thermal comfort detection on the ground of posture recognition, this study proposes a framework for thermal uncomfortable posture recognition algorithm, which mainly utilizes bone key points for thermal uncomfortable posture recognition. The entire study is divided into three modules. The first module is about extracting key points from bones. The second module is to normalize the extracted information. The final module is the pose recognition module [16–18]. This study adopts the deconstructive key point regression algorithm, as it can improve accuracy and computational efficiency, and conduct end-to-end training. The deconstructive key point regression algorithm framework is shown in Fig. 1.

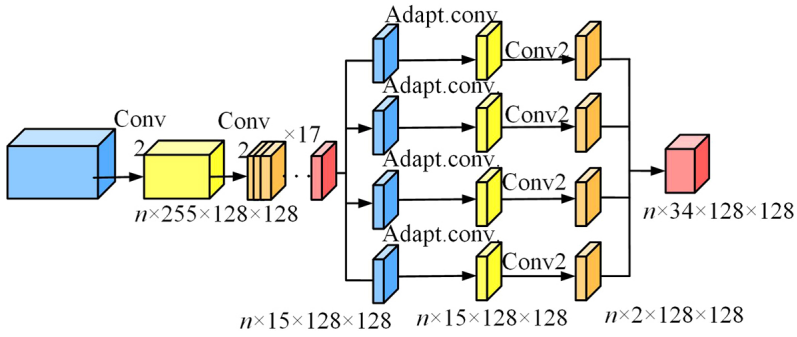

Fig. 1. Framework of deconstructive key point regression algorithm

Figure 1 shows that the input of the deconstructive key point regression algorithm is the output of HRNet. For the COCO dataset, human pose estimation consists of 17 key points, so the deconstructive key point regression algorithm regressed the coordinates of these 17 key

points separately [19, 20]. The specific approach is to divide the output feature vectors of the HRNet backbone network into 17 parts, and each feature vector is convolutionally calculated to obtain the corresponding key point features [21]. Finally, in this way, the two-dimensional offset value vector for each key point can be obtained.

3.2. Pixel information extraction on the ground of adaptive convolutional computing

Adaptive convolution computation can extract pixel information from key point regions. The convolution kernel is defined by the predicted affine transformation matrix and translation matrix generated by each pixel, as shown in (3.1).

$$G_s^q = A^q G_t + [tt...t] \tag{3.1}$$

In (3.1), G_s^q represents the final offset, A^q represents the affine transformation matrix, and the input dimension is $[G, H, W]$. The calculated channel number C maps to 4, and then this channel number is transformed into the affine transformation matrix $[H, W, 2, 2]$. $[tt...t]$ maps the number of channels to 2 through convolution operations on the input dimension. G_t represents an ordinary 3×3 two-dimensional convolution, t represents "three", with a size of 3, and the expression is shown in (3.2).

$$G_t = \begin{bmatrix} -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{3.2}$$

In (3.2), the first row represents 9 offsets of the X coordinate, and the second row represents 9 offsets of the Y coordinate. An expression for the coordinate center is shown in (3.3).

$$y(q) = \sum_{i=1}^{9} W_i X(g_{si}^q + q) \tag{3.3}$$

In (3.3), q represents the center of the X and Y coordinates. g_{si}^q represents the offset of two-dimensional coordinates. W_i represents the weight matrix of the convolutional kernel. By using multi branch regression with the same branch structure, adaptive convolution can effectively extract feature information of pixels around the center of each branch, as expressedin (3.4).

$$\begin{cases} O_1 = F_1(X_1) \\ O_2 = F_2(X_2) \\ \quad \cdots \\ O_K = F_K(X_K) \end{cases} \tag{3.4}$$

In (3.4), F is obtained through independent training of a single identical branch structure. The output feature vectors of the backbone network are represented by X. O represents the coordinate information of key points. K represents the number of bone keys. The expression of the normalization method is shown in (3.5).

$$\text{KeyPonts} = \text{Keypoints} - X_{\min} \tag{3.5}$$

In (3.5), KeyPonts represents the coordinate matrix of the original bone key points. $X_{\min}$ represents the minimum value of the coordinate system, and there is also an expression for KeyPonts as shown in (3.6).

$$\text{KeyPonts} = \text{Keypoints}/(X_{\max} - X_{\min}) \tag{3.6}$$

In (3.6), $X_{\max}$ represents the maximum value of the abscissa. The expression of KeyPonts about the Y-axis is shown in (3.7).

$$\text{KeyPonts} = \text{Keypoints} - Y_{\min} \tag{3.7}$$

In (3.7), $Y_{\min}$ represents the minimum value of the ordinate. There is also an expression for KeyPonts about the Y-axis as shown in (3.8).

