

Volume 1, Issue 1, 2024

**Print ISSN: 3007-6897
Online ISSN: 3007-6900**

Multidisciplinary Journal of Engineering and Technology



Copyright© Upubscience Publisher

Multidisciplinary Journal of Engineering and Technology

Volume 1, Issue 1, 2024



Published by Upubscience Publisher

Copyright© The Authors

Upubscience Publisher adheres to the principles of Creative Commons, meaning that we do not claim copyright of the work we publish. We only ask people using one of our publications to respect the integrity of the work and to refer to the original location, title and author(s).

Copyright on any article is retained by the author(s) under the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Authors grant us a license to publish the article and identify us as the original publisher.

Authors also grant any third party the right to use, distribute and reproduce the article in any medium, provided the original work is properly cited.

Multidisciplinary Journal of Engineering and Technology

Print ISSN: 3007-6897 Online ISSN: 3007-6900

Email: info@upubscience.com

Website: <http://www.upubscience.com/>

Table of Content

STRESS AND STRAIN CONTROL TECHNOLOGY FOR STEEL STRUCTURE WELDING	1-3
Li Li, JiangLong Meng*, ShaoBo Gu, ZhiCheng Bai, HongQiang Fu, Chao Wang	
MECHANICAL ANALYSIS OF CANTILEVERED RIGID STEEL PLATE SHEAR WALL	4-8
Jing Fan, RuiRong Liang, BaoAn Zhang*, Wei Cao, YiMing Wang, XunYin Liu	
SISM: A SELF-INTERACTIVE APPROACH TO WEB NEWS SUMMARIZATION USING DEEP LEARNING	9-16
Aleksy Nowaks	
IDENTIFICATION AND TRACKING OF AERIAL UAVS BASED ON DEEP LEARNING VISUAL ALGORITHMS	17-28
JianJun Song	
RESEARCH AND APPLICATION OF CONSTRUCTION TECHNOLOGY OF GROUND STEEL STRUCTURE OF DOUBLE-LAYER EXHIBITION HALL	29-33
Yi Zhu, YueQiang Cao*, ShenShen Zhu, ZhiCheng Bai, XiaoXia Zhao, ZhenJie Tan	
PERFORMANCE COMPARISON OF CEEMDAN-LSTM AND BASIC LSTM MODELS IN PREDICTING REALIZED VOLATILITY	34-38
ZheShuo Zhang	
CONSTRUCTION TECHNOLOGY OF LARGE-SPAN CAVITY VENTILATION SHAFT WITH SMALL CLEARANCE ON ROOF	39-42
YongYan Li, Yang Yang*, XinXin Jin, ZhiCheng Bai, Jie Luo, Kun Wu	
A COMMAND SUPPORT MODEL BASED ON BLOCKCHAIN AND THE METAVERSE	43-50
ChaoYue Song, YiLiang Han*, XuGuang Wu	
ENHANCED UNSUPERVISED IMAGE-TO-IMAGE TRANSLATION USING CONTRASTIVE LEARNING AND HISTOGRAM OF ORIENTED GRADIENTS	51-59
WanChen Zhao, WenHan Wang, Cheng Zhang, XiaoLei Qu*	
STUDY ON EQUIVALENT STATIC WIND LOAD ON LONG-SPAN ROOF BASED ON RESPONSE-RELATED SCALE	60-65
Yi Zhu, ZhiCheng Bai*, XiaoXia Zhao, YinGuang Wang, Xu Zhang	

STRESS AND STRAIN CONTROL TECHNOLOGY FOR STEEL STRUCTURE WELDING

Li Li¹, JiangLong Meng^{2,*}, ShaoBo Gu¹, ZhiCheng Bai², HongQiang Fu², Chao Wang²

¹Hangzhou Yunhe Group Investment Development Co., Ltd., Hangzhou, Zhejiang 310000, China.

²China Construction Fourth Engineering Bureau Co., Ltd., Guangzhou, Guangdong 510000, China.

Corresponding Author: JiangLong Meng, Email: 19822660870@163.com

Abstract: Stress and strain will inevitably occur during the welding process of steel structures. If not effectively controlled, it will have a significant impact on the quality, performance and safety of steel structures. This paper explores the stress and strain control technology of steel structure welding, and elaborates on the factors affecting welding and various control technical measures. This paper explores the application and effect of these technologies in practical engineering, and looks forward to future development trends. It aims to provide a comprehensive technical reference for the field of steel structure welding and promote the continuous development and progress of this field.

Keywords: Steel structure; Welding; Stress and strain control

INTRODUCTION

Steel structures are widely used in many fields such as construction, bridges and machinery due to their high strength, excellent mechanical properties and relatively low cost. Welding, as the main connection method of steel structures, will produce complex thermal processes during the welding process, resulting in stress and strain in the welded joint area. If these stresses and strains cannot be effectively controlled, they may cause problems such as welding cracks, deformation and fatigue performance degradation, seriously affecting the quality and safety of steel structures. Therefore, in-depth research on stress and strain control technology for steel structure welding has important theoretical and practical significance.

1 PROJECT OVERVIEW

The China Grand Canal Museum Project is located in Hangzhou Canal New City. This project has two underground floors with an elevation range of -2.000-13.100 m. The basement steel structure is mainly composed of core tube rigid steel columns and podium outer frame circular tube columns. The tower structure has 15 floors above ground and the podium structure has 8 floors. The tower steel components mainly include: steel columns + steel beams + lap columns + folded beams + steel plate walls + split-level trusses + cantilever trusses + cantilever walls + large-section ring beams. The podium steel structure includes orthogonal trusses, one-way trusses, core tube built-in rigid structures, split-level trusses, cantilever trusses and outer frame steel pipe columns. Project effect can be seen in Figure 1.



Figure 1 Project Effect

2 KEY FACTORS AFFECTING WELDING STRESS AND STRAIN

Welding process parameters: welding current, voltage, welding speed and other process parameters directly affect welding heat input and cooling speed, and thus affect the size and distribution of welding stress and strain; welding process also has a key impact on the installation accuracy of steel structure components. Specifically, there are 4

common impacts:

(1) Welding preparation. The weld preparation before welding is extremely critical to welding quality and installation accuracy. The weld preparation covers cleaning the welding surface, angle, size, etc., which plays a significant role in ensuring the matching accuracy of the welded joint.

(2) Welding sequence. The welding sequence during the welding process will also affect the installation accuracy. Selecting a reasonable welding sequence can reduce the accumulation of welding stress and the occurrence of deformation.

(3) Welding temperature. Temperature control during the welding process is also very important for the installation accuracy. Too high welding temperature will cause the weld joint to deform and the heat affected zone to expand, while too low welding temperature may affect the quality of the weld.

(4) Welding speed. The welding speed will also affect the installation accuracy. Too fast or too slow welding speed may affect the quality and deformation of the weld joint.

3 WELDING STRESS AND STRAIN CONTROL TECHNOLOGY

Welding is a commonly used steel structure connection method, but welding stress and strain will inevitably be generated during the welding process; welding stress refers to the stress in the weldment during and after the welding process. These stresses may cause deformation, cracking, fatigue performance degradation of the weldment, etc.; welding strain is the change in size and shape of the weldment under the action of welding stress; the purpose of welding stress and strain control technology is to take a series of measures to reduce, adjust or eliminate these adverse effects to ensure the quality and performance of the welded structure.

Among them, preheating and post-heat treatment refer to the fact that preheating can reduce the cooling rate of the welding area, reduce the temperature gradient, and thus reduce the welding stress. Post-heat treatment helps to eliminate welding residual stress and improve the performance of the welded joint. Choose a suitable welding method, such as arc welding, gas shielded welding, etc., according to the characteristics and requirements of the steel structure. Welding preheating treatment can be seen in Figure 2.



Figure 2 Welding Preheating Treatment

Just like Figure 3, design a reasonable groove form. The design of the groove affects the welding heat input and welding deformation. Use appropriate welding materials to select welding materials that match the parent material to reduce welding stress and strain.

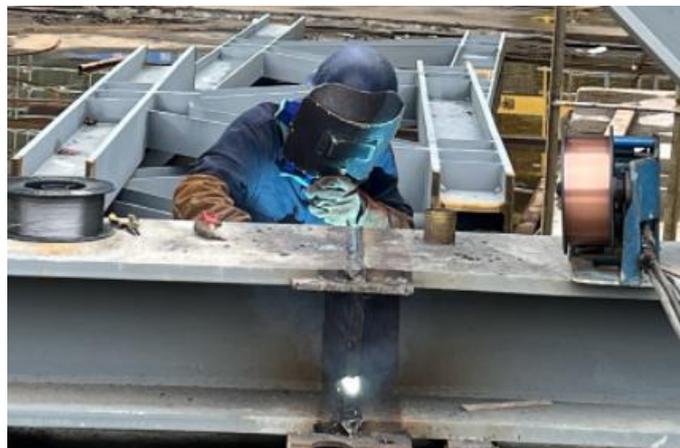


Figure 3 Groove Design

Just like Figure 4, control welding heat input, control welding heat input by adjusting welding process parameters, and reduce welding stress and strain.

Constraint control technology: Rationally design the constraint structure and consider the influence of constraints on welding deformation and stress in the design stage. Use temporary restraint measures and use temporary clamps and other measures to limit deformation during welding.



Figure 4 Constraint control

Welding deformation correction technology, mechanical correction method: use mechanical force to correct the deformed part.

Flame correction method: correct welding deformation by local heating and cooling.

4 CONCLUSION

Steel structure welding stress and strain control technology is a key link to ensure the quality and safety of steel structures. Through an in-depth understanding of the formation mechanism and influencing factors of welding stress and strain, and the comprehensive application of various control technical measures, welding stress and strain can be effectively reduced. With the continuous advancement of science and technology and the continuous growth of engineering needs, welding stress and strain control technology will continue to develop and innovate. In future research and practice, it is necessary to further strengthen multidisciplinary cooperation and exchanges, promote welding technology to a higher level, and provide more reliable guarantees for the wide application of steel structures in various fields.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Zhang Xudong. Influence of residual stress and strain in engineering structures. *Engineering Thermophysics*, 2005, 26(1): 75-78.
- [2] Wang Mingli. Residual stress and strain in mechanical structure welding process. *Welding*, 2007, 26(4): 1-5.
- [3] Li Hongzhi. Common problems and prevention of steel structure welding technology. *Steel Structure*, 2011, 42(1): 42-46.
- [4] Li Dongyu. Residual stress measurement under the influence of pin connection decompression strategy based on characteristic function method. *Light Roof Truss*, 2018, 39(1): 33-38.
- [5] Yang Jun. Technical points in the welding production engineering of building steel structures. *Today's Manufacturing and Upgrading*, 2023, (05): 64-67.
- [6] Zhang Qianhua. Precautions for steel structure welding engineering technology and measures to achieve quality control. *China Metal Bulletin*, 2023, (05): 222-224.
- [7] Li Pei, Xu Dawei. Analysis of steel structure technology in civil engineering construction. *Bulk Cement*, 2023, (02): 139-141.

MECHANICAL ANALYSIS OF CANTILEVERED RIGID STEEL PLATE SHEAR WALL

Jing Fan^{1,2}, RuiRong Liang^{1,2}, BaoAn Zhang^{1,2,*}, Wei Cao^{1,2}, YiMing Wang^{1,2}, XunYin Liu^{1,2}

¹China Construction Fourth Engineering Bureau Co., Ltd. Guangzhou, Guangdong 510000, China.

²China Construction Fourth Engineering Bureau Sixth Construction Co., Ltd. Hefei, Anhui 230000, China.

Corresponding Author: BaoAn Zhang, Email: 19822660870@163.com

Abstract: This paper studies and analyzes the stress characteristics of cantilevered rigid steel plate shear walls in detail. By establishing a mathematical model and using the principles of engineering mechanics, the stress behavior of cantilevered rigid steel plate shear walls under different loading conditions is deeply discussed, mainly analyzing the cracks after pouring concrete in the cantilevered wall. Through simulation analysis and experimental verification, the performance of the structure is evaluated, providing a theoretical basis and reference value for its application in engineering practice.

Keywords: Cantilevered rigid steel plate shear wall; Stress analysis; Cracks; Numerical simulation

1 INTRODUCTION

As a new type of structural system, the performance advantages of cantilevered rigid steel plate shear walls in terms of earthquake resistance and wind resistance have been preliminarily verified [1-2]. However, in-depth research on its stress characteristics is still relatively insufficient. Accurately understanding and grasping the mechanical behavior of cantilevered rigid steel plate shear walls under different stress states is of great significance for the design, construction and performance evaluation of the structure [3-4].

This paper aims to explore the mechanical behavior of cantilevered rigid steel plate shear wall under various stress states through theoretical analysis and simulation, and verify it with actual engineering cases, so as to comprehensively explore the stress behavior of cantilevered rigid steel plate shear wall and provide theoretical basis and technical support for its application in engineering practice.

2 STRUCTURAL CHARACTERISTICS OF CANTILEVERED RIGID STEEL PLATE SHEAR WALL

2.1 Basic Structure

Cantilevered rigid steel plate shear wall is a composite structure composed of steel plate, frame steel member and connector. Steel plate is usually made of high-strength steel with a thickness between 3mm and 6mm, which can be adjusted according to specific design requirements. Frame steel members are usually made of H-shaped steel or box-shaped steel, whose main function is to provide rigidity and support and form an integral structure with the steel plate.

The design of connectors is the key, which includes two main methods: welding and bolt connection. Welding connection is suitable for parts requiring high rigidity and strength, while bolt connection provides a certain flexibility and can better absorb and disperse stress. The arrangement of connectors needs to ensure effective force transmission between steel plate and frame steel member to avoid local stress concentration.

2.2 Working Principle

The main function of the cantilevered rigid steel plate shear wall is to resist horizontal loads, such as wind loads and seismic loads. Its working principle can be analyzed from the following aspects:

- (1) Shear action: The steel plate, as the main load-bearing element, bears shear force under horizontal loads. The steel plate provides the main anti-lateral displacement capacity through its shear stiffness and strength.
- (2) Bending action: The frame steel members mainly bear bending moment and shear force under horizontal loads. They provide additional lateral stiffness and bearing capacity through bending deformation.
- (3) Synergistic effect: The steel plate and the frame steel members form a whole through the connector. The synergistic effect of the two is the key to the excellent seismic performance of the cantilevered rigid steel plate shear wall. The shear deformation of the steel plate and the bending deformation of the frame steel members coordinate with each other to share and resist the external load together.

2.3 Advantages and Applications

Cantilevered rigid steel plate shear walls have the following advantages in high-rise buildings:

- (1) High seismic performance: Due to the synergistic effect of the steel plate and the frame steel components, the cantilevered rigid steel plate shear wall can effectively absorb and disperse seismic energy and has high seismic

performance.

(2) Light weight: Compared with traditional reinforced concrete shear walls, steel plate shear walls are lighter, which helps to reduce the self-weight of the building structure and reduce the bearing requirements of the foundation.

(3) Convenient construction: The construction process of steel plate shear walls is relatively simple, and modular components can be prefabricated in the factory and assembled on site, shortening the construction period and improving construction efficiency.

(4) High space utilization: Due to the thin thickness of the steel plate shear wall, it can provide more available space than traditional shear walls, improving the use efficiency of the building.

3 FORCE ANALYSIS

3.1 Theoretical Analysis

Under the action of horizontal load, the cantilevered rigid steel plate shear wall, as an integral structure, bears shear force, bending moment and axial force. It can be analyzed using the classical plate-shell theory and beam-column theory.

The steel plate mainly transmits shear force through shear deformation, and its shear stiffness is an important factor in determining the overall shear resistance of the shear wall. The shear stress of the steel plate can be calculated by the following formula:

$$\tau = \frac{V}{A_s} \tag{1}$$

The frame steel member mainly bears the bending moment, and its stiffness and strength determine the lateral stiffness of the shear wall. The distribution of the bending moment can be calculated by the following formula:

$$M = \frac{EI}{L^2} \tag{2}$$

Where M is the bending moment, E is the elastic modulus of the steel, I is the section moment of inertia of the frame steel member, and L is the height of the shear wall.

3.2 Concrete Crack Verification after the Cantilever Wall Support Frame is Removed

As shown in the figure 1 and 2 below, after the support frame is removed, the maximum internal force occurs at the GHJ3 support position, and the maximum internal force of the shear wall at the truss support is 171.28kN.

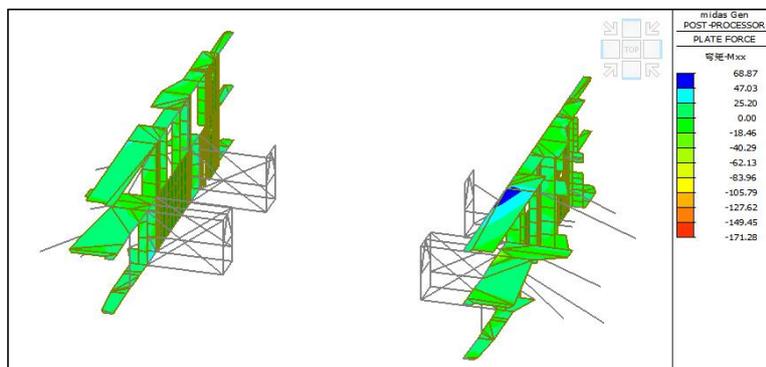


Figure 1 Internal force cloud diagram of concrete wall slab after support frame removal

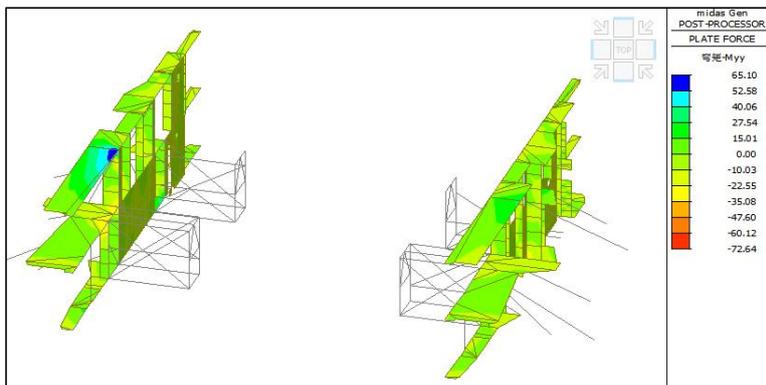


Figure 2 Internal force cloud diagram of concrete wall slab after support frame removal

Cantilever wall truss and internal extension structure are shown below Figure 3:

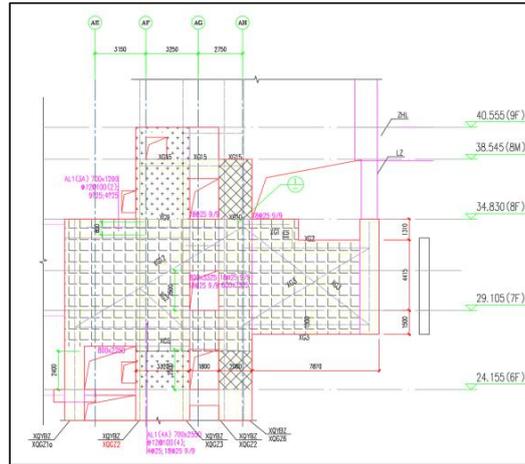


Figure 3 Cantilever wall truss and inward extension structure

The beam size at the support is: 700×2550, reinforcement information: tensile reinforcement: 18φ25, tensile reinforcement area: $3.14 \times 252 \times 18 / 4 = 8831 \text{mm}^2$.

It can be calculated that the concrete crack is $0.026 \text{mm} < 0.4 \text{mm}$, which meets the requirements of the specification. The removal of the support frame has little impact on the structure.

3.3 Calculation of Strain Value of Cantilever Shear Wall

According to the common concrete structure steel content and reinforcement ratio in the cantilever wall of this project, the corresponding strain of steel can be calculated when the crack width of the typical section concrete reaches 0.2mm-0.3mm. Assuming that there is no slip between steel and concrete, the strain of steel at this time is equal to the strain of concrete. The calculation statistical results are detailed in the table 1 below.

Table 1 Statistics of Axial Tensile Strain of Concrete (Crack Width 0.25mm)

When the width of the concrete axial tensile crack is 0.25mm, the nominal strain value of the steel				
section 600×1000			The specification allows cracks	
Strain value	Steel content 0%	Steel content 2%	0.3mm	10%
Reinforcement ratio 2.05%	9.30E-04	8.92E-04	9.23E-04	8.92E-04
Reinforcement ratio 1.63%	9.90E-04	1.02E-03	9.19E-04	8.74E-04
Reinforcement ratio 1%	1.27E-03	8.52E-04	9.05E-04	1.04E-03

The cantilever wall goes through two most unfavorable stages:

- (1) Before the cantilever wall concrete is poured, the internal force at the cantilever truss support reaches the extreme value, and crack verification is required;
- (2) After the cantilever wall is poured, the subsequent design load (design dead load and design live load) further causes the cantilever wall internal force to reach the extreme value, and crack verification is required for the cantilever strong support and the newly poured concrete.

The verification results are as follows Figure 5 and 6:

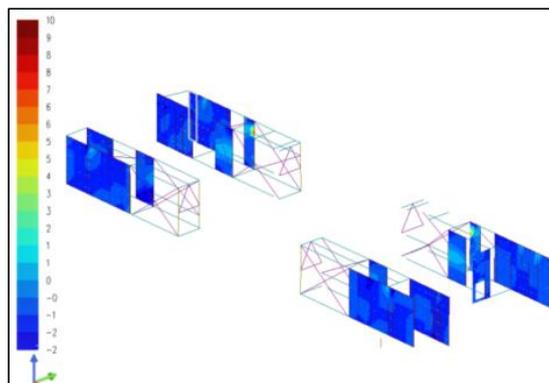


Figure 5 Shear wall stress without concrete pouring

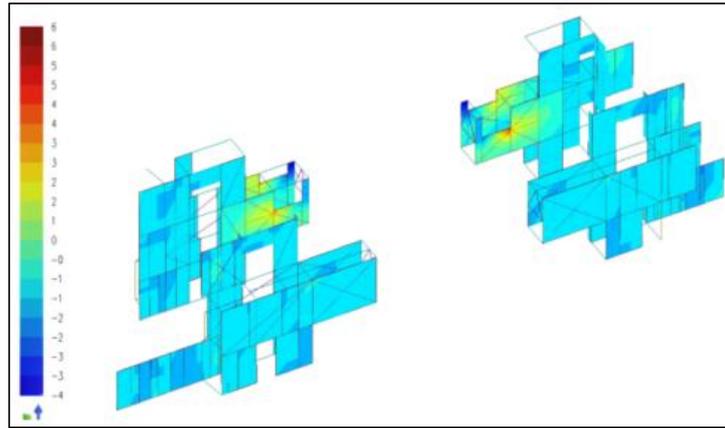


Figure 6 Shear wall stress after concrete pouring

It can be seen from the above figure that the peak stress of the shear wall in the above two stages is less than or equal to 16N/mm^2 , and the comprehensive strain of concrete is $4.44 \times 10^{-4} < 8 \times 10^{-4}$. It can be seen that the comprehensive strain of the shear wall meets the nominal strain control value, and cracks can be effectively controlled.

4 DESIGN SUGGESTIONS AND OPTIMIZATION MEASURES

4.1 Design Suggestions

- (1) The thickness of the steel plate has a significant impact on the shear resistance and overall stiffness of the shear wall. During design, the appropriate steel plate thickness must be selected based on the load conditions and seismic resistance requirements of the actual project.
- (2) The cross-sectional form and size of the frame steel components directly affect the overall stiffness and load-bearing capacity of the shear wall. It is recommended to give priority to high-strength, lightweight H-shaped steel or box-shaped steel as frame steel components.
- (3) The connector plays a key force transmission role in the cantilevered stiff steel plate shear wall. Its stiffness and strength should ensure effective force transmission between the steel plate and the frame steel components to avoid local stress concentration.

4.2 Optimization Measures

- (1) Use high-strength, lightweight steel to improve the load-bearing capacity and seismic performance of the shear wall.
- (2) Optimize construction technology and adopt advanced construction technology and equipment to improve construction efficiency and quality. It is recommended to adopt the process of modular prefabrication + automated welding.
- (3) Use topology optimization methods to optimize the structural form and component arrangement of the shear wall to improve material utilization and structural performance.

5 CONCLUSION

5.1 Conclusion

Through theoretical analysis, numerical simulation and experimental verification, this paper systematically studies the stress characteristics of cantilevered stiff steel plate shear walls and puts forward corresponding design suggestions and optimization measures. The main conclusions are as follows:

- (1) Cantilevered stiff steel plate shear walls exhibit good mechanical properties under horizontal loads and have high load-bearing capacity and stiffness.
- (2) The experimental results are reliable: Through experimental verification, the stress-strain relationship and deformation mode of the cantilevered stiff steel plate shear wall under different load conditions are highly consistent with the theoretical analysis and numerical simulation results, verifying the reliability of the model.

5.2 Shortcomings

Although this paper systematically studies the stress characteristics of cantilevered stiff steel plate shear walls, there are still some issues that need further discussion. Future research can be conducted in depth from the following aspects:

- (1) Further study the mechanical properties of cantilevered stiff steel plate shear walls under complex loads, especially the combined effects of wind load and earthquake load.
- (2) Explore the application of new high-performance materials in cantilevered stiff steel plate shear walls, such as

high-strength composite materials and ultra-high-strength steel, to further improve structural performance and reduce structural weight.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Kan Zhihao. Seismic performance design and verification of cantilever conversion structure residential buildings. *Sichuan Cement*, 2024(05): 124-128.
- [2] Li Hongqiang, Zhang Mingchao, Li Jie, et al. Design analysis of long cantilever steel structure - taking Zhengzhou International Cultural Exchange Center Hotel as an example. *Group Technology and Production Modernization*, 2024, 41(01): 42 -46.
- [3] Feng Xiaotian, Hao Jiping, Yu Jinguang. Research on the seismic performance of sandwich corrugated CFRP-steel composite plate shear walls. *Journal of Building Structures*, 2024, 45(09): 69-82.
- [4] Cao Zezheng. Research and application of steel plate concrete composite shear wall construction technology. *Construction Mechanization*, 2024, 45(05): 102-103.

SISM: A SELF-INTERACTIVE APPROACH TO WEB NEWS SUMMARIZATION USING DEEP LEARNING

Aleksy Nowaks

Faculty of Mathematics, Informatics and Mechanics, The University of Warsaw, Poland.

Corresponding Email: alexnow8392@uw.edu.pl

Abstract: The exponential growth of online information has intensified the need for efficient web news summarization techniques. This paper introduces the Self-Interactive Summarization Model (SISM), a novel approach that combines advanced deep learning methods with an innovative refined tuning process. SISM employs a two-stage strategy: extensive pre-training on a diverse dataset, followed by a specialized fine-tuning phase. We present a new, carefully curated dataset that reflects the varied nature of web news articles, enabling comprehensive model evaluation. Our experiments demonstrate SISM's superiority over existing state-of-the-art models, with significant improvements in ROUGE scores across multiple test sets. The study highlights the critical role of our refined tuning process in enhancing summarization quality and adaptability to diverse news content. SISM's performance underscores its potential to advance the field of automated web news summarization, offering more accurate and contextually relevant summaries.

Keywords: Web news summarization; Deep learning; Self-Interactive summarization model; Natural language processing

1 INTRODUCTION

In the current era of rapid internet development and explosive growth of textual data, the need for automatic text summarization systems has become increasingly urgent. Automatic summarization systems can independently and quickly determine the main content of a document to generate a summary, providing convenience for rapid information and intelligence acquisition. Extractive summarization refers to generating a summary by extracting sentences from the document that significantly contain the main information [1-2].

In extractive summarization methods, the encoder-decoder structure has been effectively applied [3-4]. Additionally, the attention mechanism in deep learning can help models discover parts containing important document information, which has a significant effect on improving summarization performance [5-6]. However, existing automatic summarization models based on the encoder-decoder structure focus too much on the decoder part [7-9], concentrating on obtaining sentences more relevant to the source document without mining richer document information in the encoder and ignoring the connections between different sentences. Since the association information between sentences plays an important role in selecting diverse information from documents, it is necessary to obtain document information as completely as possible when performing automatic summarization.

Based on these considerations, this paper proposes a Self-Interactive Summarization Model (SISM) with an encoder-decoder structure based on a self-interactive attention mechanism [10] for single-document automatic summarization. The model consists of a sentence encoder based on self-interactive attention mechanism, a document encoder, and a sentence extractor.

- 1) **Sentence encoder:** Utilizes word embeddings to generate vector representations of sentences.
- 2) **Document encoder:** Takes sentence vector representations as input, first extracts the overall document information, then uses a self-interactive attention mechanism to obtain association information between different sentences. The document encoder then reads the document sentences again, combining sentence information with inter-sentence association information to obtain a richer document representation.
- 3) **Sentence extractor:** Determines which sentences should be selected based on sentence information and encoder output containing document feature information.