$$\text{KeyPonts} = \text{Keypoints}/(Y_{\max} - Y_{\min}) \tag{3.8}$$

In (3.8), $Y_{\max}$ represents the maximum value of the ordinate. The study selected the loss function as the cross entropy loss function, and its expression is shown in (3.9).

$$\text{CEL}(W) = \frac{1}{N}\sum_{i=1}^{N} H\left(y^{(i)}, \hat{y}^{(i)}\right) \tag{3.9}$$

In (3.9), W represents the network model parameter matrix. N represents the number of training samples. $y^{(i)}$ represents the true value of the label. $\hat{y}^{(i)}$ represents the predicted value of the label. $H(y^{(i)}, \hat{y}^{(i)})$ represents the information entropy of $y^{(i)}$, $\hat{y}^{(i)}$, and its expression is shown in (3.10).

$$H(y^{(i)}, \hat{y}) = -\sum_{j=1}^{q} y_j^{(i)} \log \hat{y}_j^{(i)} \tag{3.10}$$

In (3.10), q represents the total number of classifications. $y_j^{(i)}$ is an element in $y^{(i)}$, and its value is either 0 or 1. The incorrect category is 0. The output of the $soft$ max function is shown in (3.11).

$$\hat{y}_k = \text{soft max}(o_k) = \frac{\exp(o_k)}{\sum_{i=1}^{q} \exp(o_i)} \tag{3.11}$$

In (3.11), o_k represents the output of the neural network. The training and testing stages will use accuracy indicators, and the calculation of accuracy is shown in (3.12).

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} = \hat{y}^{(i)}\right) \tag{3.12}$$

In (3.12), N represents the number of samples. $y^{(i)} = \hat{y}^{(i)}$ indicates that the actual value is equal to the predicted value.

3.3. Recognition of human thermal discomfort posture

By training a deep neural network, the model can automatically learn and recognize various poses, improving the accuracy and efficiency of recognition. Therefore, after obtaining the coordinate information of bone key points, it was realized that the features input into the neural network were much less than the image information. This study constructed three algorithms for pose recognition: fully connected network, 1D convolution, and 1D convolution + LSTM. When processing features extracted from the coordinate information of bone key points, due to the relatively small number of features, a fully connected network was first constructed for pose recognition, and the network structure is shown in Fig. 2.

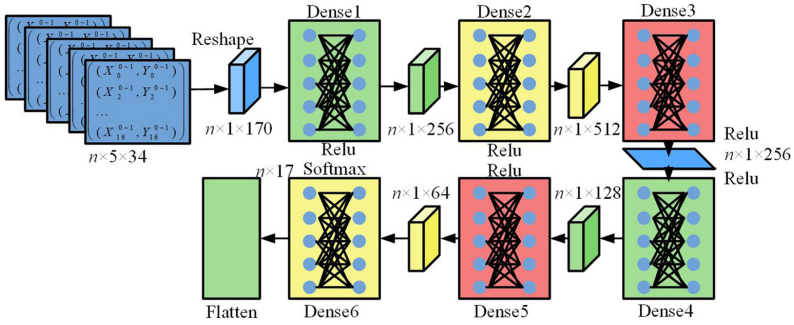

Fig. 2. Fully connected network structure diagram

In Fig. 2, the dimension of the input data is $n \times 5 \times 34$, and through the Reshape operation, the dimension of its feature vector becomes $n \times 1 \times 170$. These feature vectors are further transmitted through a series of fully connected layers (Dense1, Dense2, Dense3, Dense4, Dense5), each layer increasing the dimensionality of the feature vectors and introducing more nonlinearity. During this process, the probability of Dropout is 0.5, which helps to prevent overfitting. Finally, through the Dense6 fully connected layer, the dimension of the feature vector becomes $n \times 1 \times 17$, indicating that there are 17 pose categories. In the final step, the Flatten operation flattens the 17 dimensional feature vectors into a one-dimensional feature vector, with each element representing the score of a pose category. By adjusting the parameters of 1D convolution, spatial features in sequence data can be effectively captured and used for attitude recognition, as shown in Fig. 3.