We conducted automatic evaluations using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) method on the CNN news dataset, where each document includes an article summary written by the author as the reference summary. Experimental results show that the summaries obtained by SISM contain richer information and can better replace the original documents, improving information acquisition efficiency.

2 RELATED WORK

Text summarization has been an active area of research in natural language processing for several decades. Recent years have seen significant advancements, particularly with the application of deep learning techniques. This section provides an overview of relevant work in the field, focusing on extractive summarization methods and the use of attention mechanisms.

2.1 Extractive Summarization

Extractive summarization involves selecting important sentences or phrases from the source document to form a summary. Traditional approaches often relied on statistical methods and hand-crafted features [11]. However, the advent of deep learning has led to more sophisticated models.

Liu et al. [2] introduce a novel web text summarization approach, combining deep learning with a unique refined tuning process. Their results outperform existing methods on various datasets, highlighting the importance of their pre-training and refined tuning strategy.

Nallapati et al. [12] proposed SummaRuNNer, one of the early neural network-based models for extractive summarization. It uses a two-layer recurrent neural network (RNN) to encode sentences and documents, treating sentence extraction as a sequence labeling task. This work demonstrated the potential of neural networks in capturing complex document structures for summarization.

Cheng and Lapata [7] introduced a neural summarization model with an attention-based sentence extractor. Their model uses a convolutional neural network (CNN) to encode sentences and a recurrent neural network to represent documents, showcasing the effectiveness of hierarchical document encoding.

Zhou et al. [13] proposed a neural latent variable model for extractive summarization, which jointly learns sentence scoring and selection. This approach addresses the limitation of previous methods that treated sentence scoring and selection as separate steps.

More recently, Liu and Lapata [14] introduced BERTSUM, which leverages the pre-trained BERT model for extractive summarization. They demonstrated that fine-tuning BERT on summarization tasks can lead to state-of-the-art performance, highlighting the potential of transfer learning in summarization.

2.2 Attention Mechanisms in Summarization

Attention mechanisms have played a crucial role in improving the performance of summarization models by allowing them to focus on the most relevant parts of the input.

Rush et al. [15] were among the first to apply neural attention to abstractive sentence summarization. Although their work focused on abstractive methods, it paved the way for the use of attention in extractive summarization as well.

Nallapati et al. [16] introduced a neural attention-based model for abstractive summarization, which included a novel coarse-to-fine attention mechanism. This work demonstrated the effectiveness of hierarchical attention in capturing document structure.

Fan et al. [17] proposed a controllable abstractive summarization model with structured attention. Their approach allows for fine-grained control over summary attributes such as length and style, showcasing the flexibility that attention mechanisms can provide.

In the context of extractive summarization, Xiao and Carenini [18] introduced an extractive summarization model with a novel attention mechanism that captures both local and global context. Their approach demonstrates the importance of considering different levels of document context in summarization.

2.3 Self-Attention and Transformer-based Models

The introduction of self-attention and Transformer architectures [19] has led to significant advancements in various natural language processing tasks, including summarization.

Zhang et al. [20] proposed HIBERT, a document-level pre-training model based on the Transformer architecture for extractive summarization. Their work demonstrates the effectiveness of hierarchical Transformer encoders in capturing document structure.

Zhong et al. [21] introduced a Transformer-based extractive summarization model that incorporates sentence-level and document-level attention. Their approach achieves strong performance by effectively modeling both local and global document context.

Wang et al. [22] proposed MATINF, a general framework for joint inference of multiple NLP tasks, including extractive summarization. Their model uses a multi-task attention network to capture task-specific and shared information across different NLP tasks.

2.4 Hybrid and Novel Approaches

Recent research has also explored hybrid approaches and novel techniques to improve summarization performance.

Narayan et al. [23] introduced a hybrid extractive-abstractive summarization model that first selects salient sentences and then rewrites them abstractively. This approach combines the strengths of both extractive and abstractive methods.

Xu et al. [24] proposed a discourse-aware neural extractive model that incorporates discourse structure into the summarization process. Their work highlights the importance of considering document-level discourse information in extractive summarization.

Jia et al. [25] introduced a neural reinforcement learning approach for extractive summarization. Their model uses a novel reward function that considers both the informativeness and diversity of selected sentences.

Dong et al. [26] proposed a unified pre-training model for both extractive and abstractive summarization. Their approach demonstrates the potential of multi-task learning in improving summarization performance across different paradigms.

2.5 Evaluation and Datasets

Advances in summarization models have been accompanied by developments in evaluation metrics and datasets.

Narayan et al. [27] introduced the XSum dataset for extreme summarization, which focuses on highly abstractive single-sentence summaries. This dataset has become popular for evaluating both extractive and abstractive summarization models.

Fabrizi et al. [28] created the MNLI dataset for multi-document summarization, addressing the need for large-scale datasets in this domain.

In terms of evaluation, while ROUGE [29] remains widely used, researchers have proposed alternatives to address its limitations. Zhang et al. [30] introduced BERTScore, which uses contextual embeddings to compute similarity between generated and reference summaries. Scialom et al. [31] proposed QUESTEVAL, a reference-free metric for text generation that correlates better with human judgments than traditional metrics.

Our proposed Self-Interactive Summarization Model (SISM) builds upon these advancements, particularly in the use of self-attention mechanisms and hierarchical document encoding. By incorporating a self-interactive attention mechanism in both the sentence and document encoders, SISM aims to capture rich intra-sentence and inter-sentence relationships, addressing limitations of previous models that focused primarily on sentence-level features or used simpler attention mechanisms.

3 SISM MODEL

First, we define the extractive summarization task. Given a document D consisting of a sequence of N sentences (s_1, s_2, \dots, s_N), select a subsequence of n ($n < N$) sentences to form the summary of document D . To achieve this goal, each sentence s_i is scored and labeled with $y_i \in \{0, 1\}$, indicating whether s_i should be considered as a candidate sentence for the summary. During supervised learning, given document D and model parameters θ , the objective is to maximize the probability of each sentence label (y_1, y_2, \dots, y_N):

$$P(y|D; \theta) = \prod_{i=1}^N P(y_i|D; \theta) \quad (1)$$

We propose SISM, which consists of three parts: sentence encoder, document encoder, and sentence extractor.

3.1 Sentence Encoder

The sentence encoder's role is to obtain sentence vectors containing rich word information based on word embeddings. Assuming a sentence s consists of T words (w_1, w_2, \dots, w_T), these words are input into the sentence encoder. To capture the dependency relationships between adjacent words, a bidirectional Long Short-Term Memory (BiLSTM) network is used to process these words:

$$\begin{aligned} h_i^{\rightarrow} &= LSTM^{\rightarrow}(w_i, h_{i-1}^{\rightarrow}) \\ h_i^{\leftarrow} &= LSTM^{\leftarrow}(w_i, h_{i+1}^{\leftarrow}) \end{aligned} \quad (2)$$

To ensure that sentences of different lengths obtain sentence vectors of the same length, we use a self-interactive attention mechanism [10] for linear connection processing of matrix H . Taking H as input, the output vector weights are:

$$a = softmax(v^T tanh(W_1 H^T)) \quad (3)$$

where W_1 is a weight matrix that can be obtained through training, and v is a vector parameter. By summing the columns of matrix H using the weight vector a , we can obtain a vector representation of a sentence. However, this vector representation only contains information from a specific position in the sentence, and different positions in a sentence may contain different types of important information that should not be ignored. Therefore, we need to obtain information representing different regions of the sentence to get a vector representation containing the overall semantic information of the sentence. We extend vector v to matrix V , obtaining the output matrix weights A :

$$A = softmax(V^T tanh(W_1 H^T)) \quad (4)$$

Thus, we can obtain matrix $M = AH$, representing information from different regions of the sentence. Since each column of M contains similar information, we perform max-pooling on M column-wise to obtain the final sentence vector s .

3.2 Document Encoder

The document encoder consists of a 2-layer LSTM. The first layer reads the sentences in the document sequentially, captures their features, and obtains an initial document representation. Then, an attention mechanism is applied to obtain a vector representation containing multiple aspects of document information, including inter-sentence interaction information. The second layer reads the sentence sequence again to reduce information loss and combines the sentence vector representations with the output of the attention mechanism layer to obtain the final document representation.

Given a document $D = (s_1, s_2, \dots, s_N)$, the hidden state update method for the first layer of the encoder at time t is:

$$\begin{aligned} h_{1t}^{\rightarrow} &= LSTM^{\rightarrow}(s_t, h_{1t-1}^{\rightarrow}) \\ h_{1t}^{\leftarrow} &= LSTM^{\leftarrow}(s_t, h_{1t+1}^{\leftarrow}) \end{aligned} \quad (5)$$

At time t , the hidden state h_{1t} not only contains historical information prior to sentence s_t but also information from sentences far from s_t . To better describe the connections between sentences and obtain more useful information, we design a self-interactive attention mechanism to process the output of the first layer of the encoder.

First, we assign different weights to each hidden state of the first layer and sum them:

$$m_{1t} = \sum_{k=1}^N \alpha_{tk} h_{1k} \quad (6)$$

where α_{tk} is the normalized weight for the k -th hidden state at time t :

$$\alpha_{tk} = \frac{\exp(e_{tk})}{\sum_{i=1}^N \exp(e_{ti})} \quad (7)$$

e_{tk} is the initial weight value calculated using only H :

$$e_{tk} = v_{tk}^T \tanh(WH^T) \quad (8)$$

v_{tk} and W are trainable model parameters. Thus, each m_{1t} contains the connection between s_t and other sentences.

At time t , m_{1t} is input into the second layer of the encoder. To reduce information loss as much as possible, the sentence vector is also input and combined with m_{1t} . The update method for the hidden state of the second layer at time t is:

$$h_{2t} = LSTM(s_t || m_{1t}, h_{2t-1}) \quad (9)$$

where $||$ denotes concatenation of s_t and m_{1t} . This yields the vector representation of the document.

3.3 Sentence Extractor The sentence extractor consists of an LSTM network that can detect and compute the salience of each sentence and label it. Given document D and the hidden states of the document encoder ($h_{21}, h_{22}, \dots, h_{2N}$), the extractor makes the following prediction for the label of the i -th sentence:

$$P(y_i | s_i, D) = \text{softmax}(MLP(h_{2i}, h_{di})) \quad (10)$$

where $MLP(h_{2i}, h_{di})$ represents a multi-layer network calculated as:

$$MLP(h_{2i}, h_{di}) = W \tanh(W_1 h_{2i} + W_2 h_{di}) \quad (11)$$

where W_1, W_2 , and W are trainable neural network parameters; h_{di} is the hidden state of the sentence extractor calculated as:

$$h_{dt} = LSTM(s_t, h_{dt-1}) \quad (12)$$

h_{d0} is the last output hidden state h_{2N} of the document encoder. The loss function used during model training is:

$$\text{loss} = - \sum_{i=1}^N \log P(y_i | s_i, D) \quad (13)$$

Finally, the prediction result for whether sentence s_i should be selected as part of the summary is:

$$\hat{y}_i = \operatorname{argmax}_{y_i \in \{0,1\}} P(y_i | s_i, D) \quad (14)$$

4 EXPERIMENTAL VALIDATION

We aimed to investigate the following questions:

- 1) Compared to baseline models, can the attention mechanism in SISM improve the performance of extractive summarization?
- 2) How does the length of the generated summary (i.e., 75B, 275B, and full length (3 sentences)) affect the results?
- 3) We compared SISM with two baseline models that generate summaries by selecting significant sentences from the original document:
- 4) **LEAD**: A standard model that selects the first 3 sentences of the document as the summary [32, 33].
- 5) **NN-SE**: A neural network model for extractive summarization, including a hierarchical document encoder and an attention-based sentence extractor [7].

4.1 Dataset and Experimental Details

We constructed the model training and testing dataset based on CNN news [7]. The CNN news dataset is widely used in automatic question-answering system research. Each document contains the original news text and highlighted text written

by news editors, which can serve as true abstractive summaries and be used as reference summaries. Cheng et al. [7] used a rule-based method to label each sentence in the document with 0 or 1 (1 indicates that the sentence matches the highlighted text; 0 indicates otherwise) with a verified sentence label accuracy of 85%.

The dataset statistics are shown in Figure 1. Since over 95% of sentences in the dataset do not exceed 50 words and over 95% of documents do not exceed 60 sentences, we set the sentence length to 50 and the document length to 60. For the document encoder and sentence extractor, we use LSTM units with a size of 650. The regularization dropout rate used in the LSTM input entering the hidden layer and the sentence scoring process is 0.5. During training, we perform batch training with 20 documents per batch using the Adam optimizer with an initial learning rate of 0.001.

		Number of neurons in the hidden layer					
		5		10		15	
BBO	Average	0.880		Average		0.879	
	Stdv	0.011		Stdv		0.012	
	Best	0.902		Best		0.892	
GA	Average	0.858		Average		0.833	
	Stdv	0.009		Stdv		0.011	
	Best	0.873		Best		0.867	
PSO	Average	0.785		Average		0.785	
	Stdv	0.019		Stdv		0.022	
	Best	0.818		Best		0.812	
DE	Average	0.681		Average		0.693	
	Stdv	0.043		Stdv		0.037	
	Best	0.740		Best		0.741	
ACO	Average	0.643		Average		0.623	
	Stdv	0.043		Stdv		0.030	
	Best	0.743		Best		0.671	
BP	Average	0.658		Average		0.673	
	Stdv	0.054		Stdv		0.060	
	Best	0.747		Best		0.766	

Figure 1 Dataset Statistics

4.2 Evaluation Method

We used ROUGE [29] to evaluate the quality of the summaries generated by the models on the entire CNN test set. ROUGE is a recall-based measurement method. ROUGE-N (N=1, 2, 3, 4) can measure the recall rate of N-grams between candidate summaries and reference summaries, which can be used to measure the amount of information contained in the summaries. ROUGE-L can detect the longest common subsequence, reflecting the readability and fluency of the summaries. We use ROUGE-1 (R-1), ROUGE-2 (R-2), ROUGE-3 (R-3), and ROUGE-4 (R-4) to reflect the information content of the summaries, and ROUGE-L (R-L) to reflect the fluency of the summaries. We provide summaries of full length and fixed lengths (75B and 275B). To ensure fair comparison, we select the 3 highest-scoring sentences as the full-length summary.

5 RESULTS AND ANALYSIS

We validated the models by using ROUGE scores for summaries of 75B, 275B, and full length (3 sentences) generated by SISM and the two baseline models. The ROUGE score comparisons for summaries of different lengths for each model are shown as figure 2.

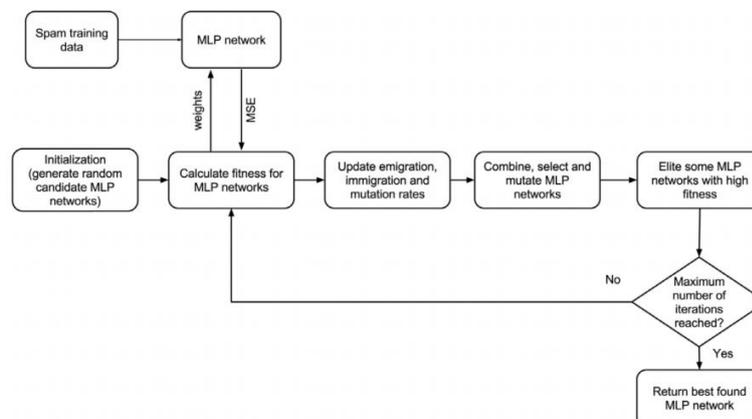


Figure 2 The ROUGE score comparisons for summaries of different lengths for each model

5.1 Summarization Model Performance

To address the first research question, we compared the ROUGE scores of full-length summaries for each model. As seen in Figure 3, among the two baseline models, the NN-SE model's ROUGE-N scores are all higher than those of the LEAD model, but its R-L score is slightly lower than that of the LEAD model. This indicates that the NN-SE model generates summaries with richer information, but the LEAD model produces more fluent and coherent summaries.

SISM outperforms the baseline models in all evaluation metrics with significant improvements. Compared to the NN-SE model, SISM's R-1, R-2, R-3, R-4, and R-L scores improved by 7.4%, 24.3%, 13.4%, 7.1%, and 7.6% respectively, with 4 of these improvements being statistically significant. This demonstrates that using the self-interactive attention mechanism combined with bidirectional LSTM in the sentence encoder and document encoder of SISM helps capture the document's main ideas and select sentences with significant meaning as summaries.

5.2 Performance of Summaries with Different Lengths

To address the second research question, we also compared the ROUGE scores of 75B, 275B, and full-length (3 sentences) summaries generated by the three models. As shown in figure 3, SISM achieves the best results for summaries of all lengths, while the NN-SE model performs slightly better than the LEAD model among the two baseline models.

When generating 75B-length summaries, SISM shows small improvements in all ROUGE scores compared to the two baseline models. For 275B-length summaries, SISM's R-1, R-2, R-3, and R-4 scores improved by 9.3%, 12.5%, 26.4%, and 24.2% respectively compared to the NN-SE model, while its R-L score improved by 4.6% compared to the LEAD model. The improvements in R-1, R-3, and R-L for SISM are statistically significant. Overall, SISM performs better when generating longer summaries.

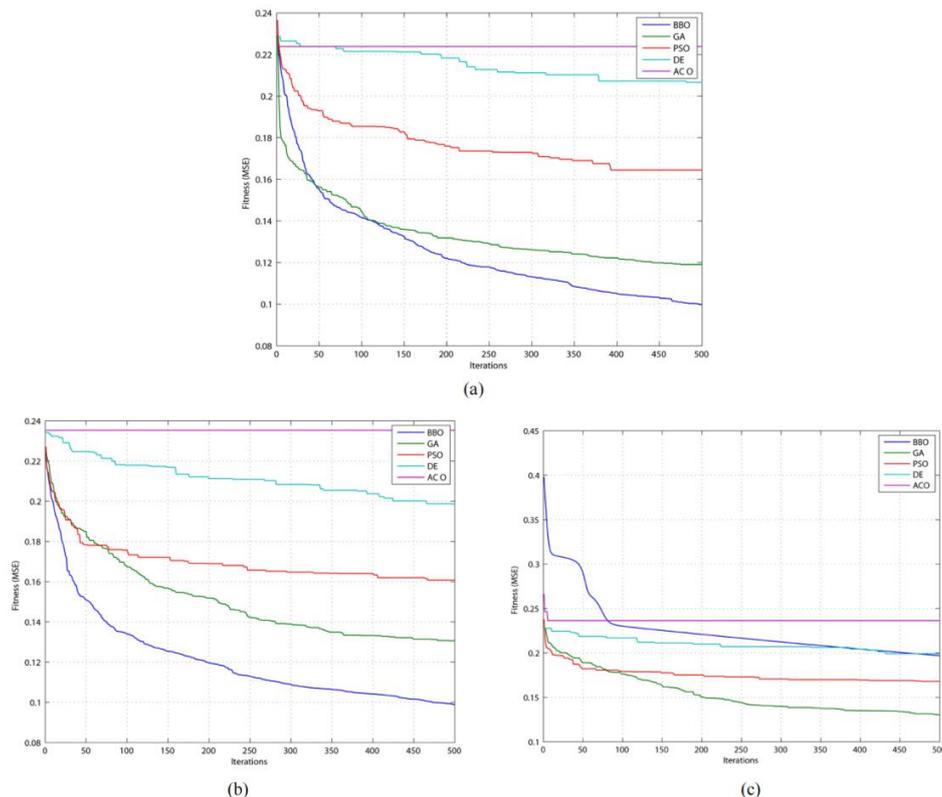


Figure 3 The ROUGE scores of 75B, 275B, and full-length (3 sentences) summaries generated by the three models

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed a text summarization model for automatically generating extractive summaries. The model utilizes an encoder-decoder structure and employs a self-interactive attention mechanism to effectively mine textual information and structural features. It produces summaries with higher ROUGE-N and ROUGE-L scores, indicating that the

model generates summaries with more information content and better fluency. Experimental results show that SISM significantly outperforms the LEAD and NN-SE baseline models, especially when generating longer summaries.

For future work, we plan to test SISM on different datasets to verify its effectiveness across various text domains. We will also leverage other features such as document topics, titles, and inter-paragraph relationships to capture richer and more significant document information, further improving automatic summarization performance.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Nenkova, A., McKeown, K. A survey of text summarization techniques. In *Mining text data*. Springer, Boston, MA. 2013: 43-76.
- [2] Liu, M., Ma, Z., Li, J., Wu, YC., Wang, X. Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. *IEEE Access*, 2024, 12: 92120-92129.
- [3] Sutskever, I., Vinyals, O., Le, QV. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 2024: 3104-3112.
- [4] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. 2014.
- [5] Bahdanau, D., Cho, K., Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*. 2014.
- [6] Luong, M. T., Pham, H., Manning, CD. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*. 2015.
- [7] Cheng, J., Lapata, M. Neural summarization by extracting sentences and words. *arXiv preprint arXiv:1603.07252*. 2016.
- [8] Nallapati, R., Zhai, F., Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [9] See, A., Liu, P. J., Manning, CD. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*. 2017.
- [10] Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., Bengio, Y. A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130*. 2017.
- [11] Erkan, G., Radev, DR. LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 2004, 22: 457-479.
- [12] Nallapati, R., Zhai, F., Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [13] Zhou, Q., Yang, N., Wei, F., Huang, S., Zhou, M., Zhao, T. Neural document summarization by jointly learning to score and select sentences. *arXiv preprint arXiv:1807.02305*. 2018.
- [14] Liu, Y., Lapata, M. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*. 2019.
- [15] Rush, A. M., Chopra, S., Weston, J. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*. 2015.
- [16] Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*. 2016.
- [17] Fan, A., Grangier, D., Auli, M. Controllable abstractive summarization. *arXiv preprint arXiv:1711.05217*. 2018.
- [18] Xiao, W., Carenini, G. Extractive summarization of long documents by combining global and local context. *arXiv preprint arXiv:1909.08089*. 2019.
- [19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*, 2017: 5998-6008.
- [20] Zhang, X., Wei, F., Zhou, M. HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. *arXiv preprint arXiv:1905.06566*. 2019.
- [21] Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., Huang, X. Extractive summarization as text matching. *arXiv preprint arXiv:2004.08795*. 2009.
- [22] Wang, H., Liu, Q., Gao, Z., Nie, Y., Liu, Z., Zheng, H. Neural multi-task learning for aspect-based sentiment analysis. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019: 1-8.
- [23] Narayan, S., Cohen, S. B., Lapata, M. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *arXiv preprint arXiv:1808.08745*. 2018.
- [24] Xu, J., Gan, Z., Cheng, Y., Liu, J. Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 5021-5031.

- [25] Jia, Y., Ye, Y., Feng, Y., Lai, Y., Yan, R., Zhao, D. Modeling discourse cohesion for discourse parsing via memory network. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics 2018, 2: 429-434.
- [26] Dong, Y., Shen, Y., Crawford, E., van Hoof, H., Cheung, JCK. BanditSum: Extractive summarization as a contextual bandit. arXiv preprint arXiv:1809.09672. 2018.
- [27] Narayan, S., Cohen, S. B., Lapata, M. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. arXiv preprint arXiv:1808.08745. 2018.
- [28] Fabbri, A. R., Li, I., She, T., Li, S., Radev, D. R. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. arXiv preprint arXiv:1906.01749. 2019.
- [29] Wang, X., Wu, Y. C., Ji, X., Fu, H. Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. *Frontiers in Artificial Intelligence*, 2024, 7: 1320277.
- [30] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., Artzi, Y. Bertscore: Evaluating text generation with bert. arXiv preprint arXiv:1904.09675. 2019.
- [31] Scialom, T., Dray, P. A., Lamprier, S., Piwowarski, B., Staiano, J. QuestEval: Summarization asks for fact-based evaluation. arXiv preprint arXiv:2103.12693. 2021.
- [32] Wasson, M. Using leading text for news summaries: Evaluation results and implications for commercial summarization applications. In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, 1998: 2.
- [33] Brandow, R., Mitze, K., Rau, LF. Automatic condensation of electronic publications by sentence selection. *Information Processing & Management*, 1995, 31(5): 675-685.
- [34] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
- [35] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. Language models are unsupervised multitask learners. *OpenAI blog*, 2019, 1(8): 9.
- [36] Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, L., Shazeer, N. Generating wikipedia by summarizing long sequences. arXiv preprint arXiv:1801.10198. 2018.
- [37] Chen, X., Liu, M., Niu, Y., Wang, X., Wu, YC. Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. *IEEE Access*, 2024, 12: 78505-78514
- [38] Wang, X., Wu, Y. C., Ma, Z. Blockchain in the courtroom: exploring its evidentiary significance and procedural implications in US judicial processes. *Frontiers in Blockchain*, 2024, 7, 1306058.
- [39] Ma, Z., Chen, X., Sun, T., Wang, X., Wu, Y. C., Zhou, M. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. *Future Internet*, 2024, 16(5): 163.
- [40] Wang, X., Wu, Y. C., Zhou, M., Fu, H. Beyond Surveillance: Privacy, Ethics, and Regulations in Face Recognition Technology. *Frontiers in Big Data*, 2024, 7, 1337465.
- [41] Kedzie, C., McKeown, K., Daume III, H. Content selection in deep learning models of summarization. arXiv preprint arXiv:1810.12343. 2018.

IDENTIFICATION AND TRACKING OF AERIAL UAVS BASED ON DEEP LEARNING VISUAL ALGORITHMS

JianJun Song

Shanghai Technical Institute of Electronics Information, Shanghai 201411, China.

Corresponding Author: JianJun Song, Email: songjianjun151@163.com

Abstract: When conducting long-range wireless charging and monitoring of Unmanned Aerial Vehicles (UAVs) in the air, remote identification and tracking of the drones in the air are required. To address this issue, a deep learning-based algorithm for aerial UAVs identification and tracking is proposed. The YOLO algorithm is utilized for UAVs identification in the air, and the Deep Sort algorithm is used for tracking the identified UAVs. A model and training dataset are constructed, and the deep learning model is trained. The trained model is then invoked to verify the identification and tracking effectiveness of the UAVs in the air.

Keywords: Deep learning visual algorithms; YOLO; Deep SORT; UAVs identification and tracking

1 INTRODUCTION

Long-distance wireless power transmission, applied to Unmanned Aerial Vehicles (UAVs), has the potential to be a breakthrough technology for efficiently charging remote-flying UAVs [1]. Currently, the use of microwave and laser for long-distance wireless charging of UAVs has become feasible [2]. The rapid advancement of technology will gradually enhance the safety and reliability of long-distance wireless charging for UAVs. To achieve long-range wireless charging for UAVs, real-time alignment and tracking of the UAVs must be performed. Therefore, utilizing visual methods for UAVs identification and tracking is an economical and feasible approach.

In recent years, the rapid development of UAVs has raised many security issues. Various Anti-Unmanned Aerial Vehicle Defense Systems (AUDS) have been developed worldwide, including radars, high-energy laser guns, and other anti-drone defense systems. However, these systems are expensive and difficult to be widely accepted in civilian scenarios [3]. The use of visual methods for UAV identification and tracking is evidently more economical and easier to apply.

To address the security issues caused by UAVs, researchers have begun to use deep learning-based image algorithms to identify UAVs, achieving good results [4]. Some researchers have improved the YOLO algorithm to enhance the identification of UAVs [5, 6]. In addition to using deep learning algorithms for aerial UAV identification, researchers have also used deep learning algorithms to predict the trajectories of UAVs [7].

Moreover, the hierarchical image-based target tracking of UAVs is also an important task. Previous researchers have conducted a significant amount of research and application on common multi-target tracking algorithms in various fields, such as droplet identification and tracking in microfluidic control and object recognition and tracking in logistics warehouses, achieving certain results [8-11]. Kumar S et al. utilized a deep learning algorithm for vehicle detection and tracking [12]. Razzok M et al. based on deep learning image algorithms for pedestrian detection and tracking [13]. Some researchers have attempted to use deep learning-based methods for the identification and tracking of UAVs [14], but overall, research and application in the identification and tracking of aerial UAVs are relatively limited.

This study proposes a deep learning-based method for the identification and tracking of aerial UAVs. The YOLO algorithm will be used to identify aerial UAVs, and the Deep SORT algorithm will be used to track the identified UAVs. A recognition and tracking model will be constructed, and a training dataset will be established to train the deep learning model. The trained model will be invoked to validate the effectiveness of aerial UAV identification and tracking.

2 METHODOLOGY

2.1 Method Overview

A method for the identification and tracking of UAVs is proposed based on the combination of YOLO algorithm and Deep SORT algorithm, as illustrated in Figure 1. The method mainly includes model construction, training deployment and invocation, image acquisition, image preprocessing, image distortion correction, recognition, and tracking.

Model Construction: A model combining the YOLO algorithm and Deep SORT algorithm is built for object detection and tracking of UAVs.

Training, Deployment, and Invocation: The constructed algorithms are trained and subsequently deployed for the recognition and tracking of aerial UAVs.

Image Acquisition: UAV images are captured using a camera for further processing.

Image Preprocessing: Preprocessing of the captured images includes tasks such as scaling, smoothing, and conversion to grayscale.

Image Distortion Correction: Image distortion correction aims to eliminate distortions that may occur during image capture, enhancing accuracy in presenting the original scene.

UAV Recognition and Tracking: Utilizing the constructed and trained deep learning model, images or videos are processed to identify UAVs, followed by visual tracking and analysis using deep learning tracking algorithms.

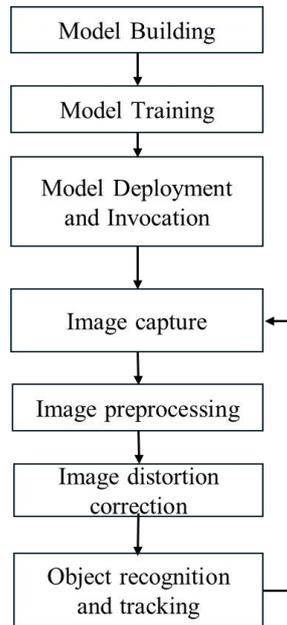


Figure 1 Method Overview

2.1.1 Experimental platform

An experimental setup, as illustrated in Figure 2, is constructed for the purpose of capturing images of aerial UAVs. This experimental platform primarily consists of a host computer, a camera, and a support frame. The system is designed to perform real-time, multi-angle imaging and processing of UAV images. The camera is responsible for capturing images of the UAVs, while the host computer runs the deep learning visual model and handles the real-time processing of the images collected by the camera.

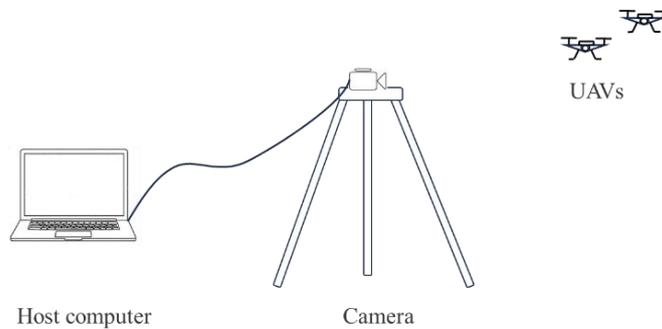


Figure 2 Experimental Setup

2.1.2 Principle of YOLO + deep SORT algorithm

Combining the YOLO and Deep SORT algorithms enables the accomplishment of multi-object detection and tracking tasks. As shown in Figure 3, initially, the YOLO algorithm performs object detection, identifying the positions and class information of target objects in each frame. Subsequently, the Deep SORT algorithm utilizes these detection results for data association and trajectory updating, thereby achieving multi-object tracking.

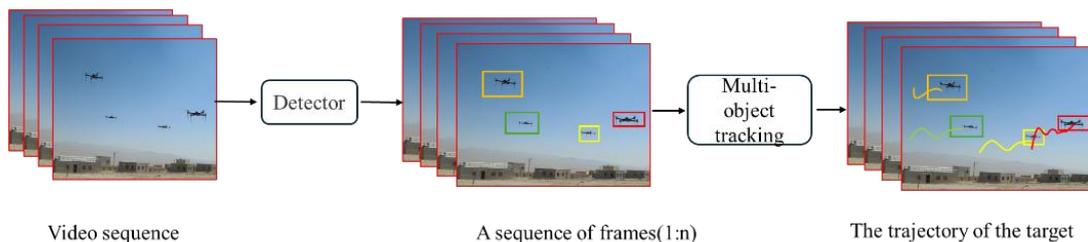


Figure 3 The Schematic Diagram of the Algorithm Principle

The tracking process of the SORT algorithm is as follows:

- (1) Utilize the YOLO algorithm for target detection in the video sequence to obtain the position and size information of target objects in each frame.
- (2) Extract features from each target object in every frame to obtain its deep feature vector.
- (3) Initialize a tracking list to store information about currently tracked target objects.
- (4) For each target object in every frame, compute its Mahalanobis distance and Hungarian matching cost with target objects in the tracking list to find the best match.
- (5) If a target object successfully matches with a target object in the tracking list, update its position and velocity information.
- (6) If a target object fails to match with any target object in the tracking list, add it to the tracking list and initialize its position and velocity information.
- (7) Repeat the above steps until the entire video sequence is processed.

The process of the Deep SORT algorithm is more complex compared to the SORT algorithm, incorporating appearance feature extraction and cascade matching.

2.1.3 Method process

Just like Figure 4, integration of YOLO and Deep SORT algorithms for UAV recognition and tracking applications:

- (1) Network Model Construction: Develop a deep learning model combining YOLO and Deep SORT.
- (2) Dataset Construction: Build a UAV dataset for training the YOLO algorithm for target recognition. Construct a REID dataset for training the Deep SORT tracking algorithm to track UAVs.
- (3) YOLO Model Training: Train the YOLO model for UAV recognition.
- (4) YOLO Model Invocation and Evaluation: Invoke the trained YOLO model for practical use, evaluation, and optimization.
- (5) Deep SORT Model Training: Train the Deep SORT algorithm model for multi-object tracking of UAVs.
- (6) Deep SORT Model Invocation and Evaluation: Invoke the trained Deep SORT model for practical use, evaluation, and optimization.
- (7) Overall Model Invocation: Invoke the models for UAV recognition and multi-object tracking in a comprehensive manner.

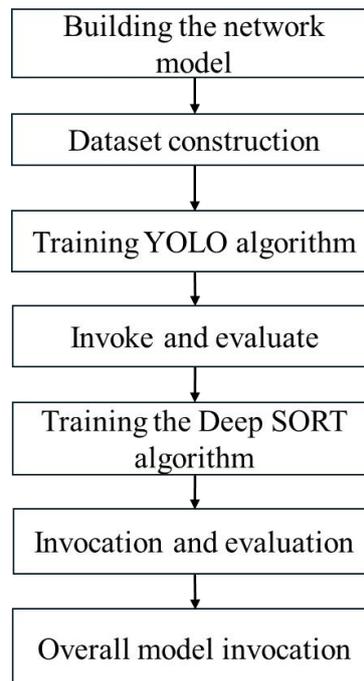


Figure 4 Method Process

2.2 UAV Recognition Based on YOLO Algorithm

2.2.1 Object recognition and performance metrics

Object detection refers to the process of using computer vision techniques to identify and localize specific objects or targets in images or videos. The following are performance metrics for object detection algorithms:

Precision: Precision evaluates how accurate the predictions are in terms of correctness:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{1}$$

Recall Rate, which evaluates how comprehensive the search is in terms of completeness:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

Intersection over Union (IOU), as shown in Figure 5, is a metric used to evaluate the overlap between predicted and ground truth bounding boxes. If the IOU is greater than a threshold, the prediction is considered positive. The calculation method is as follows:

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (3)$$

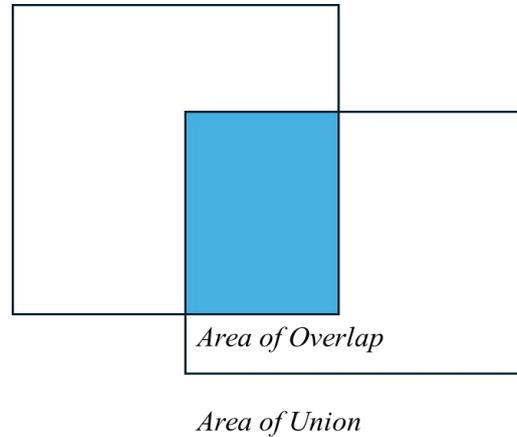


Figure 5 The Schematic Diagram of the IOU

Average Precision (AP) is a measure of how well a learned model performs on each category.

Mean Average Precision (MAP) measures the overall performance of the learned model across all categories, calculated by taking the average of AP values for all categories.

Frames Per Second (FPS) refers to the number of image frames processed per second, used to evaluate the speed and efficiency of computer vision systems in handling images or videos.

2.2.2 YOLO algorithm principle

YOLO stands as a paradigmatic example of a single-stage object detection algorithm. Figure 6 provides a schematic depiction of the YOLO algorithm's core mechanics. The process can be broadly demarcated into three principal stages:

- (1) Adjustment of the input image to a standardized size.
- (2) Acquisition of feature maps through the utilization of a convolutional neural network.
- (3) Application of Non-Maximum Suppression (NMS) to the entire set of predicted results, followed by the outputting of the final outcomes.

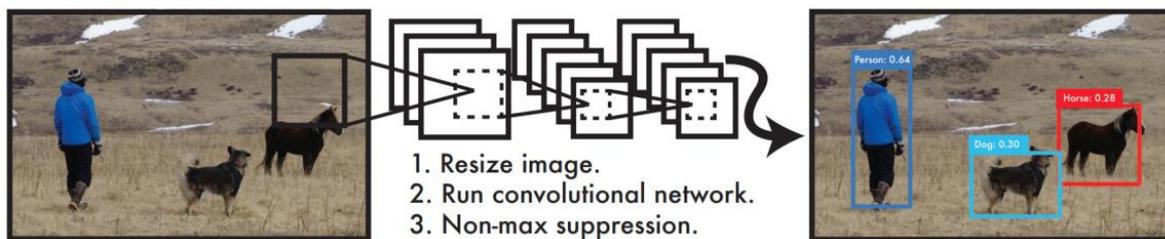


Figure 6 Schematic Diagram of the YOLO Algorithm [15]

Using the YOLOv5 algorithm as an example, the network structure is explained, which mainly consists of three parts: Backbone, Neck, and Head.

Backbone: A convolutional neural network that aggregates and forms image features at different levels of granularity. YOLOv5 uses CSPDarknet53 as its Backbone, which, while maintaining high accuracy, improves model inference speed through enhanced network structure.

Neck: A series of network layers that mix and combine image features and pass them to the prediction layer. YOLOv5 uses Path Aggregation Network (PANet) as its Neck, achieving the fusion and utilization of multi-scale features through top-down path enhancement and bottom-up feature fusion.

Head: Predicts image features, generates bounding boxes, and predicts categories. YOLOv5 adopts the Anchor-Free approach from YOLOv3, optimizing network parameters by calculating the loss between predicted boxes and ground truth bounding-boxes to achieve accurate object detection.

Figure 7 depicts the procedure of the YOLO algorithm, which consists of the following steps:

- (1) The image is divided into an $S \times S$ grid.
- (2) Each grid cell predicts M bounding boxes, with the prediction information for each bounding box including its position, confidence level, the probability of an object's presence, and the probability of belonging to a specific class.

(3) Non-Maxima Suppression (NMS) is conducted, utilizing the predicted probabilities of objects and IOU values to filter out the necessary bounding boxes for output.

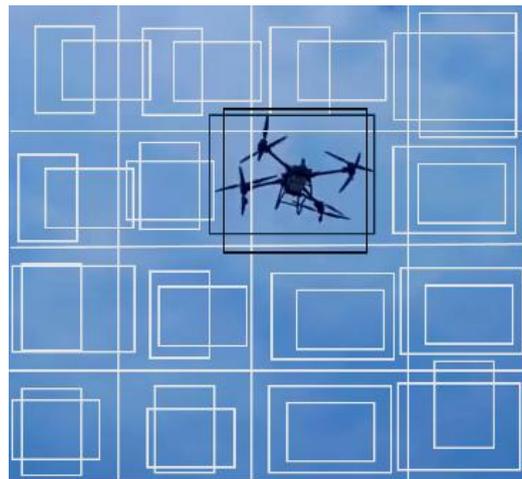


Figure 7 Schematic Diagram for UAVs Forecasting with YOLO Algorithm

2.3 Multi-Target Tracking of UAVs Based on the Deep SORT Algorithm

2.3.1 Multiple object tracking and evaluation

The Multiple Object Tracking (MOT) refers to the process of locating each target in every frame of an image and tracing their trajectories. An input video sequence yields the trajectories of targets and a unique ID for each target, with each target represented by a bounding box.

The evaluation criteria for MOT are primarily categorized into the following four types:

(1) Accuracy: This metric measures the accuracy of the target tracking performed by the algorithm. The ID Switches (IDSW) metric quantifies the number of times the MOT algorithm switches between different objects. The Multiple Object Tracking Accuracy (MOTA) metric calculates the false positive rate, false negative rate, and unmatched detection rate, combining these into a numerical value that relatively fairly reflects the overall tracking performance. Despite its limitations, this is the most widely accepted MOT evaluation metric to date.

(2) Precision: Metrics such as the Multiple Object Tracking Precision (MOTP), Tracking Distance Error (TDE), and Optimal Sub-Pattern Assignment (OSPA) are widely used as critical evaluation criteria. These metrics consider both the overlap of bounding boxes and distance measurements, describing the degree to which objects are accurately tracked. Additionally, they consider false alarms and label errors, providing a comprehensive perspective on the performance evaluation of multi-target tracking systems.

(3) Completeness: Completeness metrics indicate the extent to which the ground truth trajectories are tracked. This includes measures such as Mostly Tracked (MT), Partially Tracked (PT), Mostly Lost (ML), and Fragmentation (FM), which belong to this category.

(4) Robustness: To assess the MOT algorithm's capability to recover from occlusions, metrics referred to as Recovery from Short-Term Occlusions (RS) and Recovery from Long-Term Occlusions (RL) have been introduced.

2.3.2 Principles of deep SORT

The Deep SORT algorithm employs a target detector to detect targets, as shown in Figure 8, and associates the detected target with the Kalman filter-predicted trajectory using association metric. The so-called association metric considers appearance features and Mahalanobis distance, conducts feature extraction and similarity estimation, constructs a cost matrix using Mahalanobis distance and appearance features, and performs matching using the Hungarian algorithm.

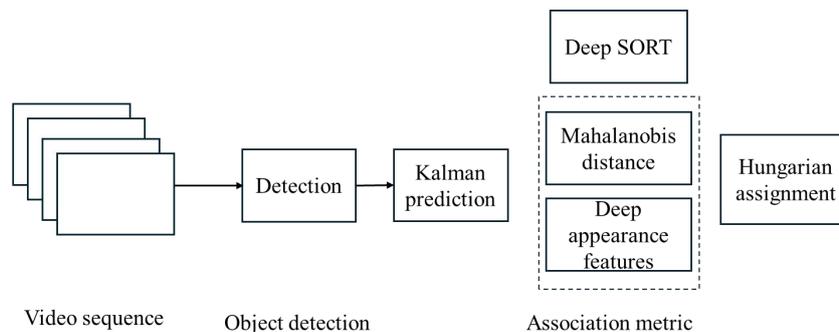


Figure 8 Schematic Diagram of the Deep SORT Algorithm Principle

The Kalman filter utilizes a linear system state equation to perform optimal estimation on the system state through observations of the system's inputs and outputs. The recursive optimal estimation theory of the Kalman filter, adopting the state-space representation method, is capable of handling multi-dimensional and non-stationary stochastic processes. The Kalman filter predicts and updates the state from one frame to the next based on the previous frame's state and uses the measurements from the current frame to update this prediction. The Kalman filter can be used to estimate the motion trajectory of targets, which can enhance the accuracy of tracking, especially in scenarios where the target moves rapidly or is obscured.

The Mahalanobis distance is a method for measuring the distance between multidimensional data points, considering the covariance of the data. Unlike Euclidean distance, Mahalanobis distance effectively handles variables with different scales and correlations. The square of the Mahalanobis distance should conform to a chi-square distribution with p degrees of freedom.

The assignment problem deals with assigning machines to tasks, workers to jobs, and so on. The goal is to determine the optimum assignment that, for example, minimizes the total cost. The Hungarian algorithm is a combinatorial optimization algorithm that solves the assignment problem in polynomial time.

3 CASE STUDY

In the context of long-range, wireless aerial charging for UAVs, accurate target and tracking of airborne UAVs are essential. This study employs a combined method using YOLO and Deep SORT for the detection and tracking of aerial multi-rotor drones. A network model was constructed, a dataset was prepared, and the model was trained and deployed. Figure 9 illustrates the scenario of target detection and tracking for UAVs.

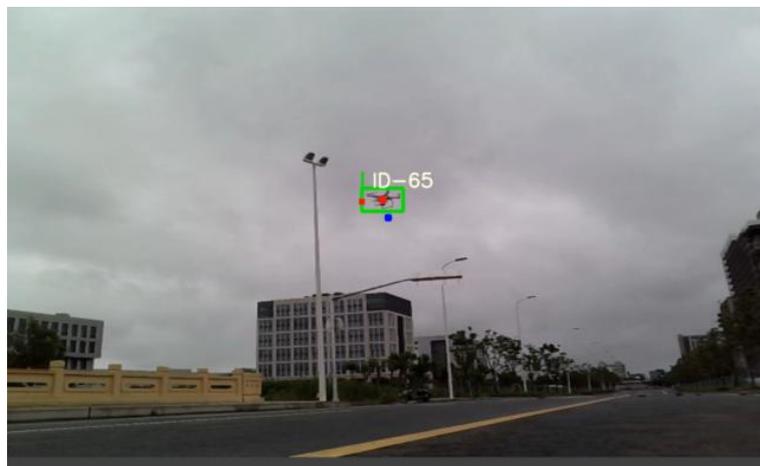


Figure 9 Identification and Tracking of UAVs

3.1 Dataset Creation

3.1.1 Algorithm dataset identification

10,000 images of UAVs were collected to establish the dataset, as shown in Figure 10. These images were annotated to mark the positions of UAVs and label their categories. Figure 11 illustrates the use of the Label-image tool for annotating UAV images.

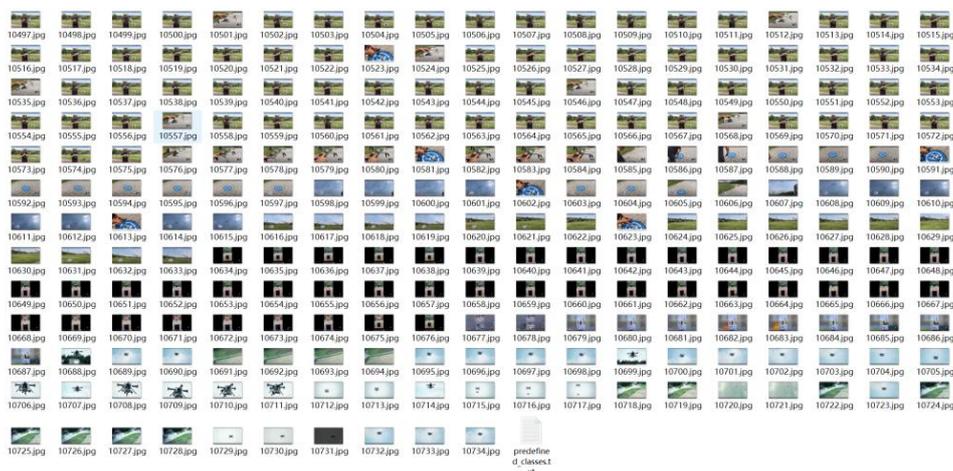


Figure 10 UAVs Dataset



Figure 11 Labeling UAVs

3.1.2 REID dataset

Figure 12 presents the UAV REID dataset, using 0009_c001_00016401_0_c01_t4.jpg as an example to elucidate the image naming convention:

0009 represents the unique identifier for each UAV, ranging from 0001 to 0020, indicating a total of 20 UAVs. c001 denotes the first camera (where “c” stands for Camera), with a total of N cameras. 00016401 indicates the 16401st frame captured by camera c1, with a video frame rate (FPS) of 25.



Figure 12 UAVs REID Dataset

3.2 Training and Analysis of YOLO Algorithm

3.2.1 Confusion matrix

The confusion matrix is an $n \times n$ matrix, where n represents the number of classes. Each row represents the actual class, and each column represents the predicted class by the model. The size of the confusion matrix depends on the number of classes involved, making it applicable to various classification tasks, whether binary or multiclass.

Figure 13 depicts the confusion matrix for drone detection. From the figure, it can be observed that when the prediction is for a UAVs, the actual probability of being a drone is 0.96; when predicted as background, the actual probability of being UAV is 0.04. This indicates that the model exhibits high accuracy in UAVs detection.

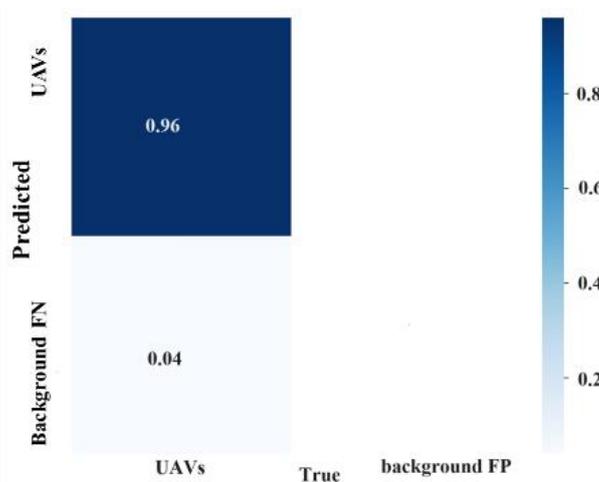


Figure 13 Confusion Matrix for UAVs Detection

3.2.2 FI-Score

The F1 score is the harmonic mean of Precision and Recall, simultaneously considering the accuracy and coverage of the model's predictions. The F1 score ranges from 0 to 1, where 1 indicates that the model's Precision and Recall are both perfect, while 0 suggests that either the Precision or Recall is extremely poor.

From Figure 14, it can be observed that when the confidence is 0.6, the F1-score for drones is 0.84. This indicates that the model performs well in predicting the "UAV" class, demonstrating high precision and recall.

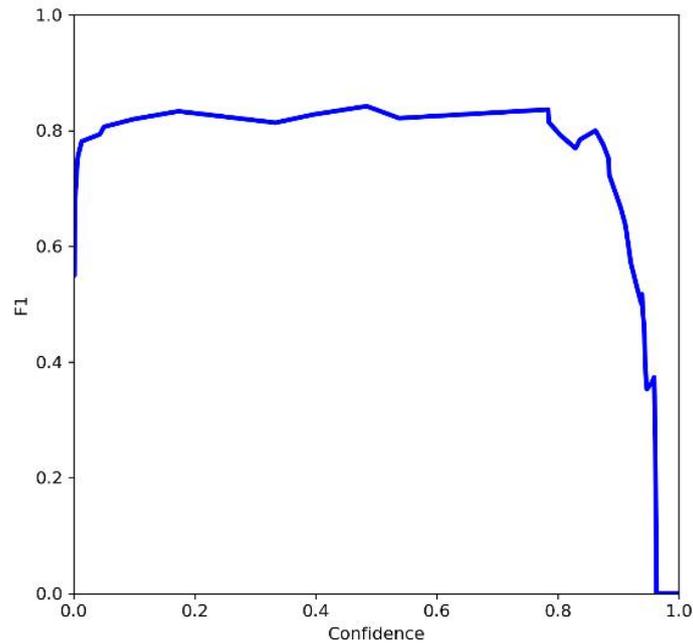


Figure 14 The F1 Score for UAV Detection

3.2.3 Precision

The blue line in the figure 15 represents the "UAV" category. The overall trend shows an increase in precision as confidence increases. Precision reaches 1 at a confidence level of 0.95.

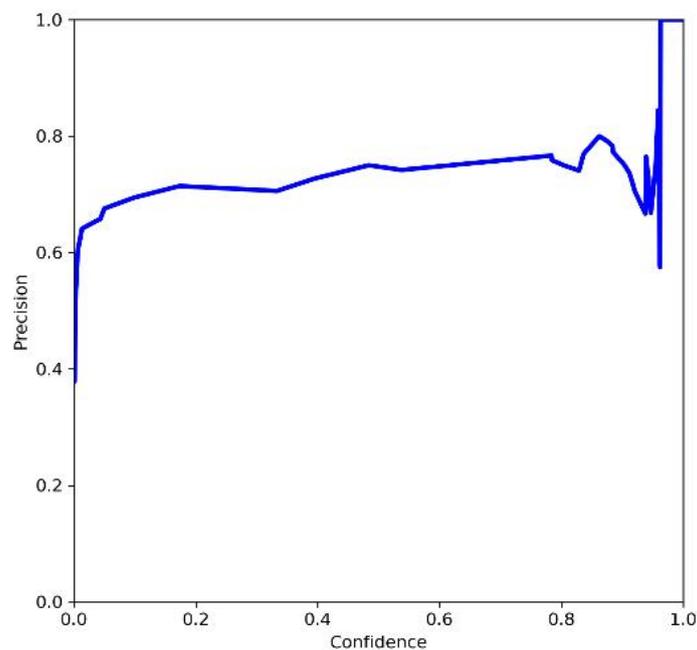


Figure 15 Precision Curve

3.2.4 Recall curve

Figure 16 displays the Recall-Confidence Curve (R curve), illustrating recall at various confidence levels. Specifically, the horizontal axis represents confidence ranging from 0.0 to 1.0, while the vertical axis represents recall, also ranging from 0.0 to 1.0.

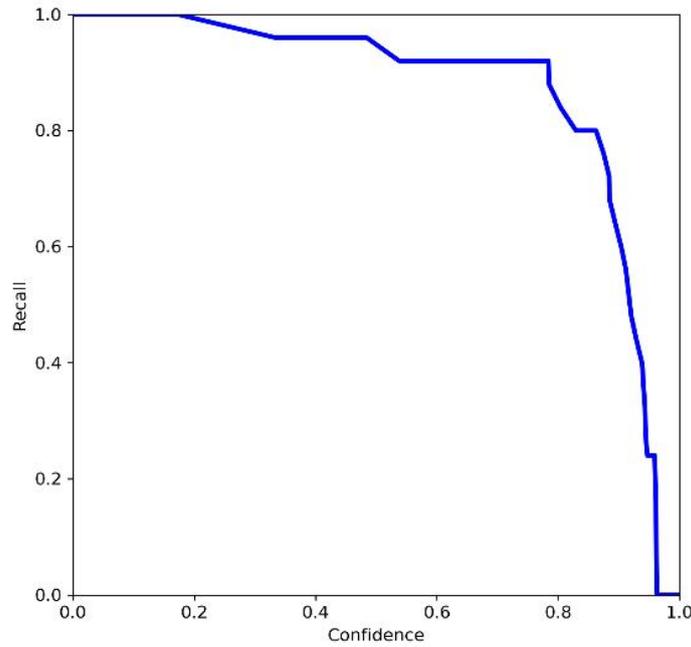


Figure 16 Recall-Confidence Curve

This curve indicates that the recall for the “UAV” category initially remains stable with, followed by a rapid decline after a certain point.

3.2.5 Precision-recall curves

Figure 17 presents the Precision-Recall Curve, the trade-off between precision and recall achieved by the at different threshold settings, the x-axis Recall (ranging from 0.0 to 1.0), while the y-axis Precision (ranging from 0.0 to 1.0). The Precision-Recall Curve for the “UAV” category exhibits a stable precision around 0.8.

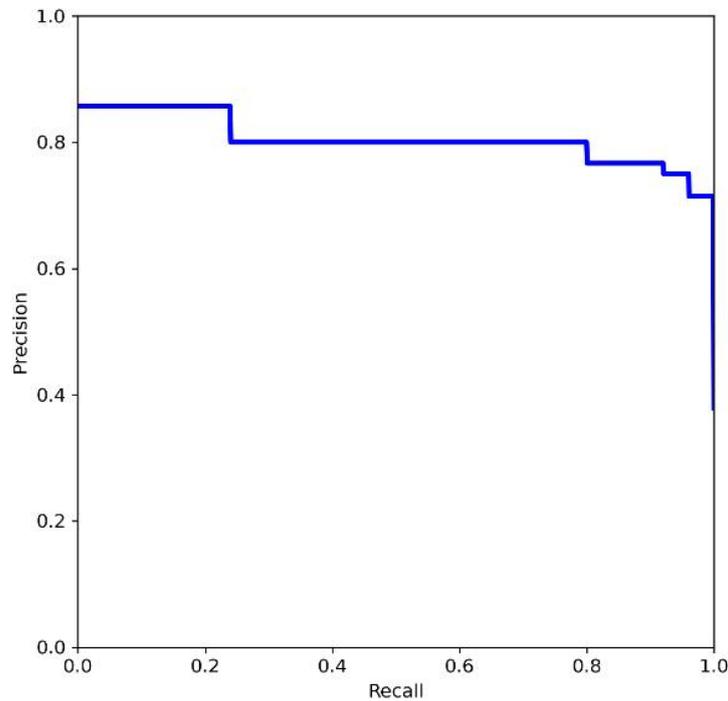


Figure 17 Precision-Recall Curve

3.2.6 Loss function

Figure 18 illustrates the loss during the training process of the YOLOv5 algorithm. The trends of different loss categories are shown in the figure:

(1) Localization loss (Box Loss): YOLOv5 utilizes generalized intersection over union (GIOU) loss as the loss function for bounding boxes. The box loss is inferred as the mean GIOU loss of the predicted boxes, and a smaller box loss value indicates more accurate box positioning.