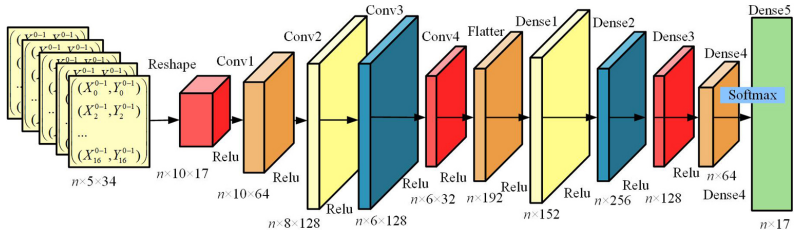

Fig. 3. 1D convolutional network architecture

In Fig. 3 the Reshape operation was performed on the input data with a dimension of$n \times 5 \times 34$, resulting in a feature vector with a dimension of $n \times 10 \times 17$. Subsequently, the feature vectors were processed using 1D convolutional structures (CONV1D_1, CONV1D_2, CONV1D3, and CONV1D_4) to obtain the feature vectors of $n \times 10 \times 64$, $n \times 8 \times 128$, $n \times 6 \times 128$, and $n \times 6 \times 32$. Then, a Flatten operation was performed on the feature vectors to obtain n * 192 feature vectors. Then, a 5-layer fully connected structure was used to process the feature vectors, obtaining the feature vectors of $n \times 512$, $n \times 256$, $n \times 128$ and $n \times 64$. In the fully connected structure, a Dropout probability of 0.5 was set to prevent overfitting. Finally, the dimension of the feature vector was obtained as $n \times 17$, indicating 17 pose categories. LSTM solves the gradient vanishing problem of RNN when processing long sequences by introducing memory units, which can store previous information and use it for the output at the current time. By combining 1D convolution and LSTM, sequence features can be captured using 1D convolution, and sequence information can be saved and transmitted using LSTM to achieve more accurate pose recognition, as shown in Fig. 4.

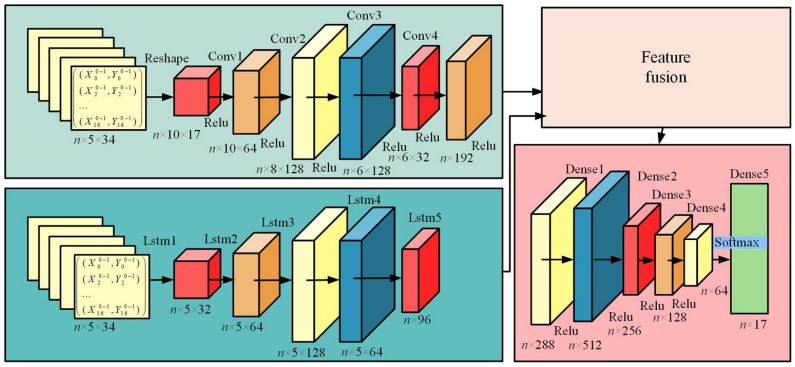

Fig. 4. 1D convolution + LSTM network structure

In Fig. 4, the dimension of the input data is $n \times 5 \times 34$, and the length of the feature vector is 5×34. The network structure is shown in 3.12, which is divided into two branches: 1D convolution branch and LSTM branch. Firstly, the input of the 1D convolution branch undergoes a Reshape operation to obtain a feature vector with a dimension of $n \times 10 \times 17$. Next, through a four layer 1D convolutional structure, the dimensions of the feature vectors are gradually transformed into $n \times 6 \times 32$. Finally, through the Flatten operation, the dimension of the feature vector is flattened to $n \times 192$. On the other hand, the input of the LSTM branch is $n \times 5 \times 34$. After processing the LSTM structures LSTM_1 to LSTM_5, the dimensions of the feature vectors obtained are $n \times 5 \times 32$, $n \times 5 \times 64$, $n \times 5 \times 128$, $n \times 5 \times 64$, and $n \times 96$. After obtaining feature vectors from two branches, they are fused to obtain a fused feature vector with a dimension of $n \times 288$. Next, through a 5-layer fully connected structure, the dimensionality of the feature vectors is gradually reduced to $n \times 17$. It is worth noting that the probability of Dropout in fully connected structures is 0.5 to prevent overfitting. Finally, the study obtained a feature vector with a dimension of $n \times 17$, which represents 17 pose categories. The output of 17 represents 17 posture categories.

4. Result analysis of human thermal uncomfortable posture recognition algorithm

To verify the recognition effect of the human thermal uncomfortable posture recognition algorithm, this study first conducted performance analysis and tested the bone extraction effect, accuracy, and loss value of the research method. Then it compares and analyzes the research algorithm with traditional recognition algorithms to verify the superiority of the research method.

4.1. Performance testing of human thermal uncomfortable posture recognition algorithm

This study conducted performance testing on the constructed human thermal uncomfortable posture recognition algorithm. To avoid errors caused by different experimental equipment, the study used the same computer for the experiment, with an Intel Xeon E5-2697 v2 CPU, 32 GB RAM, Windows 10 Home operating system, and 16 GB memory. Meanwhile, this study constructed a dataset on the ground of thermal discomfort posture. Firstly, the skeleton key point extraction module studied was tested, and the test results are shown in Fig. 5.