- (2) Confidence loss (Obj Loss): It is presumed to be the mean loss for object detection, where a smaller obj loss value indicates more accurate object detection.
- (3) Validation set localization loss (Val Box Loss): This represents the bounding Box Loss on the validation set.
- (4) Validation set confidence loss (Val Obj Loss): This represents the mean Object loss on the validation set.

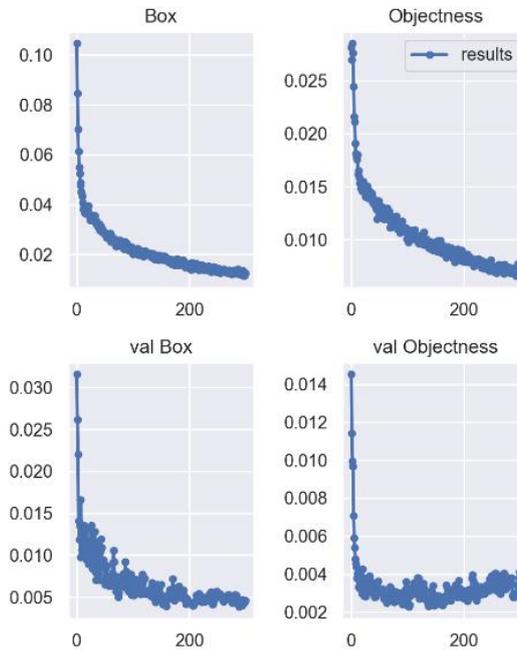


Figure 18 The Loss during the Training Process

3.2.7 Performance on the test set

Figure 19 illustrates the recognition performance of UAVs on the test dataset. The results indicate that the trained algorithm demonstrates good recognition capability, effectively identifying UAVs of various types, sizes, and within different background environments.



Figure 19 Performance of the Recognition Algorithm

3.3 Tracking Algorithm Training and Analysis

The appearance recognition network of Deep SORT was trained for 300 epochs. The value of loss function gradually decreased over the training period, as depicted in Figure 20. As training progressed, its accuracy steadily improved.

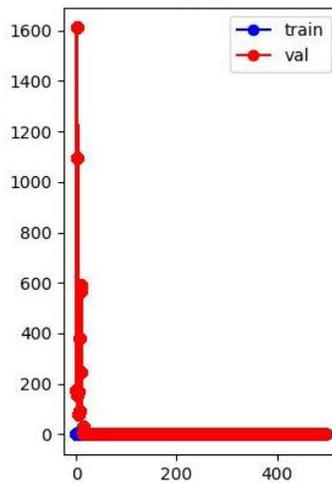


Figure 20 Loss of the Appearance Recognition Network

3.4 Experimental Results

Recognition and tracking were conducted on test videos, successfully identifying UAVs within the video sequences. The model effectively tracked the UAVs visually throughout the video. Figure 21 presents a diagram of the recognition and tracking results.



Figure 21 UAV Recognition and Tracking Results

3.5 Summary

A UAV recognition and tracking method based on the combination of the YOLO algorithm and Deep SORT algorithm is proposed. A UAV recognition and tracking model integrating the YOLO algorithm and Deep SORT algorithm is constructed. Then, a dataset specifically for training the drone recognition and tracking algorithms is curated. The constructed model undergoes training, and the training results are analyzed. Subsequently, the trained model is deployed, effectively accomplishing recognition and tracking tasks with good real-time processing performance.

4 CONCLUSION

This study proposes a deep learning-based method for aerial UAVs recognition and tracking. The YOLO algorithm is employed for UAVs detection in the air, while the Deep SORT algorithm is used to track the identified UAVs. The model is constructed, the training dataset is prepared, and the constructed deep learning model is subsequently trained. The trained models are then utilized to validate the effectiveness of UAVs recognition and tracking in aerial scenarios. The trained models successfully accomplish the tasks of recognition and tracking, demonstrating excellent real-time processing capabilities.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The author received support from Association of Fundamental Computing Education in Chinese Universities for the research. The project is titled “Research on Wildlife Conservation Applications Based on Deep Neural Network Models”, with the project number 2024-AFCEC-103.

REFERENCES

- [1] Naing KM, Zakeri A, Iliev O. Wireless energy transfer to long distance flying intelligent Unmanned Aerial Vehicles (UAVs) using reactive power transfer techniques. *Technol*, 2020, 7(9), 2458-9403.
- [2] Chittoor PK, Chokkalingam B, Mihet-Popa L. A review on UAV wireless charging: Fundamentals, applications, charging techniques and standards. *IEEE access*, 2021, 9, 69235-69266.
- [3] Sheu BH, Chiu CC, Lu WT, Lien CC, Liu TK, Chen WP. Dual-axis rotary platform with UAV image recognition and tracking. *Microelectronics Reliability*, 2019, 95, 8-17.
- [4] Samadzadegan F, Dadrass Javan F, Ashtari Mahini F, Gholamshahi M. Detection and recognition of drones based on a deep convolutional neural network using visible imagery. *Aerospace*, 2022, 9(1), 31.
- [5] Dadrass Javan F, Samadzadegan F, Gholamshahi M, Ashatari Mahini F. A modified YOLOv4 Deep Learning Network for vision-based UAV recognition. *Drones*, 2022, 6(7), 160.
- [6] Sun H, Wang X, Cao Z, Bai F, Wang X, Hao Y, Wang J. Deep learning-based counter-UAV early warning neural network. *Science Technology and Engineering*, 2021, 21(22), 9461-9469.
- [7] Shi Y, Zhu J, Ling Z. Research on UAV detection algorithm based on feature-enhanced YOLOv4. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(7), 16-23.
- [8] Durve M, Orsini S, Tiribocchi A, Montessori A, Tucny JM, Lauricella M, ... Succi S. Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications. *The European Physical Journal E*, 2023, 46(5), 32.
- [9] Xie T, Yao X. Smart logistics warehouse moving-object tracking based on yolov5 and deepsort. *Applied Sciences*, 2023, 13(17), 9895.
- [10] Bai T. Multiple Object Tracking Based on YOLOv5 and Optimized DeepSORT Algorithm. In *Journal of Physics: Conference Series*. IOP Publishing, 2023, 2547,012022.
- [11] Huang Y, Xiao D, Liu J, Tan Z, Liu K, Chen M. An improved pig counting algorithm based on YOLOv5 and DeepSORT model. *Sensors*, 2023, 23(14), 6309.
- [12] Kumar S, Singh SK, Varshney S, Singh S, Kumar P, Kim BG, Ra IH. Fusion of deep sort and Yolov5 for effective vehicle detection and tracking scheme in real-time traffic management sustainable system. *Sustainability*, 2023, 15(24), 16869.
- [13] Razzok M, Badri A, El Mourabit I, Ruichek Y, Sahel A. Pedestrian detection and tracking system based on DeepSORT, YOLOv5, and new data association metrics. *Information*, 2023, 14(4), 218.
- [14] Ma L, Meng D, Zhao S, An B. Visual localization with a monocular camera for unmanned aerial vehicle based on landmark detection and tracking using YOLOv5 and DeepSORT. *International Journal of Advanced Robotic Systems*, 2023, 20(3), 17298806231164831.
- [15] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 779-788.

RESEARCH AND APPLICATION OF CONSTRUCTION TECHNOLOGY OF GROUND STEEL STRUCTURE OF DOUBLE-LAYER EXHIBITION HALL

Yi Zhu^{1,2}, YueQiang Cao^{1,2*}, ShenShen Zhu^{1,2}, ZhiCheng Bai^{1,2}, XiaoXia Zhao^{1,2}, ZhenJie Tan^{1,2}

¹China Construction Fourth Engineering Bureau Co., Ltd. Guangzhou 510000, Guangdong, China.

²China Construction Fourth Engineering Bureau Sixth Construction Co., Ltd. Hefei 230000, Anhui, China.

Corresponding Author: YueQiang Cao, Email: 19822660870@163.com

Abstract: This study proposes an innovative construction technology of ground steel structure of double-layer exhibition hall, aiming to solve the problems of top plate cracks, steel structure installation risks and costs during construction. This technology effectively reduces the pressure of basement beams and slabs by setting a conversion steel platform on the top plate of the basement, combining intelligent hoisting and robot welding, improving construction efficiency and quality, and reducing labor costs. Through practical application in the Hangzhou Convention and Exhibition Center project, the new construction technology adopted reduces construction costs, improves construction efficiency, shortens construction period, and meets the national green building promotion requirements. In the future, this technology can be promoted and applied in a wider range of construction projects.

Keywords: Double-layer exhibition hall; Steel structure construction; Intelligent hoisting; Green building

INTRODUCTION

With the rapid development of my country's economy and the deepening of international exchanges, exhibition projects are increasing. In order to meet the needs of large-span and large-space exhibitions, exhibition and museum buildings usually adopt large-span steel structure design. The underground part is generally a reinforced concrete frame structure, and the upper part is a steel structure exhibition hall [1-3]. In order to solve the problems of top plate cracks, steel structure installation risks and engineering costs in the construction of double-layer large-span exhibition halls, this paper proposes an innovative double-layer exhibition hall ground steel structure construction technology. It is expected to be further promoted and applied in exhibition and museum buildings.

1 PROJECT OVERVIEW

Just like Figure 1, the first phase of the Hangzhou Convention and Exhibition Center project covers a total area of approximately 353,000 square meters and a total construction area of 643,200 square meters. This project has 8 exhibition halls, 1 underground floor and 1 above-ground floor (partially 2 floors); two login halls, 1 underground floor and 4 above-ground floors; middle corridor: 1 above-ground floor and partially 2 floors. The basement is a reinforced concrete + steel column structure, and the above-ground structure is mainly a steel frame + large-span truss steel roof combination structure. The steel structure includes 6 standard single-story exhibition halls and 2 double-story exhibition halls.



Figure 1 Hangzhou Conference Display Intent

2 PROCESS PRINCIPLE

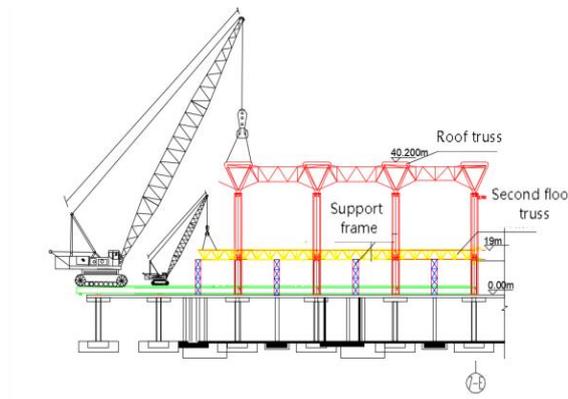


Figure 2 Schematic diagram of the construction of the above-ground steel structure of the double-story exhibition hall

In Figure 2, a steel platform is set up in the air on the top plate of the basement. The platform span is the same as the spacing between the lower steel-concrete columns, and a $2\text{m}\times 5\text{m}$ roadbed box is laid on the top. The most unfavorable working condition disturbance of the crane when driving and hoisting on the steel platform is calculated by finite element software, and a steel plate with a thickness greater than the maximum deflection of the steel platform beam is laid between the lower part of the steel platform and the steel-concrete column to achieve the padding of the concrete floor and beam. The mechanical dead load and live load are transmitted to the steel platform through the roadbed box, and then to the steel plate from the steel platform, and finally to the steel-concrete column and pile foundation, ensuring that the basement top plate is not stressed, thereby ensuring the safety of the underground structure. The ground steel structure hoisting adopts a large crawler crane for three-dimensional segmented high-precision intelligent construction. The overall construction sequence is frame structure, second-floor truss, roof truss lifting and cantilever truss, and welding is performed by welding robots.

3 KEY POINTS OF CONSTRUCTION TECHNOLOGY

3.1 Overlap Conversion Steel Platform

In order to economically and reasonably determine the specifications and dimensions of the main beam and secondary beam of the steel platform, and determine the thickness of the steel pad between the steel-concrete column and the steel platform, a force analysis of the 200t crawler crane heavy-load conversion steel platform is carried out.

200t crawler crane: The main steel beam model specification is $\square 400\times 400\times 20\text{mm}$, the secondary steel beam model specification is $\square 400\times 400\times 12\text{mm}$, the steel pad thickness is 30mm, the secondary steel beam spacing is 1.4m, and the calculation selects two spans with a column spacing of $9\text{m}\times 9\text{m}$ for force analysis. As shown in Figure 3-6.

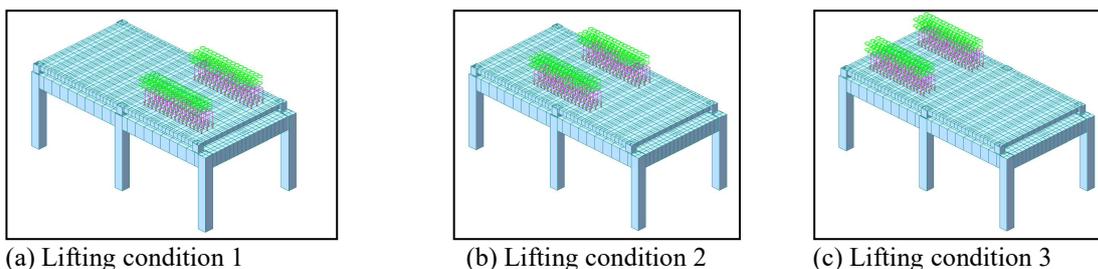


Figure 3 Simulation of Working Conditions under the Action of 200t Crawler Crane

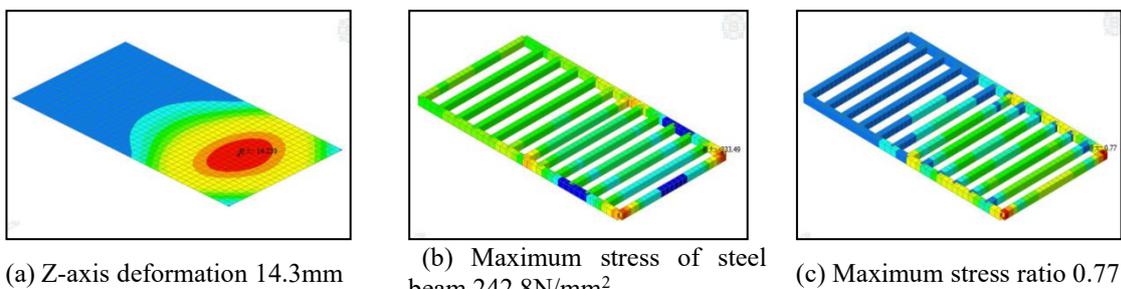


Figure 4 Force Analysis of Lifting Condition 1

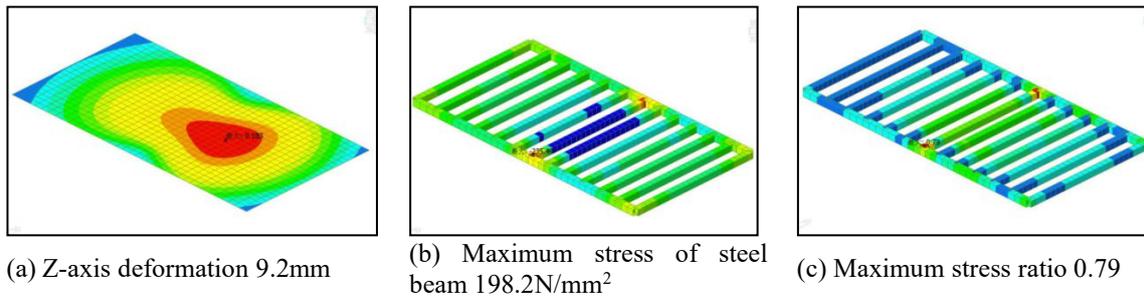


Figure 5 Force Analysis of Lifting Condition 2

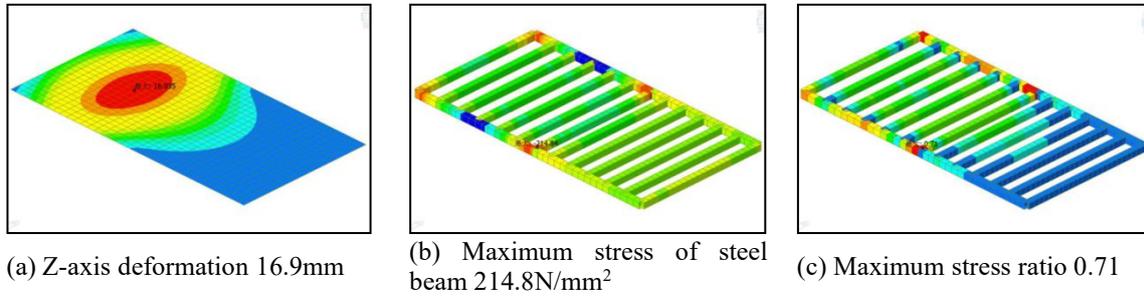


Figure 6 Force Analysis of Lifting Condition 3

After calculation, considering only the action of 200t crawler crane, the maximum z-axis deformation of 9m×9m column spacing 1 span floor reinforcement 16.9mm, the maximum stress of the platform beam is 242.8N/mm², the maximum stress ratio is 0.79, and the final load is transmitted from the platform to the concrete column.

During the design deepening and processing of the steel platform, the design drawings and detailed node drawings are issued after review, and on-site processing is carried out according to the design drawings. In order to ensure the smooth transmission of force between the steel platform and the steel-concrete column and avoid the adverse effects of eccentric compression, the axis grid is set out and the line is elastic (Figure 7), and the main beam, secondary beam and steel plate of the steel platform are welded and connected (Figure 8). The groove butt welding method is used to effectively reduce the maximum mid-span bending moment of the main and secondary beams. At the same time, in order to prevent safety accidents caused by improper operation of the crane driver, the roadbed box needs to be fully laid on the top of the steel platform (Figure 9) to enhance safety.



Figure 7 Steel Platform Axis Grid Layout



Figure 8 Steel Platform Welding



Figure 9 Roadbed Box Full Paving

3.2 Truss Pre-Assembly and Support Frame Erection

In order to improve the efficiency and accuracy of truss on-site installation, a combination of physical pre-assembly and computer virtual pre-assembly (Figure 10) is adopted to inspect the finished truss components and verify each other, so as to ensure the processing and manufacturing accuracy of the components shipped from the factory. The support frame assembled on site adopts standard lattice support (Figure 11), the support frame section is 2×2, and the vertical height is 1.5m per grid. The cross-section of the lattice frame vertical rod is P89×4, the cross-section of the web rod is P64×3.5, the bottom platform is made of HM300×200×8×12 steel to make a field shape, and the specifications of the other bottom rods are also HM300×200×8×12.

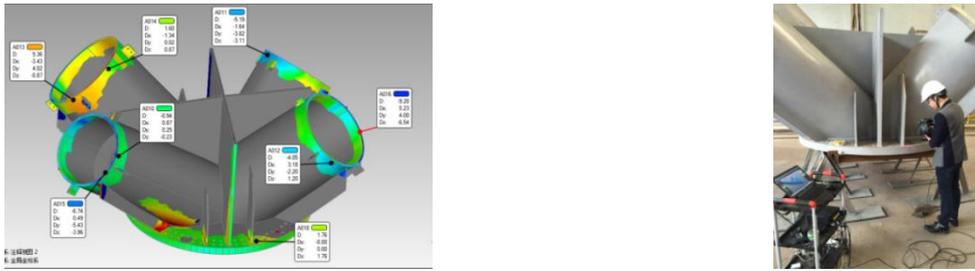


Figure 10 Analysis of Virtual Pre-Assembly Model

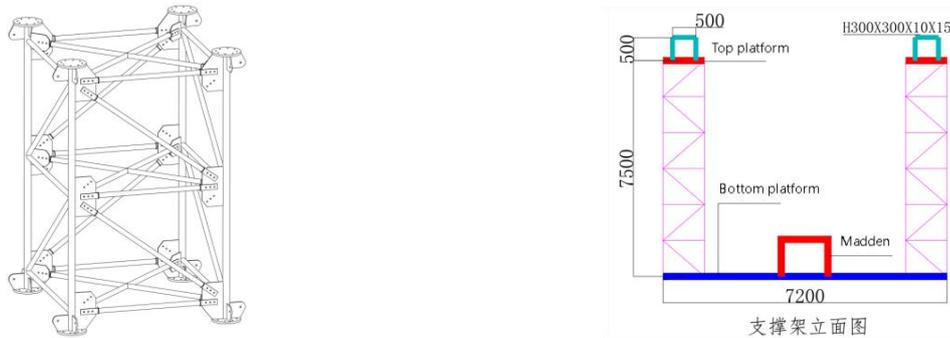


Figure 11 Schematic Diagram of Support Frame

3.3 Hoisting of Steel Structure of Double-Layer Exhibition Hall

The hoisting of steel structure of double-layer exhibition hall adopts high-precision intelligent three-dimensional construction. The installation sequence is: first hoist the lower frame, and assemble the cantilever truss of the roof, the second-floor plane truss and the inverted triangle truss of the roof off-site, and then transport them to the installation location using a flatbed truck. During hoisting, use a 200t crawler crane to hoist the roof and the second-floor truss synchronously in sections. The specific installation process is: frame structure → second-floor truss → roof truss → cantilever truss.

In addition, through the high-precision intelligent hoisting construction method combined with BIM technology and electronic monitoring, the traditional hoisting method is changed, mainly through close observation of the machine and BIM technology to calibrate the hoisting route, while manual monitoring is carried out on site to comprehensively control potential risks. The hoisting process is as follows: (1) Install cameras, proximity sensors and monitoring equipment on site; (2) Manually identify the construction environment and analyze the hoisting path; (3) In the same spatial coordinate system, establish an alarm device based on the prefabricated components and the construction site; (4) Determine the optimal hoisting path through computer image analysis; (5) Calculate the required hoisting data and guide the tower crane operator to complete the hoisting of the prefabricated components; (6) Set the initial parameters of the tower crane and adjust it to the target position; (7) Use an electronic camera to obtain images in real time during the hoisting process to determine whether the hoisting path meets the preset standards, and transmit the results to the monitoring equipment through wireless transmission; (8) When the hoisting meets the preset path, a prompt sound is issued to allow the hoisting to continue until all prefabricated components are hoisted; (9) If a deviation is found, an alarm is issued and the hoisting is stopped. The operation is restarted after adjustment using monitoring until all hoisting is completed.



Figure 12 Schematic Diagram of High-Precision Intelligent Hoisting

4 CONCLUSION

The double-story exhibition hall ground steel structure construction technology proposed in this study provides an

efficient and economical solution to the quality, safety and cost challenges faced by large-span space buildings in construction. By setting up a conversion steel platform on the top plate of the basement, combined with intelligent lifting technology and robot welding, the construction efficiency is significantly improved, the labor cost is reduced, and the safety of the underground structure is effectively guaranteed. The application of this technology in the Hangzhou Convention and Exhibition Center project has proved its advantages in shortening the construction period, improving the quality and safety of the project, and has been highly praised by the owner and related units. In the future, with the continuous improvement of requirements for green buildings and construction efficiency, this technology is expected to be promoted in a wider range of construction projects, promoting the sustainable development of the construction industry.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

References

- [1] Li Jingyuan, Wang Zhe, Yu Yinquan, et al. Analysis of key issues in the design of large-span heavy steel structure trusses for Nanchang Cultural and Sports Center. *Building Structure*, 2024, 54(13): 23-29.
- [2] Dong Zhaohai, Hou Shuangjun, Zhou Yuqing, et al. Structural design of gymnasium in Lishui Nancheng Sports Complex. *Building Structure*, 2023, 53(S2): 445-449.
- [3] Wang Hongchen, Lu Ji, Yin Longxing, et al. Design of long cantilever and large-span steel structure for Xi'an City Exhibition Center. *Building Structure*, 2024, 54(11): 84-92.

PERFORMANCE COMPARISON OF CEEMDAN-LSTM AND BASIC LSTM MODELS IN PREDICTING REALIZED VOLATILITY

ZheShuo Zhang

Wenzhou-Kean University, Wenzhou 325060, Zhejiang, China.

Corresponding Email: zhanzhes@kean.edu

Abstract: This paper presents a comparative analysis of the effectiveness of hybrid CEEMDAN-LSTM models and traditional LSTM models in predicting realized volatility in financial markets. By utilizing realized volatility data from 2004 to 2024, the study highlights significant market fluctuations during the 2008 financial crisis and the 2020 COVID-19 pandemic. The findings indicate that the CEEMDAN-LSTM model, which decomposes time series data into intrinsic mode functions (IMFs) before applying LSTM networks, outperforms the basic LSTM model in terms of predictive accuracy, particularly during periods of high volatility. This enhanced performance is evidenced by lower error metrics, such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). The research underscores the value of integrating advanced decomposition techniques with deep learning models to better capture the complex dynamics of financial markets.

Keywords: CEEMDAN-LSTM; Intrinsic mode functions (IMFs); Mean Absolute Error (MAE); Mean Squared Error (MSE)

1 INTRODUCTION

Long sequence time-series forecasting (LSTF) is gaining increasing attention and application across various fields. The main approaches involve establishing time series models and utilizing machine learning techniques. With the development of numerous methods, accurately distinguishing the strengths and weaknesses of different models and choosing appropriate methods for forecasting in different domains has become increasingly important. Therefore, this article aims to draw the following conclusions by comparing the different predictive performances of Decision Trees (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Autoregressive (AR), and Hybrid ARIMA and Recurrent Neural Networks (HAR): (i) which forecasting approach is more accurate in the same domain; (ii) which forecasting approach is more efficient in the same domain; and (iii) in which domains certain forecasting approaches are more suitable for application.

The financial market is characterized by its inherent complexity and volatility, posing significant challenges for accurate stock market prediction. Traditional forecasting methods often struggle due to the non-linear and non-stationary nature of financial time series data. To address these challenges, researchers have increasingly turned to hybrid models that combine advanced signal decomposition techniques with sophisticated neural networks. One promising approach that has emerged in recent years is the integration of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Long Short-Term Memory (LSTM) networks. CEEMDAN effectively decomposes complex time series data into simpler intrinsic mode functions (IMFs), which can then be processed by LSTM networks to capture both linear and non-linear patterns. This literature review provides an in-depth analysis of recent studies employing CEEMDAN-LSTM models for stock market prediction, highlighting their methodologies, findings, and contributions to the field.

Adebiyi et al.[1] perform a comparative analysis between ARIMA (AutoRegressive Integrated Moving Average) and Artificial Neural Networks (ANNs) models for stock price prediction. Their study highlights the strengths and limitations of both approaches. The research demonstrates that while ARIMA models are effective for linear time series data due to their reliance on past values and error terms, they often fall short in capturing the non-linear patterns inherent in financial markets. Conversely, ANNs exhibit a superior capability to model complex, non-linear relationships within stock price data, owing to their flexible structure and learning algorithms. Adebiyi et al. [1] conclude that ANNs generally outperform ARIMA models in stock price prediction tasks, especially in capturing intricate market dynamics. This comparison underscores the potential benefits of integrating neural network techniques with traditional statistical methods to enhance forecasting accuracy.

Recent studies have explored the application of CEEMDAN-LSTM models in financial time series forecasting, demonstrating their superior performance compared to traditional methods. Akşehir and Kılıç [2] propose a novel denoising approach, 2LE-CEEMDAN, which enhances the accuracy of time series forecasting by effectively decomposing complex signals into simpler components, thus facilitating more accurate LSTM modeling. Their study underscores the importance of noise reduction in improving predictive performance. Similarly, Cao et al. [3] present a comprehensive financial time series forecasting model that integrates CEEMDAN with LSTM. Their model demonstrates superior performance, attributed to CEEMDAN's ability to decompose time series into intrinsic mode functions (IMFs), which are then used as inputs for LSTM networks, effectively capturing both linear and non-linear

patterns in the data.

Further exploration of CEEMDAN-LSTM in financial forecasting by Guan [4] emphasizes the model's robustness in handling non-stationary time series. The study provides empirical evidence of improved prediction accuracy, highlighting the model's potential in real-world financial applications. Guresen et al. [5] also contribute to this area by exploring the use of artificial neural network (ANN) models in stock market index prediction. Their research highlights the potential of ANNs in capturing complex patterns within financial time series data, laying the groundwork for subsequent hybrid models like CEEMDAN-LSTM. Lin et al. [6] investigate the use of CEEMDAN-LSTM for forecasting stock index prices, finding that the model significantly outperforms conventional forecasting methods, particularly in capturing sudden market movements. They discuss the implications of using advanced decomposition techniques in financial modeling. Extending this research, Lin et al. [7] predict the realized volatility in stock price indices using a hybrid CEEMDAN-LSTM model. This approach combines the strengths of CEEMDAN in noise reduction and LSTM in sequence learning, resulting in highly accurate volatility forecasts.

In addition to the primary studies on CEEMDAN-LSTM, other researchers have proposed enhancements and comparative studies to further refine forecasting models. Assaad et al. [8] present a new boosting algorithm for time-series forecasting using recurrent neural networks (RNNs), improving forecasting accuracy by iteratively enhancing the model's performance on difficult-to-predict data points. The principles of boosting can be applied to CEEMDAN-LSTM models to refine their predictive capabilities, particularly in handling complex and non-linear financial data. Baek and Kim [9] introduce ModAugNet, a novel forecasting framework that addresses overfitting in LSTM models. Their approach involves an overfitting prevention module and a prediction module, significantly improving prediction accuracy and generalization capability. Insights from this study can be leveraged to enhance CEEMDAN-LSTM models by incorporating overfitting prevention techniques, ensuring robust performance in diverse market conditions.

Pin Lv et al. [10] investigate a hybrid model for stock index prediction based on modal decomposition techniques. Their research focuses on enhancing prediction accuracy by leveraging the strengths of various decomposition methods. The study demonstrates that using modal decomposition allows for the isolation of significant components within stock index data, thereby improving the inputs for predictive models. By integrating these decomposed components with advanced forecasting models, the hybrid approach provides a more robust and accurate prediction framework. The findings of Pin Lv et al. [10] highlight the importance of combining decomposition techniques with sophisticated modeling strategies to capture the complex dynamics of financial markets. Qi et al. [11] explore a variation of the hybrid model by integrating CEEMDAN with Wavelet Transform and GRU (Gated Recurrent Unit) networks for stock price prediction. This study highlights the effectiveness of combining multiple decomposition techniques with advanced neural networks, providing a comparative analysis with CEEMDAN-LSTM models and demonstrating the potential for further improvements in forecasting accuracy. Additionally, Sun and Liu [12] apply a CEEMDAN-ARMA-LSTM model for Air Quality Index (AQI) prediction, showcasing the versatility of CEEMDAN-LSTM frameworks beyond financial markets. Their findings suggest that integrating autoregressive moving average (ARMA) models with CEEMDAN-LSTM can enhance predictive accuracy for various types of time series data. Wang et al. [13] focus on predicting green bond indices using a CEEMDAN-LSTM model, illustrating the applicability of this hybrid approach in sustainable finance. The model's ability to handle the unique characteristics of green financial instruments is emphasized. Furthermore, Yanan et al. [14] delve into the prediction of chaotic time series using LSTM with CEEMDAN, providing insights into the model's capability to deal with highly irregular and complex data patterns, reinforcing the value of CEEMDAN-LSTM in diverse forecasting scenarios.

2 METHODS AND RESULTS

Firstly, we need to import the necessary libraries, set up the plot styles, and read data from a CSV file. Firstly, it need to import the necessary libraries. 'pandas' is used for data manipulation and analysis. 'numpy' is used for numerical calculations. 'datetime' is used for handling date and time. 'matplotlib.pyplot' is used for plotting graphs. Then, it set the plot styles, 'plt.style.use('seaborn-v0_8')' means that applies Seaborn's plotting style (version 0.8). 'plt.rcParams['figure.figsize']' means that sets the plot size to 12x6 inches. 'plt.rcParams['figure.dpi']' means sets the plot resolution to 300 DPI. Finally, the table includes columns for the trading date, closing price, opening price, highest price, and lowest price. The overall purpose of this code is to prepare the data for further analysis or visualization.

Then code set some extract the opening and closing prices of each trading day and store them in a new data frame. 'total_days', 'daily_open' and 'daily_close' correspond to the date, opening price and closing price.

RV refers to realized volatility, which refers to the fluctuation range of asset price changes that have occurred, measured by calculating the standard deviation of asset prices over a period of time. The main purpose of this code is to calculate the log returns for each trading day and calculate the RV (Realized Volatility) based on these returns and store the results in a data frame for further analysis.

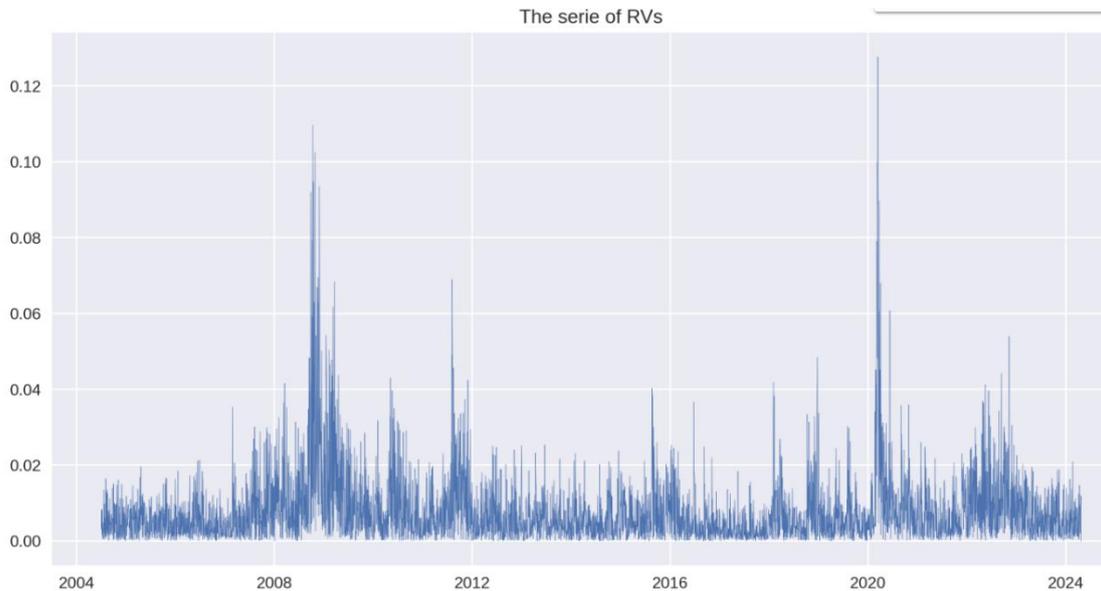


Figure 1 The Series of RVs

This represents the realized volatility from 2004 to 2024, with significant spikes occurring in 2008 due to the financial crisis and again in 2020 as a result of the COVID-19 pandemic.

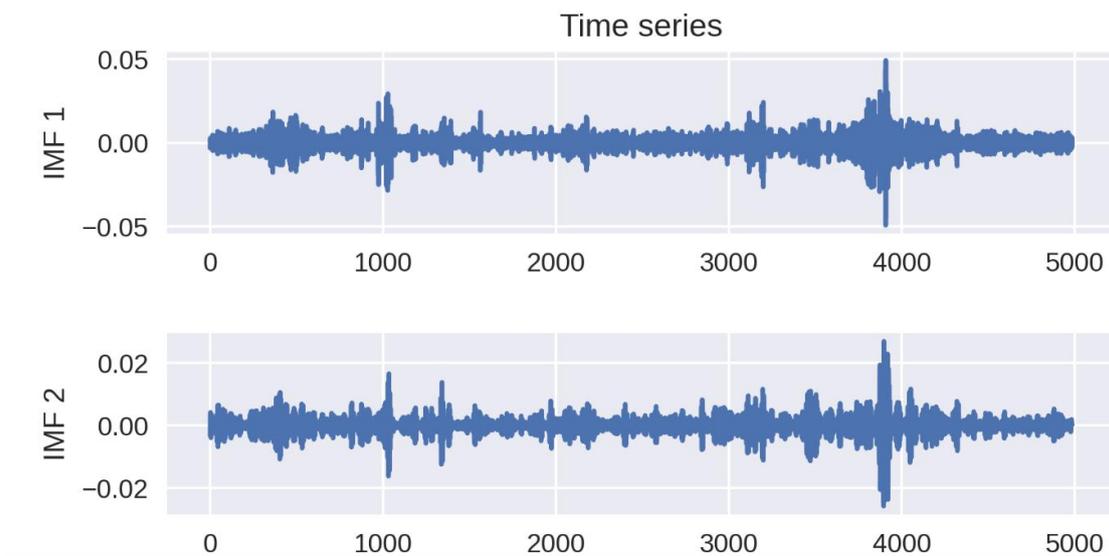


Figure 2 Time Series Diagram of IMF1 and IMF2

This image show how to use CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) to decompose time series data and visualize the decomposed results.

The picture shows the time series of the first intrinsic mode function (IMF 1). The horizontal axis represents time and the vertical axis represents the amplitude of IMF 1. This figure reflects the first level of decomposition of the original time series after CEEMDAN processing. IMF usually represents different frequency components in the time series, and the more forward IMF contains higher frequency components.

This image show how to calculate the statistical characteristics of the IMFs (Intrinsic Mode Functions) and residuals obtained from the previous decomposition, and display these characteristics in tabular form.

The table in the picture shows the statistical characteristics of each IMF and residual. Each row represents an IMF or residual, and the columns represent: count: the number of data points, mean, std (the standard deviation), skew (skewness, reflecting the symmetry of the data distribution), kurtosis (kurtosis, reflecting the sharpness of the data distribution), J-B (Jarque-Bera statistic, reflecting whether the data is close to a normal distribution), Q(10)(Ljung-Box statistic, used to test the autocorrelation of timeseries)

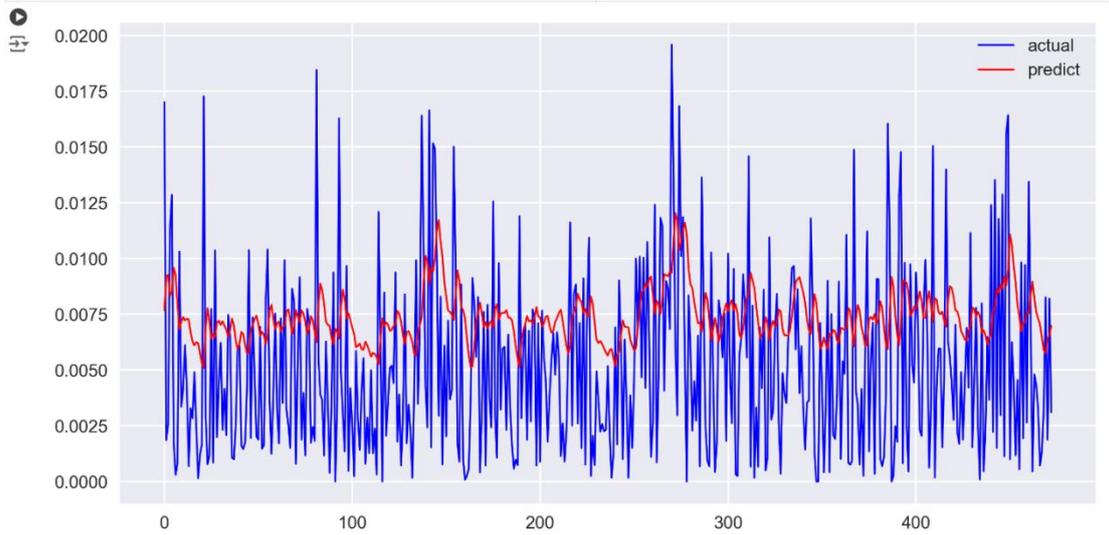


Figure 3 Actual vs Predicted Values Using Basic LSTM on Time Series Data

Table 1 Error Table of Basic LSTM on Time Series Data

	MAE	MSE	HMAE	HMSE
Basic LSTM	0.00355	1.82079	0.60725	0.53171

First, we use traditional time series models to make forecasts. The image displays a comparison between the actual values (blue curve) and the predicted values (red curve) of a time series model. These metrics indicate that there are discrepancies between the predicted and actual values, especially in certain fluctuating parts, suggesting that the prediction accuracy is not optimal.

It also uses four metrics: MAE (Mean Absolute Error), MSE (Mean Squared Error), HMAE (Harmonic Mean Absolute Error), and HMSE (Harmonic Mean Squared Error) to evaluate the model's accuracy. These values indicate the prediction effect of the model. The smaller the value, the smaller the model error.

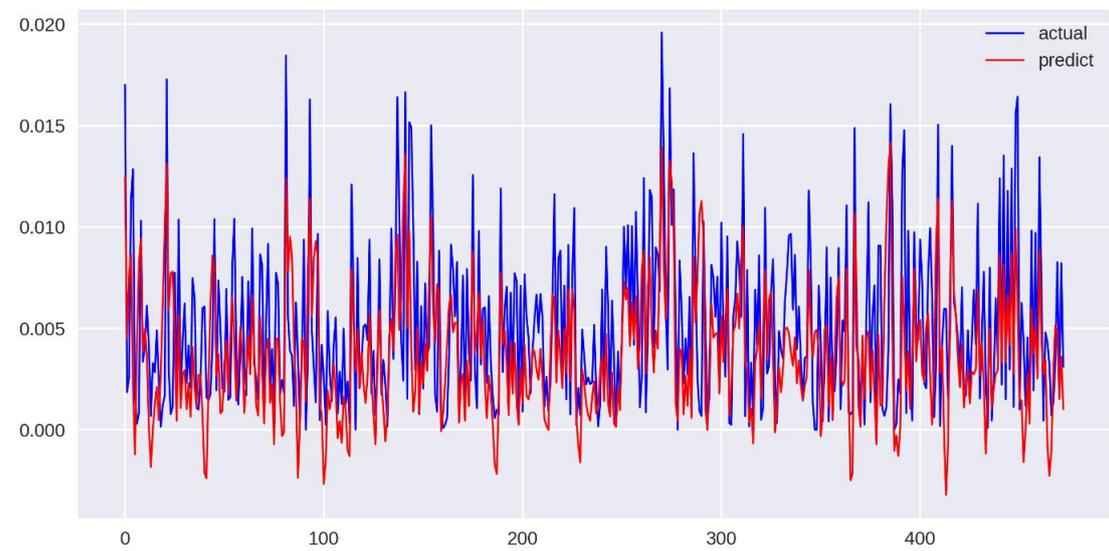


Figure 4 Actual vs Predicted Values Using CEEMDAN-LSTM Model on Time Series

The CEEMDAN-LSTM model first uses CEEMDAN to decompose the time series into several IMF components, then uses the LSTM model to predict each component, and finally recombine the prediction results to obtain the final prediction results. The prediction at this time is the sum of the predictions of all decomposition results.

As can be seen from the figure, the red forecast curve closely follows the blue actual curve. Although there are errors in some fluctuations, the accuracy of traditional time series forecasting has been greatly improved.

Table 2 Error table of Basic LSTM and CEEMDAN-LSTM Model on Time Series

	MAE	MSE	HMAE	HMSE
--	-----	-----	------	------

TimeSeries	0.00355	1.82079	0.60725	0.53171
CEEMDAD LSTM	0.00273	1.16912	0.73335	0.84252

3 CONCLUSION

This study evaluates and compares the performance of two predictive models—CEEMDAN-LSTM and Basic LSTM—in forecasting realized volatility. Realized volatility data from 2004 to 2024 was analyzed, capturing significant market fluctuations during the 2008 financial crisis and the 2020 COVID-19 pandemic.

The results indicate that the CEEMDAN-LSTM model, which combines the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Long Short-Term Memory (LSTM) networks, outperforms the Basic LSTM model. The CEEMDAN-LSTM approach decomposes the time series data into various Intrinsic Mode Functions (IMFs) and residuals, then uses LSTM to predict each component, leading to more accurate predictions, especially in periods of high volatility.

This model's effectiveness is reflected in lower error metrics, such as Mean Absolute Error (MAE) and Mean Squared Error (MSE), when compared to the Basic LSTM model. The CEEMDAN-LSTM model offers a more refined prediction of volatility, better capturing market dynamics and reducing prediction discrepancies, particularly in turbulent market conditions.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

- [1] Adebisi A A, Adewumi A O, Ayo C K. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, 2014: 1–7.
- [2] Akşehir Z D, Kılıç E. A new denoising approach based on mode decomposition applied to the stock market time series: 2LE-CEEMDAN. *PeerJ. Computer Science*, 2024, 10, e1852.
- [3] Cao J, Li Z, Li J. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, 2019, 519: 127–139.
- [4] Guan Y. Financial time series forecasting model based on CEEMDAN-LSTM. 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC), 2022.
- [5] Guresen E, Kayakutlu G, Daim T U. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 2011, 38(8): 10389–10397.
- [6] Lin Y, Yan Y, Xu J, Liao Y, Ma F. Forecasting stock index price using the CEEMDAN-LSTM model. *The North American Journal of Economics and Finance*, 2021, 57: 101421.
- [7] Lin Y, Lin Z, Liao Y, Li Y, Xu J, Yan Y. Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. *Expert Systems with Applications*, 2022, 206: 117736.
- [8] Assaad M, Boné R, Cardot H. A new boosting algorithm for improved time-series forecasting with recurrent neural networks. *Information Fusion*, 2008, 9(1): 41–55.
- [9] Baek Y, Kim H Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, 2018, 113: 457–480.
- [10] Pin Lv, Shu Y, Xu J, Wu Q. Modal decomposition-based hybrid model for stock index prediction. *Expert Systems with Applications*, 2022, 202: 117252–117252.
- [11] Qi C, Ren J, Su J. GRU Neural Network Based on CEEMDAN-Wavelet for Stock Price Prediction. *Applied Sciences*, 2023, 13(12): 7104–7104.
- [12] Sun Y, Liu J. AQI Prediction Based on CEEMDAN-ARMA-LSTM. *Sustainability*, 2022, 14(19): 12182.
- [13] Wang J, Tang J, Guo K. Green Bond Index Prediction Based on CEEMDAN-LSTM. *Frontiers in Energy Research*, 2022, 9.
- [14] Yanan G, Xiaoqun C, Bainian L, Kecheng P. Chaotic Time Series Prediction Using LSTM with CEEMDAN. *Journal of Physics: Conference Series*, 2020, 1617: 012094.

CONSTRUCTION TECHNOLOGY OF LARGE-SPAN CAVITY VENTILATION SHAFT WITH SMALL CLEARANCE ON ROOF

YongYan Li¹, Yang Yang^{1*}, XinXin Jin², ZhiCheng Bai¹, Jie Luo¹, Kun Wu¹

¹China Construction Fourth Engineering Bureau Co., Ltd. Guangzhou 510000, Guangdong, China.

²Zhejiang Provincial Architectural Design and Research Institute Co., Ltd. Hangzhou 310000, Zhejiang, China.

Corresponding Author: Yang Yang, Email: 2557841138@qq.com

Abstract: The top of the traditional cavity ventilation shaft adopts cast-in-place concrete structure. The formwork system can be removed only after the concrete reaches the strength. If the clearance of the cavity ventilation shaft is only 40cm-70cm, the construction workers cannot carry out construction in the narrow space. Failure to remove the formwork system not only affects the function of the ventilation shaft, but also causes waste of materials. For small-span cavities, the floor decking is constructed and cast in one go. However, in order to solve the fulcrum problem of large-span floor decking, the traditional method is to set concrete structural beams in the cavity as the support of the large-span floor decking. However, this method will affect the ventilation volume of the air shaft due to the horizontal components. In order to ensure that the cavity air shaft does not affect the use function, a construction method of a single floor deck connecting beam for a large-span cavity air shaft with small roof clearance is proposed. The research results show that the construction method of the small clearance and large span cavity wind shaft on the roof can effectively ensure the one-time casting effect of the roof and avoid the risk of roof leakage. This method ensures the use function of the wind shaft and the ventilation volume required by the design; the construction method of the small clearance and large span cavity wind shaft on the roof is convenient for construction, shortens the construction period, reduces costs, is green and sustainable, has reasonable force, is safe and reliable, and can effectively solve the construction difficulties encountered in the small clearance and large span cavity wind shaft; compared with the characteristics of traditional concrete structural beams as large span floor deck support, the cost saving rate of the optimized single floor deck connection beam construction of the small clearance and large span cavity wind shaft on the roof is about 88%, which is worthy of further promotion and application.

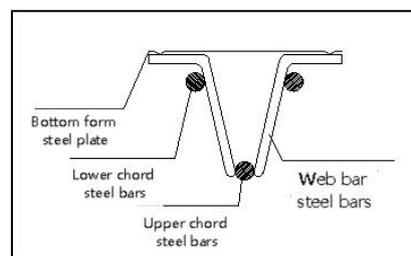
Keywords: Single floor deck connection beam; Cavity wind shaft; Floor deck

INTRODUCTION

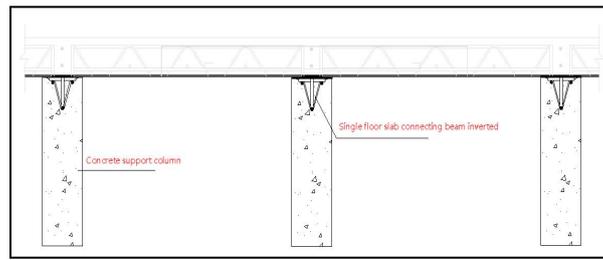
In modern architectural design, the application of large span structures is becoming more and more common, especially when small clearance cavity wind shafts are involved in roofs. In order to solve the construction problem of small clearance cavity air shaft, the floor deck can be cast in one go for small span cavity, but the support point problem of large span floor deck still needs to be overcome. The traditional method is to set concrete beams in the cavity as supports, but this method will affect the ventilation volume of the air shaft [1-3]. For this reason, a single floor deck connecting beam construction method for roof small clearance large span cavity air shaft is proposed, aiming to ensure the ventilation volume and use function of the air shaft and solve the construction problem of roof small clearance large span cavity air shaft.

1 CONSTRUCTION PRINCIPLE OF ROOF SMALL CLEARANCE LARGE SPAN CAVITY AIR SHAFT

As shown in Figure 1, the single floor deck connecting beam construction method for roof small clearance large span cavity air shaft is to use floor deck as the top plate of the air shaft, and the middle support position of the floor deck adopts the vertical component 200mm×200mm concrete column + single floor deck connecting beam inverted. The single floor slab connecting beam has 1 upper chord steel bar and 2 lower chord steel bars, which are connected by web bars and have a bottom formwork steel plate. The use of this single floor slab connecting beam will not affect the ventilation volume of the air shaft; the inverted triangle structure of the single floor slab connecting beam is used to increase the bearing capacity of the connecting beam steel bars to ensure the use function and construction safety of the cavity air shaft.



(a)



(b)

Figure 1 Construction Principle of Large-Span Cavity Air Shaft with Small Clearance on the Roof

2 MAIN PROCESS FLOW

2.1 Drawing Deepening

Before deepening, carefully study the design drawings and related technical requirements to understand the designer's design intent; confirm the key parameters such as the type, material, span, and load of the floor slab; check the main axis position to ensure that the position and size of the floor slab are consistent with the overall structure; design a reasonable connection form according to the different connection parts, refine the connection nodes, and ensure the reliability of the connection and the convenience of construction.

2.2 Concrete Column Construction

(1) Rebar binding: Before tying the reinforcement, the reinforcement should be positioned to ensure the accurate position of the reinforcement. The specifications, quantity and spacing of the reinforcement should be checked according to the drawings, and the reinforcement should be tied according to the design requirements to ensure that the column reinforcement is vertical and the stirrups are straight. After the column reinforcement is tied, the single floor deck reinforcement of the connecting beam is anchored into the column reinforcement (Note: the bottom formwork steel plate of the single floor deck is opened within the column range).

(2) Formwork reinforcement: The column formwork should be installed flat and vertical, the joints should be tight, and no mortar leakage should be allowed. The support of the formwork should be stable, and the support spacing should meet the requirements of the specifications. Back ribs and other reinforcement measures should be used to ensure the overall stability of the formwork.

(3) Concrete pouring: The mix ratio of concrete should be determined according to the design requirements and site conditions to ensure that the slump of the concrete meets the construction requirements. Concrete should be poured from the bottom of the column in layers. The thickness of each layer should not be too large to avoid concrete segregation and ensure dense pouring. After pouring, it should be maintained in time to prevent concrete cracks.

2.3 Construction of ventilation shaft wall

(1) Rebar binding: Process the rebar according to the requirements of the drawings, including cutting, bending, etc., to ensure compliance with the specifications. The rebar should be cleaned and the vertical rebar (vertical rebar) should be tied first, followed by the transverse rebar (horizontal rebar). The vertical rebar should be vertical and the horizontal rebar should be horizontal. Ensure that the rebar mesh is flat and straight. The lap length and anchorage length of the rebar should meet the design specifications.

(2) Angle iron welding: Clean the welding part before welding the angle iron to the shear wall rebar, select the appropriate welding material, select the appropriate welding method according to the specifications and materials of the angle iron and rebar, ensure that the welding current, voltage and other parameters meet the specifications, and the welding sequence should be from the center to both sides to reduce welding deformation and ensure welding quality.

(3) Formwork reinforcement: The formwork should be installed according to the layout position to ensure verticality and flatness. The formwork joints should be tight and no slurry should leak.

2.4 Floor Decking Laying

(1) Floor decking should be laid in the direction and sequence required by the design, starting from one end and gradually laying towards the other end to ensure the continuity and stability of the floor decking.

(2) The connection between floor decking should be tight to avoid gaps. During the laying process, the horizontality and verticality of the floor decking should be checked regularly to ensure that it meets the design requirements. Use pads or adjusters to make fine adjustments to ensure the flatness of the floor decking.

(3) The welding speed and temperature should be strictly controlled at the connection between the floor decking and the single floor decking of the connecting beam and the connection of the shear wall angle steel to avoid overheating and welding defects. After welding, the welding quality should be checked.

2.5 Concrete Pouring

- (1) Before pouring, a slump test should be carried out to ensure that the concrete quality meets the requirements. The wall should be poured first, and then the slab. During the pouring process, attention should be paid to the vibration of the concrete to ensure the compactness of the concrete after pouring.
- (2) After pouring, it should be covered in time and maintained according to the standards. Avoid working on the concrete before the concrete strength reaches the required level to prevent footprints.

3 QUALITY ASSURANCE MEASURES

3.1 Quality Control Standards

- (1) It is strictly forbidden to directly lift the floor decking with wire ropes during loading, unloading and installation. There should be enough fulcrums for transportation and stacking to prevent deformation.
- (2) The bent and deformed floor decking should be corrected before laying.
- (3) Cutting and hole cutting should be done with a plasma cutter. It is strictly forbidden to cut with oxygen-acetylene flame. The surroundings of large holes should be reinforced.
- (4) The floor decking should be installed, straightened, compacted and spot welded according to the drawings.
- (5) After the floor decking is laid, straightened and fixed, it should be locked with a locking machine in time to prevent the corrugated board from biting and separating due to the stacking of construction materials and personnel traffic.
- (6) After installation, clean up the construction waste in time, and the cut scraps should be collected on the ground and stacked together.

3.2 Quality Assurance Measures

- (1) Strengthen technical management and conscientiously implement national regulations, standards, operating procedures and various management systems.
- (2) Establish a complete quality management system with the project manager and project leader in charge, and set up capable and experienced professional quality inspectors to conduct quality inspection supervision and technical guidance for each process.
- (3) Strictly implement quality target management, closely link quality with job income, and implement the quality target responsibility system. The technical leader is fully responsible for quality, and the quality inspector has the veto power over construction quality, so that quality management is always under control.
- (4) The project department should hold a quality meeting for on-site production every day, conduct a comprehensive inspection of the project every week, and carry out three analysis activities, namely: analyze the quality problems, analyze the causes of quality problems, analyze the measures to be taken, and find out the problems and rectify them in time.

4 ANALYSIS OF THE CONSTRUCTION ADVANTAGES OF LARGE-SPAN CAVITY AIR SHAFTS WITH SMALL CLEARANCE ON ROOFS

The construction method of single floor slab connecting beam for large-span cavity air shafts with small clearance on roofs can effectively meet the functional requirements of the air shaft, ensure the structural safety of the roof, and save construction costs to a certain extent. For other roofs with small clearance and large-span cavity air shafts The characteristics of using traditional concrete structural beams as large-span floor slab supports are:

- (1) Easy to construct and shorten construction period: Single floor slab connecting beams are prefabricated floor slabs. Compared with concrete structures, single floor slab connecting beams only need to be installed on site, without the need for formwork erection and concrete pouring. The construction is convenient and greatly shortens the construction period.
- (2) Reduce costs and be green and sustainable: Compared with concrete horizontal components, this construction method can reduce construction costs in terms of labor; in terms of materials, it avoids the waste of concrete and formwork materials, thereby protecting the environment and achieving green development.
- (3) Reasonable force, safe and reliable: The force of a single floor slab connecting beam as a large-span floor slab support is reasonable, which can effectively ensure safety during construction.