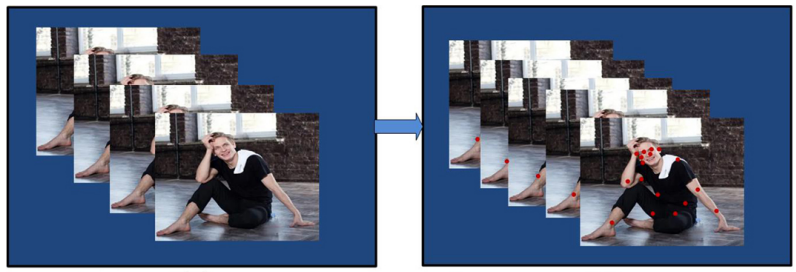

Fig. 5. Bone key extraction results: (a) Original image, (b) Bone key extraction result

Fig. 5 reflects the extraction of bone key points on the ground of the deconstructive key point regression algorithm. Fig. 5(a) is the original action sequence image, and Fig. 5(b) is the coordinate image of the bone key points marked with red circles after passing through the bone key point extraction module. This indicates that this algorithm clearly annotates the key points of the human skeleton. The experiment adopts a fully connected network algorithm, and after 2000 iterations of training, the accuracy and loss values are shown in Fig. 6.

In Fig. 6, on the training set, the loss value decreases as the number of iterations increases, and when the number of iterations reaches 600, the loss value tends to stabilize. The accuracy shows a stable increasing state as the number of iterations increases, reaching its maximum value of 87.23% when the number of iterations is 1200. On the test set, the trend of change is consistent with that of the training set, but the fluctuation is greater, and the loss value decreases with the increase of iteration times. When the number of iterations reaches 700, the loss value tends to stabilize. The accuracy shows a stable increasing state as the number of

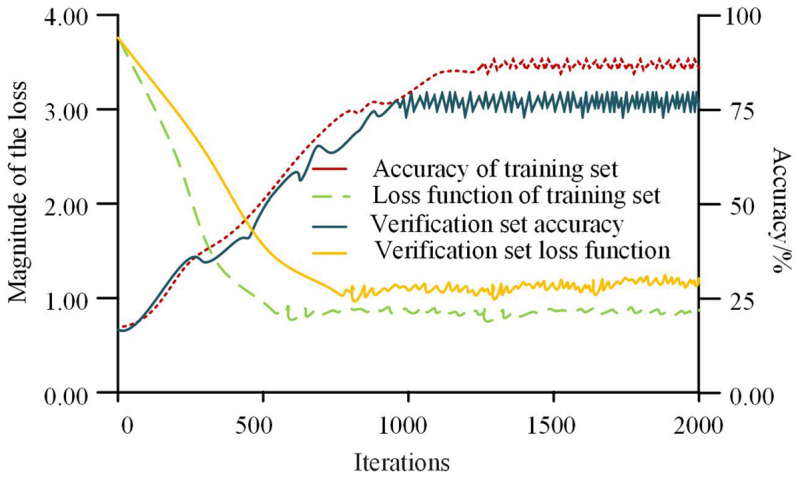

Fig. 6. Accuracy and loss changes under fully connected network algorithms

iterations increases, reaching its maximum value of 76.23% when the number of iterations is 1300. The experiment used a 1D convolutional network attitude recognition algorithm, which underwent 2000 iterations of training. The accuracy and loss values are shown in Fig. 7.

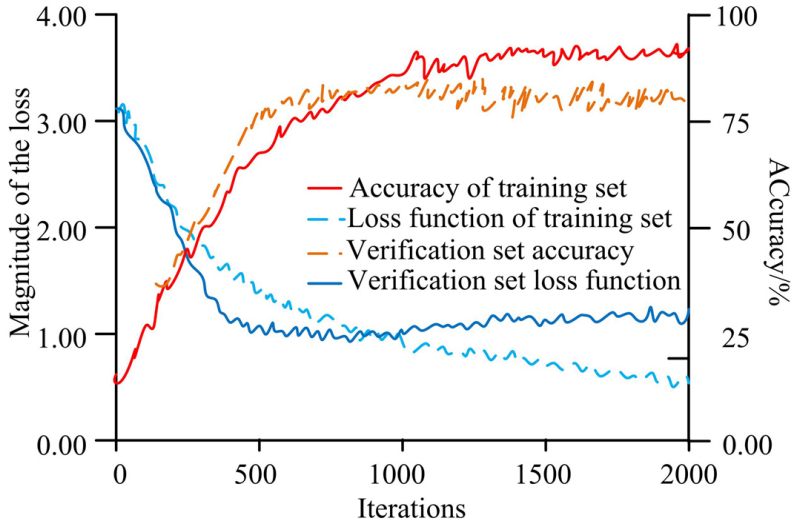

Fig. 7. Accuracy and loss value changes under 1D convolutional network pose recognition algorithm

In Fig. 7, on the training set, when the number of iterations reaches 1300, the loss value tends to stabilize. The accuracy shows a stable increasing state as the number of iterations increases, reaching its maximum value of 92.76% when the number of iterations is 1000. On the test set, the trend of curve changes is similar to that of the training set, but the fluctuation

is relatively large. The loss value decreases with the increase of the number of iterations. When the number of iterations reaches 400, the loss value tends to stabilize. The accuracy increases with the number of iterations, reaching its maximum value of 79.23% at an iteration of 400, which is 13.53% higher than the training set. This experiment uses a pose recognition algorithm that combines 1D convolution with LSTM network. After 2000 iterations of training, the accuracy and loss values are shown in Fig. 8.

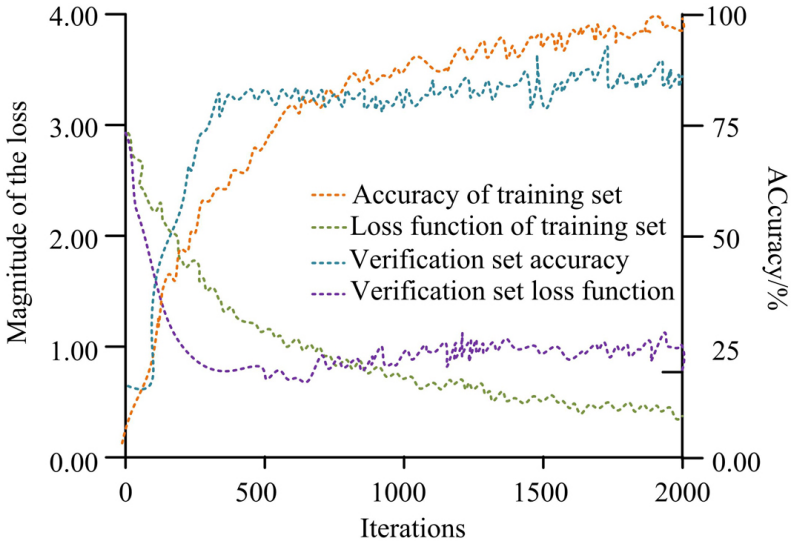

Fig. 8. Accuracy and loss changes of attitude recognition algorithm combining 1D convolutional and LSTM networks

In Fig. 8, on the training set, when the number of iterations reaches 1100, the loss value tends to stabilize. The accuracy shows a stable increasing state as the number of iterations increases, reaching its maximum value of 99.98% when the number of iterations is 1500. The fitting degree between the test set and the training set is higher, and the loss value decreases as the number of iterations increases. When the number of iterations reaches 600, the loss value tends to stabilize. The accuracy increases with the number of iterations, reaching its maximum value of 89.85% when the number of iterations is 400. Compared to the methods used in Figs. 6 and 7, the accuracy has been significantly improved. However, the highest accuracy rate has not yet reached over 90%. To achieve more ideal results, it is considered to classify the uncomfortable postures of the human body in a more detailed manner in the dataset to avoid misjudgment caused by similar actions. The classification situation and its results are shown in Table 1.

In Table 1, compared to Fig. 8, the accuracy of the study has been significantly improved, with an accuracy improvement of over 98% for the three types of datasets after classification. For the first type of dataset, the accuracy reaches 99.51%. For the second type of dataset, the accuracy rate is 98.56%. For the third type of dataset, the accuracy rate is 98.95%. Overall, the reclassification of the dataset has improved the accuracy of the algorithm.

Table 1. Accuracy under different classification situations

Posture type	Pose name	Meaning of posture	Posture description	Accuracy/%
1	Head up and wipe sweat	Heat	Place your hand on your forehead to wipe sweat	99.51%
	Arms open	Heat	Hands open	
	Pulling clothes	Heat	Shake clothes with your hands	
	Place your hand on your neck	Cold	Put your hands around your neck	
	Hold your hands in front of your chest	Cold	Cross hands in front of the chest	
2	Fan with your hands	Heat	Place your hand close to your head and fan	98.56%
	Shrug one's shoulders	Cold	Shoulder shrug up	
	Rubbing hands	Cold	Rub your hands back and forth	
	Arm contraction	Cold	Bend your arms while retracting your neck downwards	
	Haze gas	Cold	Put your hand on your mouth and exhale gas	
3	Roll up one's sleeves	Heat	Roll up the sleeves of the clothes	98.95%
	Flexing one's head	Heat	Raise your hand and scratch your head	
	Chop the ground	Cold	Step back and forth with your left and right feet	
	Walk	Moderate body sensation	Standing and walking	
	Cross legs	Cold	Overlap your legs together	

4.2. Comparative analysis of algorithms for recognizing human thermal discomfort posture

To verify the superiority of the research method, three categories of human thermal uncomfortable postures in the dataset were introduced, including GCN algorithm, 3D Convolutional Neural Networks (CNN), and PoseNet model, and their performance was compared and analyzed. The comparison results of its accuracy are shown in Fig. 9.