According to cost calculation, the cost of using concrete beams at the support point of the large-span floor slab is 1,770 yuan per unit of material and labor costs, while the cost of using a single floor slab connecting beam is 200 yuan per unit of material and labor costs, which can save 1,570 yuan, and the saving rate is about 88%.

5 ENGINEERING APPLICATION

The second phase of the Huai'an Jianhua Jiulong Peninsula project is located at Jianhua Guanyuan, southwest of the intersection of Cheng'en Avenue and Shanyang Avenue in Huai'an City, Jiangsu Province. The total construction area of the project is 186,051.04 square meters, including 5 high-rise residential buildings, 3 of which are 33 stories high and

97.8 meters high, 2 are 34 stories high and 99.9 meters high, 2 power distribution rooms and garbage rooms, 1 commercial 3-story commercial building and commercial supporting underground garage. The engineering structure is a frame shear structure. The project adopts the construction method of single floor slab connecting beam for small clearance and large span cavity air shaft on the roof, which ensures the use function of the roof air shaft. Under the premise of ensuring quality, the comprehensive benefits have been recognized by relevant units.

6 CONCLUSION

This paper proposes a construction method of single floor slab connecting beam for small clearance and large span cavity air shaft on the roof, and uses theoretical analysis method combined with on-site engineering practice to clarify the construction principle, construction process, operation points, quality assurance measures and method advantages of this method. The main conclusions are as follows:

- (1) The construction method of small clearance and large span cavity air shaft on the roof can effectively ensure the one-time casting effect of the roof and avoid the risk of roof leakage. This method ensures the use function of the air shaft and ensures the ventilation volume required by the design.
- (2) The construction method of small clearance and large span cavity air shaft on the roof is easy to construct, shortens the construction period, reduces costs, is green and sustainable, has reasonable force, is safe and reliable, and can effectively solve the construction difficulties encountered in small clearance and large span cavity air shaft.
- (3) Compared with the characteristics of traditional concrete structural beams as large-span floor deck supports, the cost saving rate of the optimized single floor deck connection beam construction for small roof clearance and large-span cavity air shaft is about 88%, which is worthy of further promotion and application.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Zhou Jian, Wang Huanhuan, Tian Chunyu, et al. Research on the design method of steel truss floor deck without dismantling bottom formwork. *Building Structure*, 2022, 52(S1): 1450-1458.
- [2] Wang Shusen, Zhou Yang, Wang Haoyang, et al. Discussion on the construction technology and process of assembled steel truss floor deck. *Sichuan Building Materials*, 2023, 49(10): 111-113.
- [3] Yang Wenfei. Application of steel truss floor deck in super high-rise buildings. *Building Technology*, 2023, 54(01): 92-95.

A COMMAND SUPPORT MODEL BASED ON BLOCKCHAIN AND THE METAVERSE

ChaoYue Song, YiLiang Han*, XuGuang Wu
College of Cryptography Engineering, Engineering University of PAP, Xi'an 710086, China.
Corresponding Author: YiLiang Han, Email: hanyil@163.com

Abstract: Traditional command support systems face issues such as deficiencies in decision support, data silos, and risks of forgery and tampering, which significantly hinder the enhancement of combat effectiveness and support capabilities. The metaverse, through tools like visualization, immersive platforms, and scenario simulations, aids in decision-making. However, it is threatened by data security concerns and asset ownership issues. Fortunately, blockchain technology offers a higher level of trust and security for the metaverse, serving as a foundational technology to address its shortcomings while effectively facilitating data flow and integration, thus eliminating data silos. Based on a needs analysis, this paper proposes a command support system model integrating the metaverse and blockchain. It provides a detailed description of the model's construction and briefly introduces the challenges and solutions associated with this integration. This work offers valuable insights for aligning future intelligent equipment supply support needs with the demands of modern warfare.

Keywords: Metaverse; Blockchain; Command support; Military application

1 INTRODUCTION

The rapid transformation of modern warfare necessitates that command support systems possess greater flexibility and responsiveness to adapt to increasingly complex battlefield environments. In this context, the efficiency of command support systems becomes a critical factor influencing the outcome of conflicts [1]. Traditional command support systems often employ a centralized hierarchical management model, where decision-making processes typically require multiple levels of approval, leading to cumbersome procedures and slow response times that hinder adaptability to rapidly changing battlefield demands. Additionally, centralized data storage presents a risk of single points of failure; should the system collapse, it could severely disrupt information transmission and order issuance, thereby diminishing the overall command effectiveness. Therefore, the decentralization of command support systems is not only an optimization of management structure but also a crucial means to enhance command efficiency.

The metaverse and blockchain technology, as two prominent trends in the contemporary internet and information technology landscape, are increasingly reshaping our societal structures and developmental models. Beyond their widespread applications in entertainment, finance, and the arts, both the metaverse and blockchain hold significant potential for advancing national defense and military development. As an emerging virtual interaction platform, the metaverse injects new vitality into command support systems through immersive experiences, data visualization, and augmented reality technologies [2]. By providing intuitive information displays and dynamic predictive analytics, the metaverse enables commanders to better comprehend complex battlefield situations. For instance, data visualization techniques allow commanders to monitor real-time distributions of resources, battlefield dynamics, and the comparative strengths of friendly and enemy forces, facilitating more accurate decision-making.

Moreover, the real-time feedback mechanisms and collaborative platforms within the metaverse enhance communication efficiency among teams, enabling rapid information sharing and significantly improving the scientific and effective nature of the decision-making process. The virtual nature of the metaverse also allows for the creation of various simulation scenarios, providing an ideal environment for equipment testing and personnel training. Through virtual reality technology, commanders and soldiers can conduct tactical drills and equipment operations in a safe environment, accumulating experience and enhancing professional skills. This immersive training not only reduces the costs associated with actual exercises but also effectively improves the overall quality and adaptability of military forces.

While the metaverse possesses many distinctive features, user data is at risk of privacy breaches and misuse. This concern is particularly critical in command support tasks, where data security is paramount; any information leakage could lead to irreversible consequences. In this context, the introduction of blockchain technology is essential. By employing decentralization and encrypted storage, blockchain enhances data security, effectively protects user privacy, and ensures that information within command support systems is safeguarded against unauthorized access or tampering. This paper proposes a model for a command support system that integrates the metaverse and blockchain. The combination of these two technologies promises to provide more efficient and secure decision support in complex battlefield environments. The paper first outlines the characteristics of the metaverse and blockchain technologies, discussing the opportunities and challenges presented by their convergence, which is expected to significantly elevate the level of intelligence in command support systems and enhance operational effectiveness. Furthermore, it designs a command support system model, detailing its overall design, technical architecture, smart contracts, and application scenarios.

2 BACKGROUND

2.1 Metaverse

Metaverse is not a simple collocation of meta and universe, it is a compound word that transcends both, and is a three-dimensional virtual world that simulates participation in political, economic, and cultural activities [3]. Neal Stephenson coined the term "Metaverse" in his 1992 science fiction novel *Snow Crash* and conceived of a parallel virtual world. "and conceived of a parallel virtual world. With the Metaverse being a fusion of multiple disciplines and an integrated use of existing IT technologies [4]. With the advancement and application of technologies such as virtual reality and artificial intelligence, the concept of metaverse has evolved, such as second life [5] and 3D virtual world [6]. Usually, a metaverse is an intermingled multi-user shared space that integrates the ternary physical world, the human world, and the digital world into a single, self-sustaining, inter-temporal space, which is based on the fusion of technologies that enable multi-sensory interactions with virtual environments, digital objects, and people [7]. In particular, the metaverse has multiple features and infinite possibilities such as virtual, realistic immersive experience, intermingling, socialization, and interaction [8].

Among them, realistic immersive experience is to create an environment where it is almost impossible to distinguish the difference between the virtual and the real through a high degree of realism and immersive feeling, which is one of the key elements that enable the explosion of the metaverse. With the development of a variety of technologies such as somatosensory devices, digital smell, digital taste, brain-computer combination, etc., players are able to obtain an infinitely close to the reality in which they can feel, and this realism allows users to better integrate into it and enjoy a distinctive experience, and at the same time promotes the development of social exchanges, education and training, entertainment and culture, and other fields.

Intermingling refers to combining the virtual world with the real world. In a perfect meta-universe, the virtual world is seamlessly connected to the real world, allowing users to travel freely between the two worlds.

2.2 Military Metaverse

The military applications of the Metaverse, analogous to its manifestations in fields such as fashion, libraries, and healthcare, are collectively referred to as the "military Metaverse" [9]. Scholars describe the military Metaverse as the convergence of military domains and Metaverse technologies, integrating knowledge and techniques from various disciplines, including military science, computer science, artificial intelligence, data science, intelligence studies, design, and engineering, to construct virtual battlefields and simulate combat scenarios, thereby supporting military operations [10]. The military Metaverse is not merely a singular virtual world; its complexity lies in its capacity to accurately replicate a variety of scenarios that may arise in complex real-world environments, requiring the integration of advanced artificial intelligence technologies to simulate the behaviors and reactions of both friendly and adversarial actors. To fully understand the military Metaverse, it is essential to study it from multiple perspectives.

From a technological standpoint, while the Metaverse shares many similarities with the internet, it should be viewed as an application of the internet rather than the internet itself [11]. The Metaverse relies on the internet to enhance user experiences from "2D" to "3D," allowing users to engage with virtual worlds in a more realistic and immersive manner. Its technologies, applications, and business models differ significantly from those of the internet. Therefore, the Metaverse should not be simply regarded as the next generation of the internet but rather as one of the trends and directions for its further development. A singular scenario cannot constitute the military Metaverse; instead, it is an integrated network composed of numerous simulation platforms and virtual worlds, designed to provide a comprehensive and realistic training environment through high integration and seamless connectivity [9].

From the perspective of military training and education, human involvement plays a crucial role in the development of military capabilities, which is achieved through education, training, and management, with training being the primary avenue. Over the years, military forces have utilized various isolated and cumbersome training systems [12]. Consequently, the military has sought to link training with virtual worlds. The establishment of the Simulator Networking (SIMNET) broke new ground in how the entire industry and military conduct operations, serving as the first demonstration of a widespread simulator network for collective training and mission rehearsals [13]. Following this, the emergence of Distributed Interactive Simulation (DIS) and High-Level Architecture (HLA) standards accelerated the integration of different simulation training, allowing combat personnel to experience battlefield "fog" and environments in a synthetic space [14]. Despite effective training outcomes, interoperability and collaboration remain insufficient, posing significant challenges in real battlefield scenarios. In Orson Scott Card's novel *Ender's Game*, soldiers engage in training within a realistic immersive world [15], reflecting a contemporary desire for virtual training and exploration of virtual worlds. While various training centers are valuable, the COVID-19 pandemic disrupted physical institutions, prompting combat personnel and commanders to seek increased opportunities for distributed learning, such as through developed wargaming and faster-than-real-time tools [16]. Due to its realistic immersive user experience and rich multimedia capabilities, the Metaverse emerges as an ideal model for educational, training, and skill development tasks [17], offering tailored training feedback for individuals.

From the perspective of operational simulation and rehearsal, the military Metaverse can simulate various complex combat environments and scenarios, assisting commanders in formulating scientifically effective operational plans while assessing the feasibility and risks of various strategies. Different simulation environments can interconnect,

allowing soldiers to experience diverse military actions—from ground combat to aerial strategies and naval tactics—in a coherent and unified virtual space [18], thus equipping combat personnel to respond to real-world situations and rapidly enhance combat skills, thereby improving command and control capabilities. Additionally, personalized training programs can be developed to facilitate realistic training and enhance soldiers' operational readiness and response capabilities in emergencies [18].

The Metaverse is poised to lead the future paradigm of internet development, forming a value chain for the Metaverse [19]. Brown identifies three key advantages of the Metaverse that have garnered the attention of the Department of Defense: "first, establishing tighter global connections and networks; second, creating high-fidelity environments that simulate the real world; and third, providing immersive experiences". The impact of the military Metaverse extends beyond mere simulation and modeling; it significantly influences the interactions between humans, machines, and network computing, intersecting importantly with defense cyberspace initiatives in areas such as data, digital twins, and artificial intelligence [20]. Literature reviews indicate that the military applications of the Metaverse hold substantial potential across various directions, including military education and training, command decision-making and situational awareness, logistical support and exercises, and optimization of joint operational systems.

2.3 Blockchain

Blockchain technology possesses characteristics such as decentralization, distribution, immutability, transparency, anonymity, smart contracts, and high availability, providing new solutions for secure, reliable, and efficient data exchange and value transfer [21].

Decentralization means that in a blockchain network, there is no central authority or server controlling and managing the entire network; instead, all nodes collaboratively maintain and manage all transactions and data. This characteristic eliminates the risk of single points of failure, reduces the likelihood of system attacks, and ensures a high level of transparency and openness.

Immutability refers to the inability to modify, delete, or tamper with data once it is recorded on the blockchain. This is due to the decentralized distributed ledger technology, where all transaction data is recorded across multiple nodes, each possessing a complete copy of the ledger. Any modification would be detected and rejected by other nodes. Additionally, each block in the blockchain contains the hash of the previous block; any change would alter the hash values of all subsequent blocks, thereby compromising the integrity of the entire blockchain.

Anonymity indicates that users' identities and transaction information in the blockchain network are anonymous and cannot be easily tracked or identified, providing enhanced privacy protection.

Traceability means that blockchain technology allows for the tracing and recording of the origin and destination of every historical transaction through its chain structure and storage system, ensuring the authenticity and credibility of each on-chain transaction.

3 METAVERSE AND BLOCKCHAIN CONVERGENCE DEVELOPMENT NEEDS ANALYSIS

Blockchain provides a secure and reliable infrastructure for the metaverse. Its transparency ensures that all transactions and interactions within the metaverse are recorded and publicly verifiable, enhancing user satisfaction and experience. The immutability of blockchain preserves original transaction records. Furthermore, blockchain can confirm ownership of users' virtual assets, such as virtual land, goods, and digital currencies, enabling secure transactions in a decentralized environment and enhancing property security, thereby increasing the overall safety and trustworthiness of the metaverse. The demands of the metaverse drive continuous advancements in blockchain technology. The distributed storage and encryption features of blockchain align with the metaverse's need for data security and foster innovative developments in data privacy protection. Additionally, the necessity for interoperability among multiple virtual environments and applications propels the evolution of blockchain's smart contract services and protocol configurations.

4 CONSTRUCTING THE MODEL OF COMMAND SUPPORT SYSTEM

4.1 System Architecture

The command support system model based on blockchain and the metaverse utilizes decentralized and tamper-proof blockchain technology to record operational and logistical data, ensuring information security and transparency. Additionally, it leverages the virtual environment of the metaverse for simulations and exercises to enhance the operational capabilities of individuals and teams. This system model facilitates real-time data sharing and dynamic adjustments, contributing to the optimization of operational strategies while reducing resource consumption and casualty rates, ultimately establishing an efficient and reliable military support system. As shown in Figure 1, the architecture of this model consists of six layers, arranged from bottom to top as follows: foundational layer, data integration layer, network layer, consensus layer, contract layer, and application layer. This model combines the security features of blockchain with the immersive training capabilities of the metaverse, creating a comprehensive and efficient support platform that significantly improves resource utilization and operational effectiveness.

The foundational layer includes the blockchain platform and the metaverse platform. The blockchain platform provides decentralized, tamper-proof, and transparent management of data, ensuring the secure recording and storage of military information. Notably, important data is only uploaded as hashed values without including metadata on-chain.

Additionally, logistics allocation and equipment management operations are recorded in logs to ensure data integrity, facilitating traceability and auditing. Meanwhile, the metaverse platform establishes military simulation systems and data connectivity networks, enabling seamless integration of various intelligence, sensor data, and real-time imagery. This allows for global observation and real-time monitoring in a virtual environment, where troops can simulate various tactical scenarios and conduct live drills to better assess outcomes and make optimal decisions [22]. Furthermore, the metaverse provides a real-time collaborative platform for command teams, enabling close coordination between commanders and operational personnel through technologies such as voice, visual aids, and virtual AI assistants.

The data integration layer is responsible for consolidating data from both the blockchain and the metaverse to support command decision-making. It collects real-time data, including training performance, simulation results, and group competition outcomes, and standardizes this data to enhance its usability and shareability. Additionally, data analysis and visualization are conducted to facilitate intuitive access for commanders, enabling rapid responses.

The network layer employs standardized communication protocols to ensure interoperability and data sharing with the blockchain platform, metaverse platform, and other systems. Security measures such as firewalls and encrypted transmissions are implemented, along with an authentication mechanism to ensure that only authorized users can access the system.

The consensus layer utilizes the Practical Byzantine Fault Tolerance (PBFT) consensus algorithm, allowing the network to function normally and maintain data consistency even in the presence of malicious nodes. Smart contracts are deployed in the contract layer, where code defines the rules and logic for core operations, automating the execution of protocols related to military support. The automated execution of smart contracts enhances system efficiency and reliability, reducing operational costs.

The application layer, or user interface layer, supports multi-device access (computers, tablets, and other mobile devices) to meet operational needs across various scenarios. Its user-friendly interface facilitates information retrieval, allowing users to quickly grasp the current situation. The layer features multifunctional modules for command scheduling, task allocation, supply support, and training communication, enabling users to select relevant modules based on their needs and improve work efficiency. Additionally, it supports real-time feedback on operations, provides data analysis insights, and offers recommendations and guidance to assist decision-making.

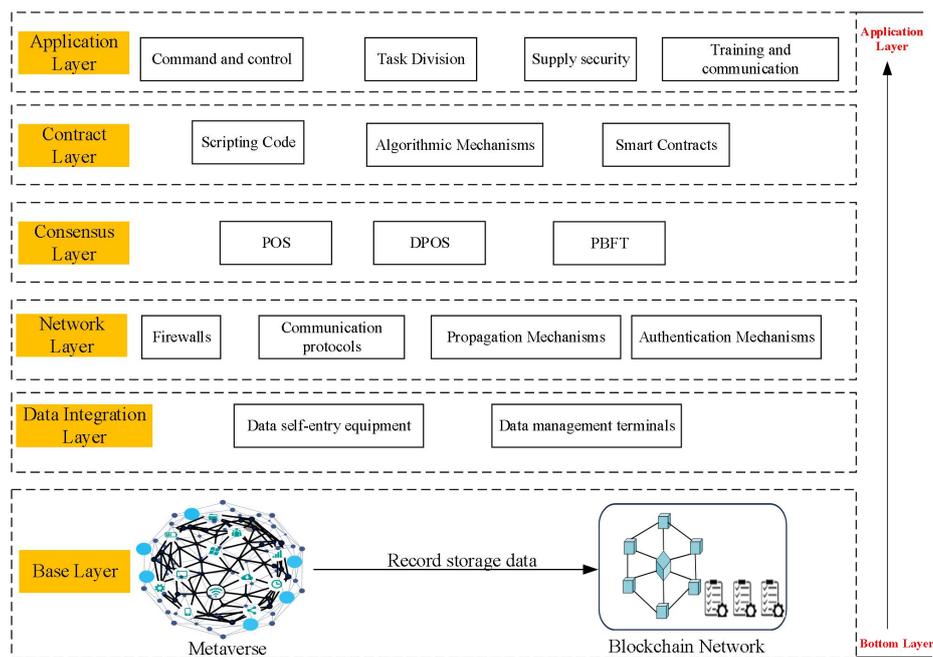


Figure 1 The Architecture of Command Support Model Based on Blockchain and Metaverse

4.2 Alliance Chain Design

Based on an analysis of command support requirements, the blockchain is designed with reference to the characteristics of the body of the information source (logistics department). For intuitive understanding, the business process of equipment supplies is simplified in Figure 2. The supply chain for equipment utilizes a consortium blockchain, with the logistics department serving as the administrator. Nodes in the blockchain network include material producers, transporters, suppliers, quality inspection units, storage units, end-users, and maintenance units, all collaboratively managing and maintaining a traceable system throughout the entire lifecycle of equipment supplies, from production to transmission, distribution, usage, and maintenance.

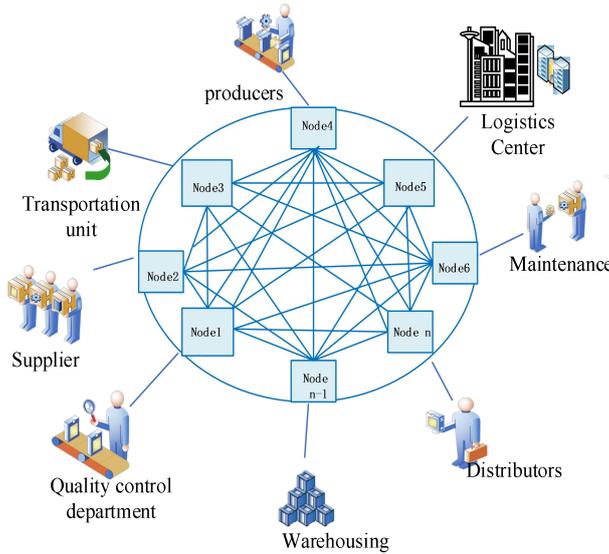


Figure 2 Blockchain-Based Equipment Material Flowchart

As shown in Figure 3, each block consists of a block header and a block body. The block header is the core component, containing elements such as timestamps, hash values, and version numbers. Additionally, the block body contains numerous transactions. This paper describes the changes in transaction data within the block, using equipment supplies as an example to illustrate the formation of a traceable supply chain.

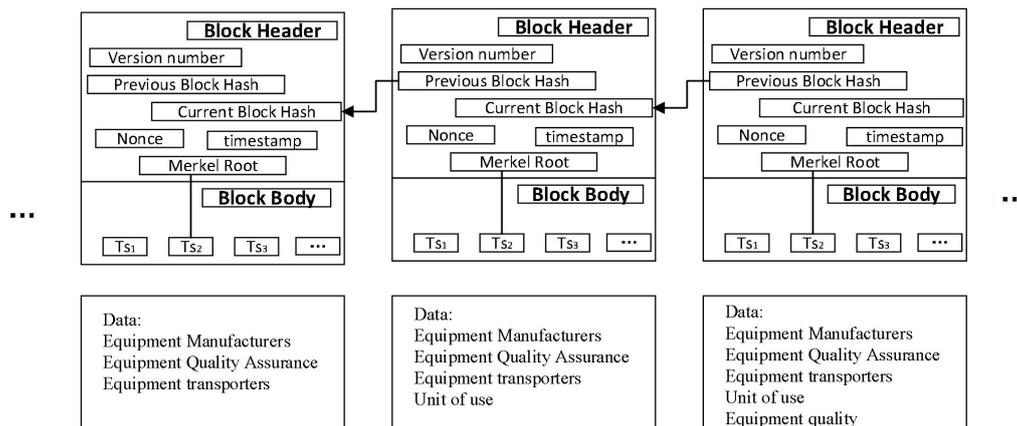


Figure 3 Structure of the Block

4.3 Smart Contracts

In the smart contract code, the necessary protocols and operations are defined. For instance, the automated execution of predefined contracts and rules allows for the allocation of supplies, equipment, and resources to various units, effectively reducing human error, enhancing efficiency, and promoting fairness. Additionally, access control policies are stipulated within the smart contract execution protocols.

Furthermore, when equipment malfunctions or operational errors occur, a device diagnostic request can be sent through the blockchain network. The smart contract then automatically executes the fault diagnosis and report generation process, while also receiving diagnostic responses from suppliers or other non-original supplier nodes. The diagnostic results are updated in real time to reflect the support requirements, and supply and maintenance requests are generated automatically. This automated processing ensures accurate delivery of support needs, thereby improving the efficiency of logistical support and the speed of response.

4.4 Application Scenarios

With the rapid advancement of technology, unmanned combat platforms are playing an increasingly important role in modern warfare [23]. However, these platforms heavily rely on network transmission for information exchange, which can lead to challenges such as information latency, limitations in human-machine interaction, and difficulties in real-time decision-making due to complex environments and network congestion. In this context, metaverse technology offers a more intelligent and efficient solution for unmanned combat operations. By establishing a virtual network

environment in cyberspace and utilizing high-speed internet for data transmission, it enables real-time, efficient information flow. This allows unmanned combat platforms to accurately acquire and process the necessary information, particularly when quick decisions and responses are required during operations. Additionally, metaverse technology enhances operator perception, understanding, and control of unmanned combat platforms through virtual reality devices and intelligent interaction systems, thereby improving combat effectiveness. Furthermore, by integrating data from various sensors and smart devices and conducting real-time analysis, it provides decision-makers with accurate and timely guidance. The decentralized and tamper-proof characteristics of blockchain ensure that all interactions and data are recorded securely. Any information tampering can be immediately detected, resulting in greater security and reliability for unmanned combat platforms. Moreover, the redundancy of data across nodes enhances the platform's robustness and resilience to attacks. The integration of blockchain and metaverse technology offers heightened security to ensure seamless command and realistic virtual environments with intelligent equipment. As both metaverse and blockchain technologies continue to evolve, they are expected to bring further innovations and breakthroughs to unmanned combat operations in the future.

Robust logistical support capabilities are essential for securing victory in warfare. In recent years, the U.S. military has studied the role of big data in the logistics of equipment management [24], while some Chinese scholars have explored the characteristics of logistical support under information conditions and its impact on logistics enhancement. However, issues with data statistics have faced criticism from personnel, highlighting challenges such as poor integration of information systems and lack of standardized protocols [25]. Emerging frontier technologies, such as military metaverse technology and artificial intelligence, empower logistics commanders to obtain real-time supply demands from various units in combat zones. This enables timely adjustments to support plans and facilitates scientific, rapid allocation based on need. By leveraging the extensive metaverse network, information can be rapidly integrated, supply routes can be planned, and logistical efficiency can be optimized [26]. Additionally, within the metaverse, the establishment of military simulation systems and data connectivity networks allows for the seamless integration of various intelligence, sensor data, and real-time imagery. This enables global observation and real-time monitoring in a virtual environment. Troops can simulate various tactical scenarios and conduct real-time drills in the metaverse, leading to better evaluation of outcomes and optimal decision-making [27]. Furthermore, the metaverse provides a real-time collaborative platform for combat command teams, enhancing coordination between commanders and operational personnel. Through technologies such as voice, visual aids, and virtual AI assistants, remote command support can be effectively executed.

5 METAVERSE AND BLOCKCHAIN CONVERGENCE CHALLENGES

5.1 Face Challenges

5.1.1 Immature technologies and key equipment

At the current stage, although blockchain and metaverse technologies exhibit certain potential and advantages in the military domain, further enhancement and development are necessary. For instance, military blockchain encounters scalability and performance issues when facing large-scale transactions. Additionally, in high-real-time scenarios, such as military command and control systems, the characteristics of blockchain, including its consensus mechanism and data transaction confirmation time, impose limitations on real-time and high-frequency trading. Meanwhile, metaverse technology still faces bottlenecks in network bandwidth and computing power, which hinders its ability to meet the demands of large-scale, high-concurrency military applications. Furthermore, military virtual training requires a high level of realism and immersion, yet current technical support for simulating complex environments and interactive objects in the metaverse is insufficient. There remains considerable room for improvement in providing precise and reliable data support for physical simulation and human behavior modeling in complex environments.

5.1.2 Security issues

The integration of military metaverse and blockchain technology also introduces a series of security challenges. Firstly, military data is sensitive, making the assurance of data privacy and security a top priority. This integration involves storing and transmitting a significant amount of sensitive information, including personal identification details, operational plans, and strategic intelligence. Without appropriate measures to protect this sensitive data, user privacy may inadvertently be exposed to anyone accessing blockchain information. Moreover, in the military metaverse environment, a considerable amount of personal behavior data is collected and analyzed to provide customized experiences, which can be easily intercepted by malicious actors. Therefore, it is crucial to safeguard this data from unauthorized access and tampering. Additionally, the confidentiality requirements of military networks and terminals have become a focal point, as the information involved pertains to national security and military secrets, necessitating extremely high security standards. There is also a need to address potential risks of attacks and breaches.

5.1.3 Standardization issues

The integration of military metaverse and blockchain is complex. During integrated joint operations, multiple military blockchain units, networks, and terminals must collaborate. Therefore, it is essential to consider standardization and cross-platform compatibility during the integration process. This approach will facilitate the construction of a new integrated combat platform, creating a highly flexible and configurable simulation environment for future warfare. Such an environment will allow military training and actual combat to align more closely with real-world scenarios, thereby enhancing overall combat capabilities. Establishing unified standards can resolve issues related to the integration of existing equipment with blockchain technology and inconsistencies in data formats. This is vital for ensuring the normal

operation of command control systems, communication systems, and for defending against various network attacks.

5.2 Solution

In advancing the integration and development of military metaverse and blockchain technologies, nations and armed forces can adopt a multifaceted strategy. First, they should enhance talent cultivation by establishing relevant academic programs in higher education institutions that focus on virtual reality, augmented reality, and blockchain. This will strengthen theoretical education and practical training for students. Additionally, military organizations should promote internal professional training to improve personnel's understanding and operational skills regarding these new technologies. Collaborations with enterprises and organizing competitions can also attract more talented individuals to the military metaverse and blockchain sectors. Second, there should be strong support for equipment research and development to encourage technological innovation. Third, enhancing civil-military integration is essential. Lastly, it is crucial to refine the legal and regulatory framework. Through these measures, a powerful and intelligent future military force can be established, ensuring national security while remaining at the forefront of technological advancement.

6 CONCLUSION

This paper designs a military operations support model based on metaverse and blockchain technologies, detailing its architecture, blockchain design, smart contracts, and application scenarios. It then discusses the challenges faced by the integration of military metaverse and blockchain, along with potential solutions. Furthermore, the application potential of blockchain in equipment maintenance is significant, with future research objectives including the reduction of communication and computational overhead, as well as the enhancement of scalability.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

- [1] Hudson Jr J F F. Mission Assurance in Joint All-Domain Command and Control. *Air & Space Power Journal*, 2021, 35: 18-32.
- [2] Wang Y, Su Z, Zhang N, et al. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, 2022, 25(1): 319-352.
- [3] PARK S M, KIM Y G. A Metaverse: Taxonomy, Components, Applications, and Open Challenges. *IEEE Access*, 2022, 10: 4209-4251.
- [4] Wang H, Ning H, Lin Y, et al. A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges. *IEEE Internet of Things Journal*, 2023, 10(16): 14671-14688.
- [5] SANCHEZ J. Second life: An interactive qualitative analysis. *Proceedings of Society for Information Technology and Teacher Education International Conference 2007*, 2007: 1240-1243.
- [6] DIONISIO J D N, III W G B, GILBERT R. 3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities. *ACM Comput. Surv*, 2013, 45(3).
- [7] WANG H, NING H, LIN Y, et al. A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges. *IEEE Internet of Things Journal*, 2023, 10(16): 14671-14688.
- [8] Park S M, Kim Y G. A metaverse: Taxonomy, components, applications, and open challenges. *IEEE access*, 2022, 10: 4209-4251.
- [9] Zhao T, Wu L, Tao J Y, et al. Metaverse Concept and Its Military Application. *Journal of System Simulation*, 2023, 35(7): 1405-1420.
- [10] Solly R, McArdle J. *Unlocking the military potential of the metaverse*, 2023.
- [11] Xu M, Ng W C, Lim W Y B, et al. A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges. *IEEE Communications Surveys & Tutorials*, 2022, 25(1): 656-700.
- [12] SOLLY R, MCARDLE J. *Unlocking the Military Potential of the Metaverse*. 2022.
- [13] COSBY L N. SIMNET: an insider's perspective//*Distributed Interactive Simulation Systems for Simulation and Training in the Aerospace Environment: A Critical Review: Vol. 10280. SPIE*, 1995: 63-76.
- [14] Tucker A, Tucker A. Ender's Wargames: Drones, data and the simulation of war as weapon and tactic. *Virtual Weaponry: The Militarized Internet in Hollywood War Films*, 2017: 145-185.
- [15] GAMAGE K A, WIJESURIYA D I, EKANAYAKE S Y, et al. Online delivery of teaching and laboratory practices: Continuity of university programmes during COVID-19 pandemic. *Education Sciences*, 2020, 10(10): 291.
- [16] JAGATHEESAPERUMAL S K, AHMAD K, AL-FUQAHA A, et al. Advancing education through extended reality and internet of everything enabled metaverses: applications, challenges, and open issues. *arXiv preprint arXiv:2207.01512*, 2022.
- [17] BLOWERS M, JAIMES N, WILLIAMS J. Benefits and challenges of a military metaverse//*Disruptive Technologies in Information Sciences VII. SPIE*, 2023, 12542: 171-180.

- [18] BUHALIS D, LEUNG D, LIN M. Metaverse as a disruptive technology revolutionising tourism management and marketing. *Tourism Management*, 2023, 97: 104724.
- [19] GEORGE A s, FERNANDO M, GEORGE A S, et al. Metaverse: The Next Stage of Human Culture and the Internet. 2021, 08: 1-10.
- [20] FAWKES A, CHESHIRE N. *Military Metaverse CONOPS*, 2023.
- [21] V ALI, A A NORMAN, S R B AZZUHRI. Characteristics of Blockchain and Its Relationship With Trust. *IEEE Access*, 2023, 11: 15364-15374.
- [22] KIM J. Advertising in the Metaverse: Research Agenda. *Journal of Interactive Advertising*, 2021, 21(3): 141-144.
- [23] Zhang P, Zhang C, Gai W. Research on application and development trend of multi-domain cooperative combat for unmanned combat platform//2021 2nd International Symposium on Computer Engineering and Intelligent Communications (ISCEIC). *IEEE*, 2021: 297-301.
- [24] Stanley-Lockman Z. Revisiting the revolution in military logistics: Technological enablers twenty years on. *Disruptive and Game Changing Technologies in Modern Warfare: Development, Use, and Proliferation*, 2020: 197-222.
- [25] Pawelczyk M. Contemporary challenges in military logistics support. *Security and Defence Quarterly*, 2018, 20(3): 85-98.
- [26] Solly R, McArdle J. *Unlocking the military potential of the metaverse*, 2023.
- [27] Zhang Z, Guo Y, Zhao X, et al. *Military Metaverse: Key Technologies, Potential Applications and Future Directions*. *Journal of System Simulation*, 2023, 35(7): 1421-1437.

ENHANCED UNSUPERVISED IMAGE-TO-IMAGE TRANSLATION USING CONTRASTIVE LEARNING AND HISTOGRAM OF ORIENTED GRADIENTS

WanChen Zhao, WenHan Wang, Cheng Zhang, XiaoLei Qu*
College of Instrument Science and Optoelectronic Engineering, Beihang University, Beijing 100191, China.
Corresponding Author: XiaoLei Qu, Email: quxiaolei@buaa.edu.cn

Abstract: Image-to-Image Translation is a vital area of computer vision that focuses on transforming images from one visual domain to another while preserving their core content and structure. However, this field faces two major challenges: first, the data from the two domains are often unpaired, making it difficult to train generative adversarial networks effectively; second, existing methods tend to produce artifacts or hallucinations during image generation, leading to a decline in image quality. To address these issues, this paper proposes an enhanced unsupervised image-to-image translation method based on the Contrastive Unpaired Translation (CUT) model, incorporating Histogram of Oriented Gradients (HOG) features. This novel approach ensures the preservation of the semantic structure of images, even without semantic labels, by minimizing the loss between the HOG features of input and generated images. The method was tested on translating synthetic game environments from GTA5 dataset to realistic urban scenes in cityscapes dataset, demonstrating significant improvements in reducing hallucinations and enhancing image quality.

Keywords: Image-to-image translation; Photorealism; GANs

1 INTRODUCTION

Image-to-Image Translation has emerged as a pivotal area of research within computer vision and machine learning, focusing on the intricate task of transforming images from one visual domain to another while preserving their inherent content and structure. This capability has wide-ranging applications, including but not limited to converting sketches into lifelike photographs [1], colorizing grayscale images [2], and altering scenes from day to night [3]. The overarching goal of image-to-image translation is to learn a robust mapping between two distinct visual domains, enabling seamless transitions that retain the essence of the original imagery.

Generative Adversarial Networks (GANs) have become the cornerstone of this transformation process, offering a powerful deep learning framework tailored for image generation tasks. Early breakthroughs in this field predominantly relied on supervised learning with paired datasets, where each image in the source domain has a corresponding counterpart in the target domain. A seminal work in this domain is pix2pix [4], which employs conditional GANs to learn a direct mapping from input images to output images by leveraging paired training data. Despite its effectiveness, this approach is constrained by the necessity of paired datasets, which are often difficult, time-consuming, and costly to obtain, particularly in complex or large-scale applications.

To circumvent the challenges associated with paired datasets, researchers have developed innovative methods that facilitate translation between unpaired datasets. CycleGAN [5] stands out as a landmark approach, utilizing dual generators (G and F) and dual discriminators. Generator G transforms images from style X to style Y, while generator F performs the reverse operation. The discriminators are tasked with differentiating between real and generated images in both styles. Given the absence of paired data, adversarial loss alone cannot guarantee accurate translation, prompting the introduction of a cycle-consistency loss. This loss ensures that an image can be reconstructed after passing through both generators (i.e., $F(G(X)) \approx X$ or $G(F(Y)) \approx Y$), thereby maintaining the content and structural integrity of the images. While CycleGAN has proven to be a versatile and effective solution, it is not without its limitations, including the high computational burden of training two sets of generators and discriminators and the tendency to produce artifacts in the generated images.

The inherent weakness of cycle-consistency loss, which can lead to image artifacts or hallucinations, has spurred the development of more refined techniques. StyleGAN [1], for instance, introduces a mapping network that projects the input latent vector into a style vector space, allowing for fine-grained control over the generative features and significantly reducing instability during training. This approach not only enhances the aesthetic quality of the generated images but also allows for greater manipulation of the image's stylistic elements. Sem-GAN [6] further advances this concept by integrating semantic labels into the training process, ensuring that the generated images are both visually convincing and semantically consistent, thus addressing the challenge of maintaining meaningful content across translations.

Attention mechanisms have also played a crucial role in advancing image-to-image translation by addressing the challenge of capturing high-level semantic information. Attention GAN [7], for example, uses attention-guided generators to create attention masks, which are then combined with the input image to generate high-quality target images. This selective focus on critical image regions helps to preserve important details and improve the overall

translation quality. Conversely, SPA-GAN [8] enhances the discriminator's ability to discern fine details by incorporating attention mechanisms, which in turn improves the generator's capacity to produce more realistic images. In addition to these methods, Unsupervised Image-to-Image Translation (UNIT) [9] leverages a shared latent space assumption, mapping images from different domains into a common feature space. This approach provides an effective solution for more complex unsupervised cross-domain image translation tasks, enabling the model to learn shared features that are domain-invariant. Similarly, CUT [10] employs contrastive learning to maximize the similarity between source and target domain images in the latent space while preserving their distinct characteristics. This technique has proven particularly effective in maintaining the unique aspects of each domain while ensuring a coherent translation.

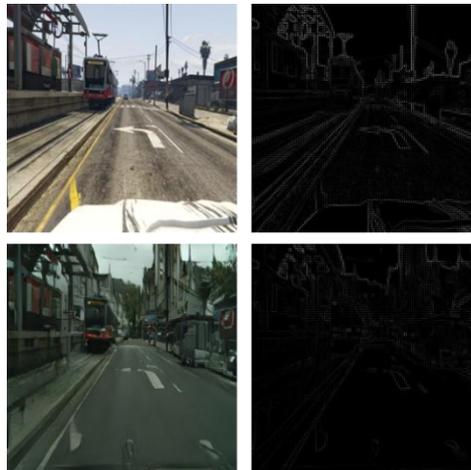


Figure 1 Generation of Image Hallucinations and Their HOG Feature Maps. The First Row Shows the Original Images along with their HOG Feature maps, while the Second Row Displays the Images Generated by the CUT Model and their Corresponding HOG Feature Maps

Despite the significant advancements achieved by these methods, several challenges persist. For instance, while the translation quality of the main subject in an image is often high, the quality of background translation can be inconsistent, leading to unwanted artifacts or hallucinations. This issue is particularly problematic in photorealism tasks, such as translating game screenshots into real-world images, where even minor semantic errors—such as misinterpreting the sky as a building—can result in glaring inaccuracies. Moreover, the reliance on semantic supervision poses a practical limitation, as many real-world datasets lack the necessary semantic labels, rendering these approaches less effective in unsupervised settings.

To address these challenges, we have developed a novel unsupervised image-to-image translation method based on the CUT model, incorporating a unique loss function grounded in Histogram of Oriented Gradients (HOG) features [11]. HOG features divide an image into small cells and compute histograms of gradient directions within each cell, providing robustness to variations in lighting and geometric transformations. By calculating the HOG features of both the input and generated images and computing the L2 loss between them, our method ensures that the semantic structure of the generated image is preserved, even in the absence of semantic label supervision. This approach not only mitigates the issue of hallucinations but also enhances the overall quality and fidelity of the translated images.

To validate the effectiveness of our proposed method, we conducted a series of experiments on image translation from GTA5 [12] to Cityscapes [13], a challenging task that involves translating synthetic game environments into realistic urban scenes. The experimental results, supported by comprehensive visualizations, demonstrate that our method effectively reduces hallucinations in image generation, significantly improving the quality and realism of the translated images. Our approach represents a meaningful step forward in unsupervised image-to-image translation, offering a robust solution to some of the most persistent challenges in this field.

2 RELATED WORK

2.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) [14], introduced by Goodfellow et al. in 2014, have revolutionized the field of generative modeling by presenting a novel framework where two neural networks, a generator and a discriminator, engage in a competitive training process. The generator aims to create images that are indistinguishable from real data, while the discriminator's goal is to differentiate between real and generated images. This adversarial interplay leads to the generator producing increasingly realistic images, and the discriminator becoming more adept at detecting fakes. The success of this dynamic has led to the development of numerous GAN variants, each designed to address specific challenges or enhance the model's capabilities.

Among the earliest significant improvements was the Deep Convolutional GAN (DCGAN) [15], which introduced a more stable architecture for GANs by leveraging convolutional layers and replacing deterministic pooling functions

with strided convolutions. This architectural shift not only stabilized training but also improved the visual quality of the generated images, making DCGANs a foundational model for many subsequent GAN-based studies.

Conditional GANs (cGANs) [16], another pivotal advancement, extended the GAN framework by incorporating auxiliary information such as class labels, enabling the generation of images conditioned on specific input data. This capability has made cGANs highly effective for tasks requiring controlled image synthesis, such as image-to-image translation, where the generated image must adhere to specific characteristics dictated by the input.

StyleGAN [1], developed by Karras et al., represents a significant leap forward in GAN research. By introducing a new generator architecture that allows for the control of style at various levels of the synthesis process, StyleGAN enables the fine-tuning of generated images in ways that were previously impossible. This model has set new standards for photorealism in image synthesis, and its innovations, such as the mapping network and style mixing regularization, have been widely adopted in subsequent works.

Research into GANs has also delved deeply into the theoretical underpinnings and optimization strategies to overcome challenges like mode collapse and unstable training dynamics. Techniques such as Wasserstein GANs (WGANs) [17], which use the Earth Mover's Distance (Wasserstein distance) as a loss function, have provided more robust convergence properties and reduced the risk of mode collapse. Additionally, the incorporation of spectral normalization has been shown to stabilize the training of GANs by constraining the Lipschitz constant of the discriminator, further contributing to the reliability of GAN training processes [18].

Furthermore, GANs have found extensive application in the domain of image-to-image translation, significantly advancing the capabilities of this field. For instance, GAN-based models have been employed in translating low-resolution images to high-resolution counterparts [19], enhancing the quality and details in tasks like super-resolution. Additionally, GANs have been pivotal in domain adaptation, where models are trained to translate images from one domain (e.g., sketches) to another (e.g., photorealistic images) while preserving essential content. Their flexibility has also been demonstrated in cross-modal translation tasks, such as converting grayscale images to color [1] or transforming aerial images into map-like representations [20]. The versatility of GANs in handling diverse image translation tasks underlines their profound impact on both academic research and practical applications, making them a cornerstone technology in the advancement of computer vision.

2.2 Image-to-Image Translation

Image-to-Image Translation is a critical research area in computer vision, focused on transforming an image from one domain into another while preserving its core content and structure. The field has been significantly advanced by the application of Generative Adversarial Networks (GANs), with the pix2pix framework by Isola et al. serving as a landmark model. Pix2pix demonstrated the power of conditional GANs (cGANs) in supervised image translation tasks, where paired datasets—each containing an input and a corresponding target image—are used to train the model. This framework has been particularly successful in tasks such as image inpainting, where missing parts of an image are filled in, and in style transfer, where an image's artistic style is altered while maintaining its underlying structure.

However, the reliance on paired datasets poses significant limitations, as collecting such data can be challenging in many scenarios. To address this, Zhu et al. proposed CycleGAN, which introduced cycle consistency as a key concept to enable translation between two domains without requiring paired data. CycleGAN's ability to learn mappings between domains using unpaired datasets has made it widely applicable in various tasks, including artistic style transfer and domain adaptation, where direct correspondences between images are unavailable.

Following CycleGAN, numerous methods have been developed to enhance the performance and applicability of image-to-image translation models. For example, DiscoGAN [21] and DualGAN [22] introduced additional constraints, such as identity loss and dual learning mechanisms, to improve the stability and quality of translations in unpaired settings. These models have contributed to a better understanding of how to maintain content consistency while achieving the desired stylistic transformation.

Multi-Modal UNsupervised Image-to-Image Translation (MUNIT) [23] and StarGAN [24] represent significant strides in handling multi-domain and multi-modal translation. MUNIT allows for the separation of content and style, enabling the generation of diverse outputs from a single input image by recombining different content and style codes. This flexibility has been particularly useful in applications where multiple plausible outputs exist for a given input, such as in image editing and creative content generation. StarGAN, on the other hand, unifies the process of image translation across multiple domains within a single model, providing a scalable solution for tasks requiring multi-domain transformations.

Recent innovations like the Contrastive Unpaired Translation (CUT) model [10] have introduced contrastive learning into the image-to-image translation pipeline. CUT maximizes mutual information between the input and output domains, enabling high-quality translation even when only a single image from each domain is available. This approach has proven particularly effective in one-sided translation tasks, where traditional methods might struggle due to limited data availability.

Moreover, Stephan R. Richter et al.'s work on utilizing intermediate representations for adversarial training has highlighted the importance of perceptual supervision at multiple levels within a neural network [25]. By enforcing adversarial constraints at different perceptual layers, this method has achieved state-of-the-art results in tasks requiring fine-grained control over the translation process, such as semantic segmentation and high-resolution image synthesis.

In addition to these advancements, researchers have also explored ways to improve the interpretability and control of image-to-image translation models. Approaches like disentanglement learning, where models learn to separate different factors of variation within the data, have allowed for more precise control over the generated outputs [26]. This has opened up new possibilities for user-guided image translation, where specific aspects of the image can be altered according to user-defined parameters, thereby enhancing the practicality and usability of these models in real-world applications.

These developments illustrate the ongoing progress in image-to-image translation research, driven by the continuous refinement of GAN-based methods and the exploration of new learning paradigms that address both practical challenges and theoretical concerns in the field.

3 METHOD

Our overall architecture is based on the structure of CUT. Building upon the CUT model, we further introduce HOG feature loss to supervise the training of the generator. By leveraging the robustness of HOG features against style variations, we ensure that the generated realistic images remain consistent with the original simulated images in terms of content and spatial structure, thereby achieving a better realistic effect for the simulated images. A schematic diagram of this structure is shown in Figure 2.

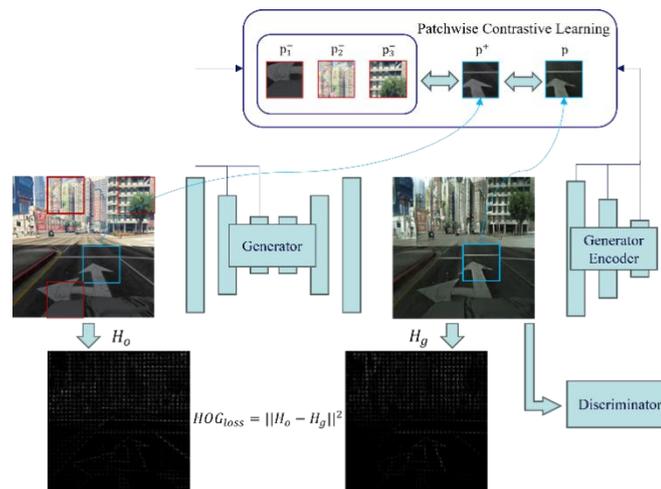


Figure 2 The Architecture of our Proposed Method

3.1 CUT Model

CUT is specifically designed to address the challenges of unpaired image translation, especially when paired samples are unavailable. Traditional unpaired image translation methods often rely on cycle consistency loss, which necessitates translating images back and forth between the source and target domains. However, this approach has limitations in preserving image details and content structure and incurs significant computational costs. CUT overcomes these challenges by introducing contrastive learning and a one-sided translation strategy.

The core idea of CUT is to preserve content consistency by maximizing the mutual information between corresponding local regions (patches) of the input and output images. Specifically, CUT employs a contrastive loss mechanism, which pulls positive samples (corresponding patches) closer together while pushing negative samples (non-corresponding patches) farther apart. This ensures that the generated images retain the local features of the original images. To achieve this, CUT introduces a patch-based discriminator that evaluates local regions of the image rather than the image as a whole. This allows the model to learn domain translation at a global scale while maintaining fine-grained details at a local scale.

In its one-sided translation strategy, CUT focuses solely on translating from the source domain to the target domain, eliminating the need for reverse mapping. This strategy significantly reduces computational overhead and avoids potential artifacts introduced by cycle consistency constraints. Additionally, CUT employs a technique known as "contrastive self-supervision," which enhances model stability and translation quality by introducing self-supervised mechanisms in the absence of paired data. This allows CUT to generate high-quality translations even with unidirectional datasets.

The CUT method is particularly suitable for scenarios where paired data is difficult to obtain, such as artistic style transfer and medical image processing. In artistic style transfer, CUT can preserve the content of the original image while achieving nuanced style transformations. In medical imaging, it can effectively enhance low-quality images to high-quality ones, playing a critical role in medical diagnosis.

Overall, the CUT method, by combining contrastive learning, one-sided translation, and self-supervision mechanisms, significantly enhances the effectiveness of unpaired image translation. Its robust ability to retain local details and efficient computational performance make it a highly promising tool in the field of image-to-image translation.

3.2 HOG Loss

As depicted in Figure 1, while the CUT model utilizes a patch-based discriminator to ensure that the images generated by the generator retain fine details and achieve overall stability, the problem of hallucination—where the generated images may contain unrealistic or extraneous features—still persists. To address this issue and further tighten the constraints on the generated images, we have incorporated HOG (Histogram of Oriented Gradients) loss into the training process of the generator.

The introduction of HOG loss is strategically aimed at harnessing the robustness of HOG features, which are particularly effective in capturing and preserving structural details and edge orientations despite variations in style. By aligning the HOG features of the generated images with those of the input images, this loss function promotes structural invariance between the source and target domains. This alignment helps in mitigating hallucinations, ensuring that the generated images are not only stylistically transformed but also maintain the structural integrity of the original content. HOG loss operates by comparing the HOG features extracted from both the generated and reference images. It encourages consistency in these features, penalizing deviations that could result in visual artifacts or inconsistencies. This method enhances the generator’s capability to produce more accurate and realistic outputs, thereby reducing the likelihood of hallucinated elements.

The calculation of HOG loss is as follows:

$$\text{HOG}_{\text{loss}} = \left\| \left\| H_o - H_g \right\| \right\|^2 \# \quad (1)$$

where H_o denotes the HOG features of the original images and H_g denotes the HOG features of the generated images.

4 EXPERIMENT AND RESULTS

4.1 Dataset

We conduct our experiments using the following two datasets:

The GTA5 dataset [12] is a synthetic dataset tailored for computer vision tasks. It is derived from the popular video game Grand Theft Auto V (GTA V), where the in-game urban environment is utilized to simulate real-world street scenes. This dataset offers high-resolution images with pixel-level annotations that include various urban objects such as buildings, vehicles, roads, and sidewalks. The Cityscapes dataset [13] is a real-world dataset specifically curated for visual understanding tasks in urban settings. It is extensively used in autonomous driving research. This dataset comprises high-resolution images captured from 50 different cities across Germany, with detailed annotations covering 30 classes, 19 of which are commonly employed for evaluation. Both datasets consist of 2,500 images each.



Figure 3 Illustration of the Datasets. The First Row Represents the GTA5 Dataset, while the Second Row Depicts the Cityscapes Dataset

4.2 Evaluation Metrics

In this paper, we use three evaluation metrics—the Inception Score (IS) [27], Kernel Inception Distance (KID) [28], and Fréchet Inception Distance (FID) [29]—to assess the realism of generated images. Each of these metrics provides unique insights into different aspects of image generation performance.

The Inception Score (IS) evaluates both the clarity and diversity of generated images. It measures how well the generated images can be classified into distinct categories and how diverse these generated samples are. Specifically, IS uses a pre-trained Inception v3 model [30] to compute the Kullback-Leibler (KL) divergence between the conditional

label distribution $p(y | x)$ for each generated image x and the marginal label distribution $p(y)$ across multiple generated images. A higher IS indicates that the images are not only of high quality but also exhibit a diverse range of classes.

$$IS = \exp\left(\frac{1}{N} \sum_{i=1}^N \text{KL}(p(y|x_i) || p(y))\right) \# \quad (2)$$

Where KL denotes the Kullback-Leibler divergence and N denotes the number of samples.

The Kernel Inception Distance (KID) measures the similarity between the feature distributions of real and generated images. Unlike IS, which focuses on classification, KID uses a kernel function to compare the distributions of features extracted from a pre-trained Inception model. The metric calculates the Maximum Mean Discrepancy (MMD) between these feature distributions, providing an unbiased estimate of the distance between the real and generated image distributions. A lower KID score suggests that the generated images closely match the real image distribution in feature space.

$$KID = \frac{1}{M^2} \sum_{i=1}^N \sum_{j=1}^M \varphi(x_i)\varphi(x_j) - \frac{2}{MN} \sum_{i=1}^N \sum_{j=1}^M \varphi(x_i)\varphi(y_j) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^M \varphi(y_i)\varphi(y_j) \# \quad (3)$$

Where φ represents the feature map, and M and N denote the number of real and generated samples, respectively.

The Fréchet Inception Distance (FID) assesses the distance between the feature distributions of real and generated images. It calculates the Fréchet distance (or Wasserstein distance) between the Gaussian distributions fitted to the feature vectors of both real and generated images. The metric takes into account both the mean and covariance of these distributions. A lower FID score indicates that the generated images have feature distributions that are more similar to those of real images, reflecting higher image quality and better alignment with the real image distribution.

$$FID = \left\| \mu_r - \mu_g \right\|_2^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}) \# \quad (4)$$

Where μ_r and μ_g are the means, and Σ_r and Σ_g are the covariances of the real and generated image features, respectively.



Figure 4 Visual Comparison with Different State-of-the-art Methods in the GTA5-to-Cityscapes Translation Experiment. From Left to Right, the Sequence Displays the Original Image, Followed by the Results of CycleGAN, SRC, TSIT, AttentionGAN, CUT, and our Proposed Method

4.3 Implementation Details

In this study, we utilized the PyTorch [31] deep learning framework to train the network on a single Nvidia A100 GPU (80G). The models were trained with a batch size of 16 and a learning rate of 0.0002. We employed the Adam optimizer, known for its effective learning rate adaptation, to enhance convergence and stability. The training was conducted over 200 epochs to ensure thorough learning and model refinement. The weight of the proposed HOG loss (λ_{HOG}) was set to 10, balancing the contribution of HOG features in the loss function and improving feature representation by emphasizing gradient-based features. When computing the HOG features of the image, we set the window size to 256 by 256 pixels, the block size to 6 by 6 pixels, the block stride to 2 by 2 pixels, the cell size to 2 by 2 pixels, and the number of bins to 9. These settings were chosen to ensure a clear and accurate representation of the HOG features. All input images were resized to 256×256 pixels to ensure consistency across the dataset, standardizing the input dimensions for uniform training and evaluation. In alignment with the experimental setup for photorealism, we generated images in the Cityscapes style from GTA5 game images. We divided 2,000 images from each dataset as the training set and 500 images as the test set.

4.4 Comparison with State-of-the-art Methods

Table 1 Statistical Comparison with State-of-the-art Methods

	IS	KID	FID
GTA5	-	0.1350	153.0678
CycleGAN	2.6099	0.0813	116.2996
SRC	2.7234	0.0646	98.9844
TSIT	2.6935	0.0477	88.7268
Attention GAN	2.7617	0.0411	82.2881
CUT	2.3735	0.0321	69.2619
Ours	2.6870	0.0310	65.7156

To validate the effectiveness of our proposed method, we performed a comparative analysis using several widely adopted networks for image-to-image translation tasks, including CycleGAN [5], SRC [32], TSIT [33], Attention GAN [7] and CUT [10]. These methods are open-source and have been extensively validated.

As shown in Table 1, our method demonstrates superior performance in terms of the IS metric, which assesses the clarity and diversity of the generated images. Additionally, our method achieved results of 0.0310 for the KID metric and 65.7156 for the FID metric, both of which measure the distance between generated images and the real domain. These results surpass those of all other methods, demonstrating that our approach excels in the task of photorealism compared to other image-to-image translation methods.