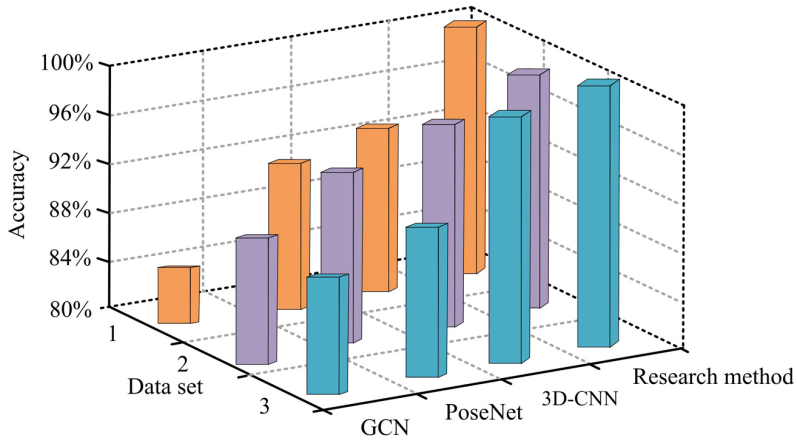

Fig. 9. Comparative analysis of accuracy of four methods

In Fig. 9, for the first type of thermal uncomfortable posture, the research method has the highest accuracy, reaching 98.23%. It is 7.23% higher than the 3D-CNN algorithm, 10.23% higher than the PoseNet algorithm, and 17.23% higher than the GCN algorithm. For the second type of thermal uncomfortable posture, the accuracy of the research method is still the highest, reaching 97.89%, significantly higher than the other three algorithms. For the third type of thermal uncomfortable posture, the accuracy of the research method is better than that of the 3D-CNN algorithm, reaching 99.65% and 97.34%, respectively.

To further investigate the impact of algorithms on energy consumption, experiments were conducted to compare the energy consumption results of different algorithms recognizing poses. Three types of buildings A, B, and C were set up in the experiment, and the results are shown in Fig. 10.

Figure 10 shows that the research algorithm has the lowest energy consumption rate, with energy consumption of around 82.3% for the three types of buildings, while the 3D-CNN, PoseNet, and GCN algorithms regulate energy consumption of over 85% for the three types of buildings. It can be seen that research algorithms can accurately evaluate human thermal posture and have extremely precise energy regulation.

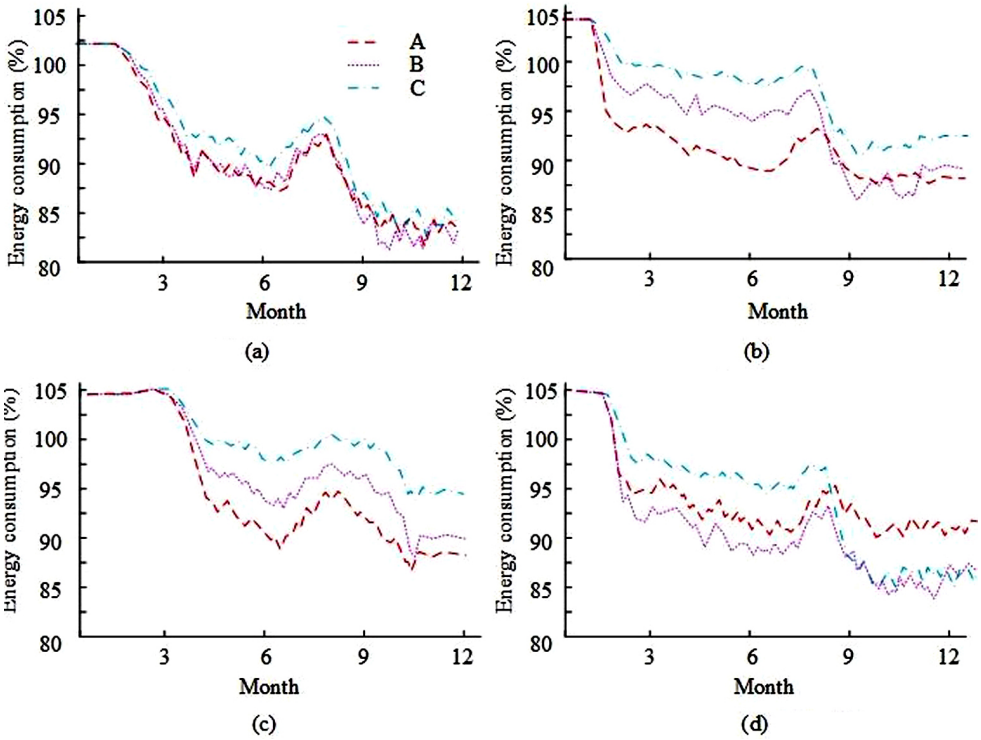

Fig. 10. Comparison of different algorithms for regulating building energy consumption: (a) Research method, (b) 3D-CNN, (c) PoseNet, (d) GCN

5. Conclusions

To reduce energy consumption in green buildings, this study proposes a new algorithm for recognizing human thermal uncomfortable postures. The results showed that the experiment used a pose recognition algorithm combining 1D convolution and LSTM network. On the training set, when the number of iterations was 1100, the loss value tended to stabilize. When the iteration number is 1500, the accuracy reaches its maximum value, which is 99.98%. On the test set, when the number of iterations is 600, the loss value tends to stabilize, and when the number of iterations is 400, the accuracy reaches its maximum value, which is 89.85%. After classifying the uncomfortable postures of the human body in the construction dataset, the accuracy rate for the first type dataset reached 99.51%. For the second type of dataset, the accuracy rate is 98.56%. For the third type of dataset, the accuracy reached 98.95%, and classification improved the accuracy of the algorithm. In the comparison of the four algorithms, for the first type of thermal uncomfortable posture, the research method has the highest accuracy, reaching 98.23%. For the second type of thermal uncomfortable posture, the accuracy of the research method reached 97.89%, significantly higher than the other three algorithms. For the third type of thermal uncomfortable posture, the accuracy

of the research method reached 99.65%. The experimental results validate the superiority of the research method, indicating that the research algorithm has high accuracy. The green building energy consumption reduction method on the ground of human thermal uncomfortable posture recognition algorithm has positive significance and practical value. However, this study still has certain limitations, such as insufficient data samples and optimization of algorithm performance, which require further exploration. In future research, the range of data samples will be expanded to improve the generalization ability of the algorithm. In terms of action definition, due to the different behaviors of each person in a hot and uncomfortable environment, for different regions and climates, more parameters need to be set for model learning in future research. The different climates in different regions have high requirements for regulating the energy consumption of green buildings. One building with strong climate adaptability can automatically adjust its internal environment according to different climate conditions to maintain internal air quality and comfort, effectively reduce energy consumption, reduce carbon emissions, and improve the sustainability of the building. The use of research algorithms can effectively predict human posture behavior under different climates, providing solutions for building design.

References

[1] Z. Wu, Y. Zhao, and N. Zhang, "A literature survey of green and low-carbon economics using natural experiment approaches in top field journal", *Green and Low-Carbon Economy*, vol. 1, no. 1, pp. 2–14, 2023, doi: 10.47852/bonviewGLCE3202827.

[2] A.M. Usman and M.K. Abdullah, "An assessment of building energy consumption characteristics using analytical energy and carbon footprint assessment model", *Green and Low-Carbon Economy*, vol. 1, no. 1, pp. 28–40, 2023, doi: 10.47852/bonviewGLCE3202545.

[3] Y. Fang, B. Luo, T. Zhao, D. He, B.B. Jiang, and Q.L. Liu, "ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting", *CAAI Transactions on Intelligence Technology*, vol. 7, no. 4, pp. 744–757, 2022, doi: 10.1049/cit2.12145.

[4] Q. Men, E.S.L. Ho, H.P.H. Shum, and H. Leung, "Focalized contrastive view-invariant learning for self-supervised skeleton-based action recognition", *Neurocomputing*, vol. 537, no. 7, pp. 198–209, 2023, doi: 10.1016/j.neucom.2023.03.070.

[5] Y. Yuan, Z. Lu, Z. Yang, M. Jian, L. Wu, Z. Li, and X. Liu, "Key frame extraction based on global motion statistics for team-sport videos", *Multimedia Systems*, vol. 28, no. 2, pp. 387–401, 2022, doi: 10.1007/s00530-021-00777-7.

[6] W. Zhuang, Y. Chen, J. Su, B. Wang, and C. Gao, "Design of human activity recognition algorithms based on a single wearable IMU sensor", *International Journal of Sensor Networks*, vol. 30, no. 3, pp. 193–206, 2019, doi: 10.1504/IJSNET.2019.100218.