4.5 The Visualization Results

As shown in Figure 4, GTA5 represents the original image (a game screenshot), while CycleGAN, SRC, TSIT, Attention GAN, CUT, and Ours represent the images generated by each respective method. It is evident that our method outperforms the others in terms of detail accuracy and correctness in the generated targets. Our method avoids the screen tearing or color distortion that appears in images produced by other methods. Notably, in some methods, the front of the car either disappears or is transformed to blend with the ground, and in the sky, some methods generate vegetation instead. Our method, however, does not exhibit these critical errors, resulting in significantly higher quality images.

5 CONCLUSION

In this paper, we propose an image-to-image translation method based on HOG feature loss, achieving state-of-the-art results in experiments aimed at making GTA5 images appear more realistic within the Cityscapes dataset. Evaluation metrics and visual results demonstrate that our method not only enhances the quality of generated images but also significantly reduces the artifacts commonly seen in other approaches. Moreover, our method exhibits robustness across various challenging scenarios, highlighting its potential for broader applications in image translation tasks. The reduction in visual artifacts not only improves the aesthetic quality but also ensures greater consistency with real-world data. Future work could explore extending this method to a wider range of image-to-image translation tasks, while further enhancing the quality and accuracy of the generated images.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Richardson E, Alaluf Y, Patashnik O, et al. Encoding in style: a stylegan encoder for image-to-image translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 2287-2296.
- [2] Kiani L, Saeed M, Nezamabadi-pour H. Image colorization using generative adversarial networks and transfer learning. *2020 International Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2020: 1-6.
- [3] Li X, Guo X. SPN2D-GAN: semantic prior based night-to-day image-to-image translation. *IEEE Transactions on Multimedia*, 2022, 25: 7621-7634.
- [4] Isola P, Zhu J Y, Zhou T, Efros A A. Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 5967-5976.
- [5] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2223-2232.
- [6] Cherian A, Sullivan A. Sem-GAN: Semantically-consistent image-to-image translation. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019: 1797-1806.
- [7] Tang H, Liu H, Xu D, et al. Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 34(4): 1972-1987.
- [8] Emami H, Aliabadi M M, Dong M, et al. SPA-GAN: Spatial attention GAN for image-to-image translation. *IEEE Transactions on Multimedia*, 2020, 23: 391-401.
- [9] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 2017, 30.
- [10] Han J, Shoeiby M, Petersson L, et al. Dual contrastive learning for unsupervised image-to-image translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 746-755.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, 1: 886-893.
- [12] Richter S R, Vineet V, Roth S, Koltun V. Playing for data: Ground truth from computer games. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.
- [13] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 3213-3223.
- [14] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2014, 27.
- [15] Radford A. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [16] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8798-8807.
- [17] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs. *Advances in Neural Information Processing Systems*, 2017, 30.
- [18] Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
- [19] Zhang B, Gu S, Zhang B, et al. Styleswin: Transformer-based GAN for high-resolution image generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 11304-11314.
- [20] Wyawahare M, Ekbote N, Pimperkhede S, et al. Conversion of Satellite Images to Google Maps Using GAN. *International Conference on Innovations in Computational Intelligence and Computer Vision*. Singapore: Springer Nature Singapore, 2022: 103-117.
- [21] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks. *Proceedings of the International Conference on Machine Learning*. PMLR, 2017: 1857-1865.
- [22] Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised dual learning for image-to-image translation. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2849-2857.
- [23] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 172-189.
- [24] Choi Y, Choi M, Kim M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8789-8797.
- [25] Richter S R, AlHajja H A, Koltun V. Enhancing photorealism enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 1700-1715.
- [26] Gonzalez-Garcia A, Van De Weijer J, Bengio Y. Image-to-image translation for cross-domain disentanglement. *Advances in Neural Information Processing Systems*, 2018, 31.
- [27] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, 2016, 29.
- [28] Bińkowski M, Sutherland D J, Arbel M, et al. Demystifying MMD GANs. *arXiv preprint arXiv:1801.01401*, 2018.
- [29] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems*, 2017, 30.

- [30] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [31] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 2019, 32: 8024-8035.
- [32] Jung C, Kwon G, Ye J C. Exploring patch-wise semantic relation for contrastive learning in image-to-image translation tasks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 18260-18269.
- [33] Jiang L, Zhang C, Huang M, et al. TSIT: A simple and versatile framework for image-to-image translation. Proceedings of the European Conference on Computer Vision (ECCV), 2020: 206-222.

STUDY ON EQUIVALENT STATIC WIND LOAD ON LONG-SPAN ROOF BASED ON RESPONSE-RELATED SCALE

Yi Zhu^{1,2}, ZhiCheng Bai^{1,2*}, XiaoXia Zhao^{1,2}, YinGuang Wang^{1,2}, Xu Zhang^{1,2}

¹China Construction Fourth Engineering Bureau Co., Ltd. Guangzhou 510000, Guangdong, China.

²China Construction Fourth Engineering Bureau Sixth Construction Co., Ltd. Hefei 230000, Anhui, China.

Corresponding Author: ZhiCheng Bai, Email: 19822660870@163.com

Abstract: This paper proposes an equivalent static wind load calculation method based on response-related scale for the wind load calculation of large-span roof structures. The effectiveness of this method is verified by wind tunnel tests and finite element analysis of the roof of the east entrance of Hangzhou Convention and Exhibition Center. The research results show that the equivalent static wind load method based on response-related scale has high accuracy in reflecting the overall response characteristics of the structure, and its error in calculating the node displacement response is the smallest, only 4.6%, which is better than the traditional single-objective equivalent method and least squares method.

Keywords: Long-span roof; Static wind load; Response-related scale; Equivalent load; Hangzhou convention and exhibition center

INTRODUCTION

With the acceleration of urbanization and economic development, large-scale exhibitions play an increasingly important role in modern urban construction [1,2]. As the core area of Hangzhou Airport Economic Demonstration Zone, Hangzhou Convention and Exhibition Center project is not only the cover of the airport and an important window of Hangzhou city, but also a convention and exhibition base with complete facilities in the Yangtze River Delta region. In the design of large-scale roof structures of exhibitions, the accurate calculation of wind loads has an important impact on structural safety and economy. The traditional static wind load calculation method has certain limitations in large-span roof structures and it is difficult to fully reflect the complex response characteristics of the structure [3,4]. To this end, this study proposes an equivalent static wind load calculation method based on response-related scale to improve the accuracy and reliability of wind load calculation of large-span roof structures.

1 PROJECT OVERVIEW

The Hangzhou Convention and Exhibition Center project is located in Nanyang Street, Xiaoshan District, Hangzhou City, Zhejiang Province. It is located on the bank of Qiantang River and close to the Qiantang River estuary. Hangzhou Metro Line 1 crosses from the central corridor from east to west. The total land area of the project is about 740,000 m², and the total construction area is about 1.34 million m², of which the underground construction area is about 500,000 m² and the indoor net exhibition area is about 300,000 m². The volume ranks first in Zhejiang Province and fifth in the country. Hangzhou convention and exhibition center renderings can be seen in Figure 1.

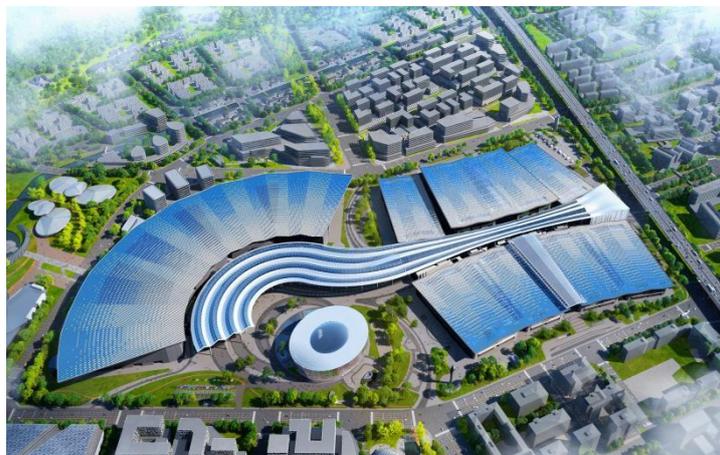


Figure 1 Hangzhou Convention and Exhibition Center renderings

2 RESEARCH METHODS

In order to calculate the equivalent static wind load of the large-span roof, a calculation method based on the response-related scale is proposed. The equivalent target that best reflects the overall effect of the structure is selected

through the influence range of the response of a certain node of the structure. On this basis, the equivalent static wind load of the structure is determined by the dynamic load response correlation method. Taking the roof of the east entrance of the Hangzhou Convention and Exhibition Center as an example, a rigid model pressure wind tunnel test was carried out, and the equivalent static wind load was calculated by different methods.

2.1 Dynamic Load Response Correlation Method

The motion equation of the large-span spatial structure based on the node degree of freedom is formula (1):

$$[Q]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{F(t)\} \quad (1)$$

Where, $[Q]$ 、 $[C]$ 、 $[K]$ are the mass matrix, damping matrix and stiffness matrix of the corresponding node degree of freedom respectively; $\{x\}$ 、 $\{\dot{x}\}$ 、 $\{\ddot{x}\}$ are the displacement, velocity and acceleration of each node degree of freedom respectively; and $\{F(t)\}$ are the wind load time history of each node degree of freedom respectively.

Define $[K]\{x\}$ in formula (1) as the structural wind vibration force, which comprehensively reflects the combined effect of wind load, damping force and inertia force. From the definition, it can be seen that the structural wind vibration force and the structural displacement response are linearly related. The response of the large-span spatial structure can be regarded as a quasi-static response under the action of the structural wind vibration force. Therefore, after obtaining the response time history of each degree of freedom of the structure through random vibration analysis, the equivalent static wind load of the large-span roof structure can be obtained based on the structural wind vibration force using the LRC method. Since the equivalent wind vibration force contains the structural dynamic characteristics, this method is called the dynamic load response correlation method.

The equivalent static wind load in the modal space can be expressed as formula (2)-formula (4):

$$F_{eq} = \bar{F} + \tilde{F}_{eq} \quad (2)$$

$$\tilde{F}_{eq} = g \sum_j W_j \tilde{F}_{eq,j} \quad (3)$$

$$\tilde{F}_{eq,j} = \omega_j^2 [Q] \sigma_{qj} \varphi_j \quad (4)$$

$$W_j = \frac{\sum \rho_{q,ij} \sigma_{Tq,i} \varphi_j}{\sigma_T} \quad (5)$$

Where: F_{eq} is the equivalent static wind load of the structure; \bar{F} is the average wind load of each degree of freedom; \tilde{F}_{eq} is the pulsating equivalent static wind load; g is the peak factor, which can be taken as 2.5 for the main structure; $\tilde{F}_{eq,j}$ is the root mean square of the j -th order pulsating wind load; W_j is the participation coefficient of the j -th order wind load; ω_j is the i -th order natural circular frequency of the structure; φ_j is the i -th order vibration mode; σ_{qj} is the i -th order generalized displacement root mean square; $\rho_{q,ij}$ is the correlation coefficient between the i -th order modal response and the j -th order modal response; σ_T is the root mean square of the equivalent target; $\sigma_{Tq,i}$ is the i -th order root mean square value of the equivalent target; $\sum(\cdot)$ is the summation symbol.

To calculate the modal wind load participation coefficient of formula (5), the equivalent target must be selected first. Equivalent targets are generally selected based on engineering needs and experience, such as structural displacement or base force response. For large-span spatial structures, due to the complexity of structural response, there are often more than one representative equivalent target. When multiple equivalent targets are selected, the equivalent target can be replaced by a consistent equivalent target. The consistent equivalent target is expressed as a linear combination of all equivalent targets, namely:

$$T = I^T \{T_i\} \quad (6)$$

Where: T is the consistent equivalent target; $\{T_i\}$ is the column vector composed of each equivalent target; $(\cdot)^T$ represents transposition, I represents the distribution coefficient vector of each equivalent target, calculated as follows:

$$I = [P_{ij}]^{-1} \{1\} \quad (7)$$

Where $[P_{ij}]$ is the correlation coefficient matrix of each equivalent target time series.

2.2 Response Related Scale

The response related scale reflects the affected size of the response at a specified position under random load. Since the vertical structural size of the roof structure is much smaller than the plane size, the area can be expressed as follows.

$$E_i = \sum_j |\rho_{ij}| M_j \quad (8)$$

Where: E_i represents the response correlation scale of the i node position; ρ_{ij} represents the correlation coefficient of the structural response at i and j ; M_j represents the node subordinate area at i position, $|\cdot|$ represents the absolute value. Since the responses at different positions may have a negative correlation, in order to ensure that the affected area is a positive value, the absolute value of the correlation coefficient is used when calculating the response correlation

scale.

From the definition of E_i , it can be seen that the larger E_i , the response at i position is, the higher the correlation between the response at other positions is, and the stronger the integrity of the response at i position is. Calculating the equivalent static wind load of the structure with the response at the maximum position as the equivalent target can reflect the overall characteristics of the large-span space structure. For large-span space structures, since the structure is in an over-static form, the roof may have multiple local extreme values. For such cases, the equivalent static wind load should be calculated with the responses at all local extreme value locations as multiple targets.

3 ENGINEERING EXAMPLES

3.1 Wind Tunnel Test

The overall layout of the exhibition hall of the Grand Convention and Exhibition Center is in the form of "fishbone + semi-enclosed" and the roof has a unique shape. In order to ensure its safety performance, its wind load is determined by wind tunnel tests. The following figure 2 shows the wind tunnel test model and ground roughness simulation picture (north-facing flow). The wind tunnel test uses a 1:300 model with a ground roughness of Class B. The wind tunnel test takes north as the 0° direction and rotates clockwise to complete a total of 36 wind load tests in wind directions.

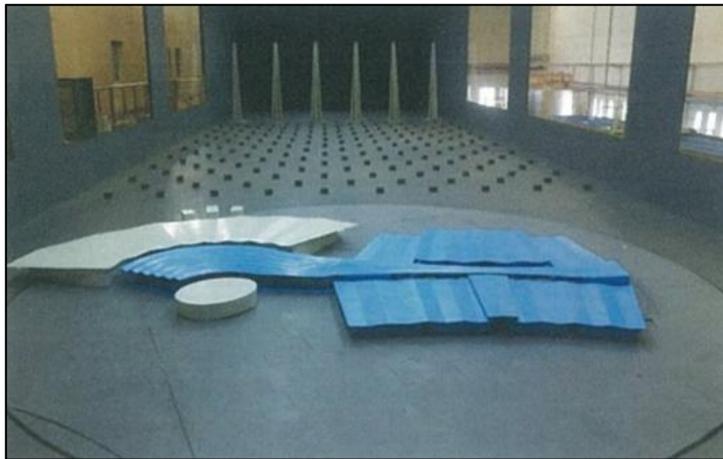


Figure 2 Wind tunnel test of Hangzhou Convention and Exhibition Center

From the wind tunnel test results, the middle corridor is disturbed by the surrounding venues, and the wind load is relatively small in most wind directions. The mean value of the wind pressure coefficient is defined as formula (9), and the root mean square value of the wind pressure coefficient is defined as formula (10).

$$W_p = \frac{p}{0.5\rho v^2} \quad (9)$$

$$\widetilde{W}_p = \frac{\tilde{p}}{0.5\rho v^2} \quad (10)$$

Where: W_p is the mean value of the wind pressure coefficient; \widetilde{W}_p is the root mean square value of the wind pressure coefficient; p is the mean wind pressure at the measuring point; \tilde{p} is the root mean square value of the wind pressure at the measuring point. v is the incoming wind speed at a height of 10m corresponding to the prototype; ρ is the air density.

When the corridor is at 90° , the interference in front is small. The following figure 3 and 4 shows the mean value of the wind pressure coefficient and the root mean square value of the wind pressure coefficient in the middle corridor under this wind direction. Since the roof of the east entrance is significantly higher than the interfering buildings on both sides, and the elevation of the edge of the entrance is higher than the elevation of the middle corridor, the wind load at the end is large, while the wind load in the middle is small due to the existence of reattachment. The maximum negative wind pressure coefficient of the entrance can reach -1.7; the root mean square value of the wind pressure coefficient can reach 0.3.

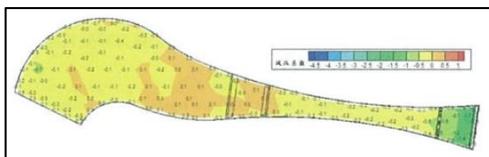


Figure 3 Average Value of Wind Pressure

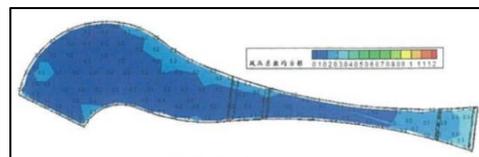


Figure 4 Root Mean Square Value of Wind Pressure Coefficient

Coefficient

3.2 Structural Response and Response-Related Scales

The total length of the corridor can reach 960m. In order to ensure the coordination of the structure under the action of temperature, the structure is divided into multiple substructures for design during structural design. Among them, the wind load at the east entrance is large, so this part of the structure is selected for key analysis. The roof of the east entrance is trapezoidal, with a structural length of 86m and a maximum span of 95m. The typical vertical vibration mode of the structure and the first 500 natural vibration periods of the structural period are obtained through finite element analysis.

The 90° wind direction flow load is applied to the structure, and the response of the structure under the action of average wind and pulsating wind is shown in the figure 5 below. Obviously, the position where the maximum displacement of the structure occurs is the same as the distribution form of wind load. The maximum average displacement and the maximum pulsating displacement of the structure are both located at the center of the east end of the login port, and the maximum average displacement and the maximum pulsating displacement are 55mm and 10mm respectively.

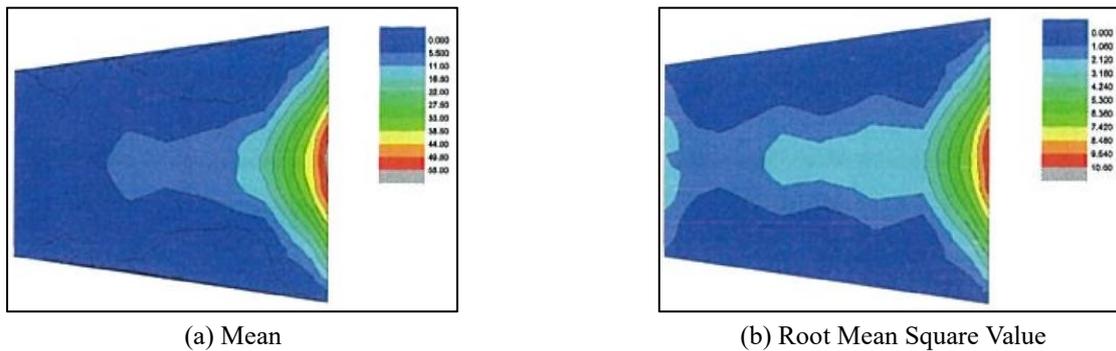


Figure 5 Wind-Induced Displacement Response

The relevant scale of the structural displacement response is calculated using formula (8), and the relevant scale distribution diagram at any node of the model can be obtained, as shown in the figure 6 below. The maximum position of the response-related scale is located near the center of the roof plane, and three local extreme values appear, which deviate greatly from the maximum position of the structural displacement, indicating that the maximum displacement of the structure is a local response caused by local loads. If the maximum displacement of the structure is used as a single equivalent target, it cannot fully reflect the structural response to other areas with smaller displacements.

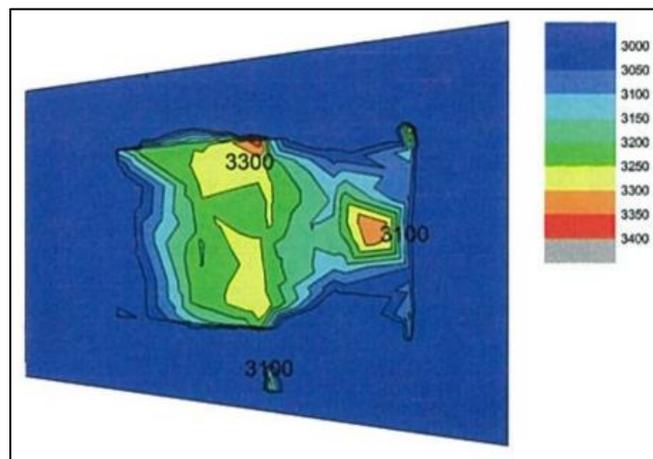


Figure 6 Displacement Response Related Scale

Equivalent static wind load: In order to compare the difference of equivalent static wind loads with different equivalent targets, the dynamic load response related method is used with the maximum displacement as the equivalent target, the dynamic load response related method is used with the structural displacement at the extreme position of the response integral scale as the multi-target equivalent, and the least square method is used with the response of all roof nodes as the equivalent target to calculate the equivalent static wind load. The distribution of equivalent static wind load is shown in the figure 7-9 below.

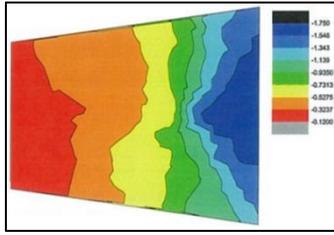


Figure 7 Equivalent Static Wind Load with a Single Displacement Extreme Value as the Equivalent Target

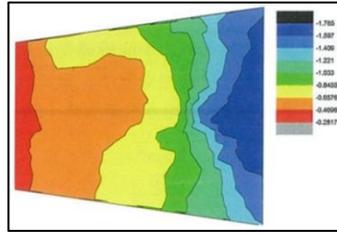


Figure 8 Equivalent Static Wind Load with all Displacement Extreme Values as the Equivalent Target

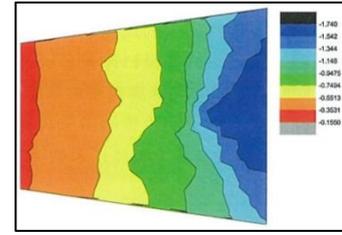


Figure 9 Equivalent Static Wind Load with the Displacement Corresponding to the Position of the Response-Related Scale Extreme Value as the Equivalent Target

From the overall distribution point of view, the equivalent static wind loads corresponding to different methods are close to the distribution form of the average wind load of the structure. The wind load on the west side is small and the wind load on the east side is large, but there are obvious differences in the distribution range of local wind loads. When the maximum displacement is used as a single target equivalent, the correlation between the wind load on the west side and the maximum displacement on the east side is poor, so the range of the small wind load on the west side is significantly higher than the multi-target equivalent static wind load.

Each equivalent static wind load is applied to the structural model, and the structural displacement response at the 1-12 node positions is calculated respectively, as shown in the figure 10 below. It can be seen that different equivalent static wind loads can accurately reflect the extreme displacement response of point 11, but the displacements of positions 1-9 obtained based on the single-target equivalent static wind load are quite different from the maximum displacement, and the maximum relative error at point 2 can reach more than 50%. The accuracy of the response-related scale method with fewer equivalent targets is close to that of the least squares method with all responses as equivalent targets, and the error at point 2 is the largest, about 7%.

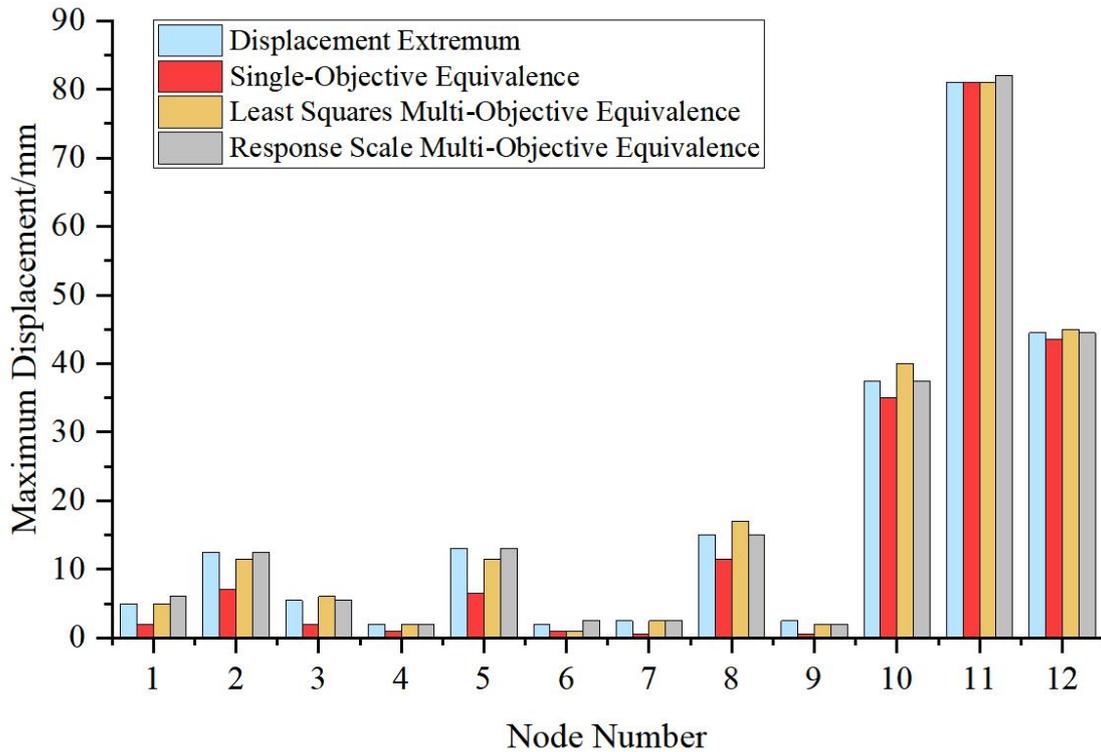


Figure 10 Comparison of Node Displacements Corresponding to Different Equivalent Static Wind Loads

The following table 1 shows the vertical resultant forces of the structure corresponding to different equivalent static wind loads. The calculation result based on the single-target equivalent has the largest error, while the relative error based on the response-related scale method is the smallest.

Table 1 Vertical Resultant Force Corresponding to Different Equivalent Static Wind Loads

Equivalent target	Equivalent method	Vertical load resultant/ 10^6N	Maximum vertical load/ 10^6N	Relative error
Maximum displacement	Dynamic load response	3.75	5.26	-28.7%

	correlation method			
Total displacement of the roof	Least square method	5.82	5.26	10.6%
Displacement of the maximum position of the response-related scale	Dynamic load response correlation method	5.50	5.26	4.6%

Based on the above research, a large-span roof wind load calculation method based on response correlation scale is proposed. The response correlation scale reflects the range of influence of the overall structure on the response of a certain node position of the structure. The larger the index, the stronger the overall structure response of the position. Since the large-span roof structure is an over-static system, multiple response correlation scale extreme values are often required to determine the equivalent target that fully reflects the overall response of the structure. Based on the selected equivalent target, the equivalent static wind load of the structure is determined by the dynamic load response correlation method. The experimental analysis based on the roof of the east entrance of Hangzhou Convention and Exhibition Center shows that both the multi-objective equivalent static wind load and the single-objective equivalent static wind load can reflect the wind load of the main wind-exposed area on the surface of the structure, but the distribution shapes of the equivalent static wind loads of different methods are different in the area with small average wind load, and the multi-objective equivalent results are greater than the single-objective equivalent results; the equivalent static wind load method based on the response-related scale is used, and the accuracy of the calculated node displacement response is close to that of the least squares equivalent static wind load calculation method, but the vertical force accuracy is higher than the least squares method, and the maximum error is only 4.6%.

4 CONCLUSION

This paper proposes an equivalent static wind load calculation method based on the response-related scale for the calculation of the wind load of the east entrance of Hangzhou Convention and Exhibition Center. Through wind tunnel tests and finite element analysis, this method has significant advantages in reflecting the overall response characteristics of large-span roof structures. The research results show that the equivalent static wind load method based on the response-related scale can not only accurately reflect the wind load distribution in the main wind-exposed area, but also effectively control the calculation error of the displacement response of each node, with a maximum error of only 4.6%. Therefore, this method has broad application prospects and promotion value in the calculation of wind loads on large-span roofs.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Cai Meng, Greens McCann, Tang Jiani. Concept and practice review of international green exhibitions. *Journal of Shanghai University of International Business and Economics*, 2015, 22(04): 37-47.
- [2] Chen Nan, Wang Limin, Zhao Pengfei, et al. Steel structure roof design of the entrance hall of Hongdao International Convention and Exhibition Center. *Building Science*, 2024, 40(03): 1-9.
- [3] He Yuanming, Huang Yongjun, Mao Tongxiang. Node design of the roof of the entrance hall of Shenzhen International Convention and Exhibition Center. *Building Structure*, 2020, 50(02): 30-35.
- [4] Chen Fubin, Li Qiusheng, Fu Jiyang, et al. Experimental study on frequency domain characteristics of wind loads on large-span roofs. *Vibration and Shock*, 2012, 31(05): 111-117.