[7] W. Zhu, Y. She, J. Hu, B. Wang, J. Mu, and S. Li, "A hybrid relative navigation algorithm for a large-scale free tumbling non-cooperative target", *Acta Astronautica*, vol. 194, no. 5, pp. 114–125, 2022, doi: 10.1016/j.actaastro.2022.01.028.

[8] C. Wang, Z. Zhang, and Z. Xi, "A human body based on sift-neural network algorithm attitude recognition method", *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 1, pp. 129–133, 2020, doi: 10.1166/jmihi.2020.2867.

[9] X. Gao, Y. Yang, Y. Zhang, M. Li, J. Yu, and S. Du, "Efficient Spatio-Temporal Contrastive Learning for Skeleton-Based 3-D Action Recognition", *IEEE Transactions on Multimedia*, vol. 25, no. 1, pp. 405–417, 2023, doi: 10.1109/TMM.2021.3127040.

[10] C.L. Quan, L. You, F. Shen, et al., "Pose recognition in sports scenes based on deep learning skeleton sequence model", *Journal of Intelligent and Fuzzy Systems*, vol. 53, no. 3, pp. 1–10, 2021, doi: 10.3233/JIFS-189834.

[11] C. Pan, H. Cao, W. Zhang, X. Song, and M. Li, "Driver activity recognition using spatial-temporal graph convolutional LSTM networks with attention mechanism", *IET Intelligent Transport Systems*, vol. 15, no. 2, pp. 297–307, 2021, doi: 10.1049/itr2.12025.

[12] X. Wen, "Energy consumption monitoring model of green energy-saving building based on fuzzy neural network", *International Journal of Global Energy Issues*, vol. 44, no. 5-6, pp. 396–412, 2022, doi: 10.1504/IJGEI.2022.125405.

[13] Q. Wang, Y.J. Hu, J. Hao, N. Lv, T. Li, and B. Tang, "Exploring the influences of green industrial building on the energy consumption of industrial enterprises: A case study of Chinese cigarette manufactures", *Journal of Cleaner Production*, vol. 231, no. 10, pp. 370–385, 2019, doi: 10.1016/j.jclepro.2019.05.136.

[14] Y. Wang, M. Lin, K. Xu, S. Zhang, and H. Ma, "Energy consumption analysis of glass house using electrochromic window in the subtropical region", *Journal of Engineering Design and Technology*, vol. 19, no. 1, pp. 203–218, 2021, doi: 10.1108/JEDT-12-2019-0348.

[15] J. Xu, "Research on energy consumption control method of green building based on BIM technology", *International Journal of Industrial and Systems Engineering*, vol. 40, no. 3, pp. 399–414, 2022, doi: 10.1504/IJISE.2022.122248.

[16] Y. Singh and L. Kaur, "Effective key-frame extraction approach using TSTBTC–BBA", *IET Image Processing*, vol. 14, no. 4, pp. 638–647, 2020, doi: 10.1049/iet-ipr.2018.6361.

[17] S. Xu, S. Sun, Z. Zhang, F. Xu, and J.H. Liu, "BERT gated multi-window attention network for relation extraction", *Neurocomputing*, vol. 492, no. 1, pp. 516–529, 2022, doi: 10.1016/j.neucom.2021.12.044.

[18] X. Gu, L. Lu, S. J. Qiu, Q. Y. Zou, and Z.Y. Yang, "Sentiment key frame extraction in user-generated micro-videos via low-rank and sparse representation", *Neurocomputing*, vol. 410, no. 14, pp. 441–453, 2020, doi: 10.1016/j.neucom.2020.05.026.

[19] H. Xie, Y. Zhong, Z. Yu, A. Hussain, and G. Chen, "Automatic 3D human body landmarks extraction and measurement based on mean curvature skeleton for tailoring", *The Journal of the Textile Institute*, vol. 113, no. 8, pp. 1677–1687, 2022, doi: 10.1080/00405000.2021.1944513.

[20] R. Liao, Y. Zhang, Y. Wang, and D. Dai, "Multi-view face pose recognition model construction based on a typical correlation analysis algorithm", *International Journal of Biometrics*, vol. 13, no. 2-3, pp. 289–304, 2021, doi: 10.1504/IJBM.2021.114654.

[21] A.Z.O. Al-Hijazeen, M. Fawad, M. Gerges, K. Koris, and M. Salamak, "Implementation of digital twin and support vectormachine in structural health monitoring of bridges", *Archives of Civil Engineering*, vol. 69, no. 3, pp. 31-47, 2023, doi: 10.24425/ace.2023.146065.

Received: 2023-12-15, Revised: 2024-02-27