

Research on Adaptive Coding and Denoising System Based on Sound Quality and Audio Effect Balance Optimization

Yechi Zhang*

Zhuoyue Hornors College, Hangzhou Dianzi University, Hangzhou 310018, China

* Corresponding author: Yechi Zhang (Email: 24270606@hdu.edu.cn)

Abstract: This article proposes a comprehensive evaluation index, cost-effectiveness analysis model, adaptive encoding scheme, and improved denoising strategy for optimizing audio storage and noise removal. To improve audio quality and storage efficiency, and to quantify the balance between storage efficiency and sound quality fidelity for different audio formats, this article establishes a sound storage balance quantization model. By constructing a comprehensive evaluation quantization function and combining multidimensional factors to evaluate the balance between storage efficiency and sound quality fidelity. Based on a three-level indicator system (basic technology, performance evaluation, and comprehensive evaluation), entropy weight method is used to calculate indicator weights, and TOPSIS method is applied for comprehensive scoring. The results indicate that AAC format has the best sound storage balance performance, while WAV and FLAC have superior sound quality but lower storage efficiency, MP3 focuses more on storage efficiency in low bandwidth environments. At the same time, this article establishes a multidimensional audio evaluation model that quantifies sound quality using factors such as mean square error (MSE), while quantifying the size of audio files through factors such as audio duration, sampling rate, bit depth, number of channels, and compression ratio. Based on relevant features, the sound quality performance of speech and music content was evaluated, and the comprehensive and relative effects of sampling rate, bit depth, and compression algorithm on sound quality were analyzed. Based on this, an audio performance balance index was designed to quantify the cost-effectiveness between sound quality and file size. The sorting results show that music-44100Hz-MP3_320kpp.mp3 and speech-44100Hz-MP3_320kpp.mp3 have the highest cost-effectiveness, and recommend the best parameter configurations for voice and music audio.

Keywords: Entropy Weight Method TOPSIS, Multidimensional Audio Evaluation Model, MDCN, LSTM Architecture.

1. Introduction

In the field of digital media, audio processing technology has developed into a key technological foundation. Audio signal processing mainly faces three technical difficulties: optimizing storage efficiency, eliminating noise, and optimizing timbre, which directly affect the end-user experience. To cope with the large capacity characteristics of audio data, the key is to maintain sound quality integrity while achieving efficient storage and transmission. The file size and sound quality of various audio formats are affected by different parameters such as sampling rate, bit depth, and compression algorithm, and different types of audio often require different parameter configurations[1]. In practical applications, audio quality is often affected by indoor and outdoor environmental noise and equipment noise. Due to the different time-frequency characteristics of these noises, professional noise reduction algorithms are required for processing. This article uses mathematical modeling to solve audio storage optimization and noise removal problems, in order to improve overall audio quality[2]. Design a comprehensive evaluation metric to quantitatively analyze the balance between storage efficiency and sound fidelity of different audio formats such as WAV, MP3, AAC, etc. The evaluation criteria need to comprehensively consider file size, sound quality loss, encoding and decoding complexity, and applicable scenarios to provide a scientific basis for audio format selection. Based on the provided audio files, establish a model to deeply analyze the impact mechanism of core parameters such as sampling rate, bit depth, and compression

algorithm on audio quality and file size. We need to design a scientific cost-effectiveness index to evaluate the balance between sound quality and file size, and based on this, systematically rank different parameter combinations, ultimately providing optimal parameter recommendations for both voice and music content.

2. Establishment of a Quantitative Model for Sound Storage Balance

2.1. Construction of comprehensive evaluation function

In practical applications of audio encoding, storage efficiency and sound quality fidelity often exhibit a mutually restrictive relationship. In the quantitative model of sound storage balance, this article constructs a comprehensive evaluation function to scientifically evaluate this balance relationship. This function not only considers the independent performance of two key indicators, but also introduces balance evaluation to comprehensively quantify the performance of audio formats.

This article uses a weighted product to highlight the importance of each indicator, while introducing a balance term to measure the coordination between two key indicators. This design approach ensures the comprehensiveness of the evaluation system while also considering practicality in practical applications. The specific function expression is as follows:

$$z(x, y) = x^a y^b \left(1 - \frac{|ax-by|}{ax+by}\right) \quad (1)$$

The schematic diagram of this function is shown in Figure 1.

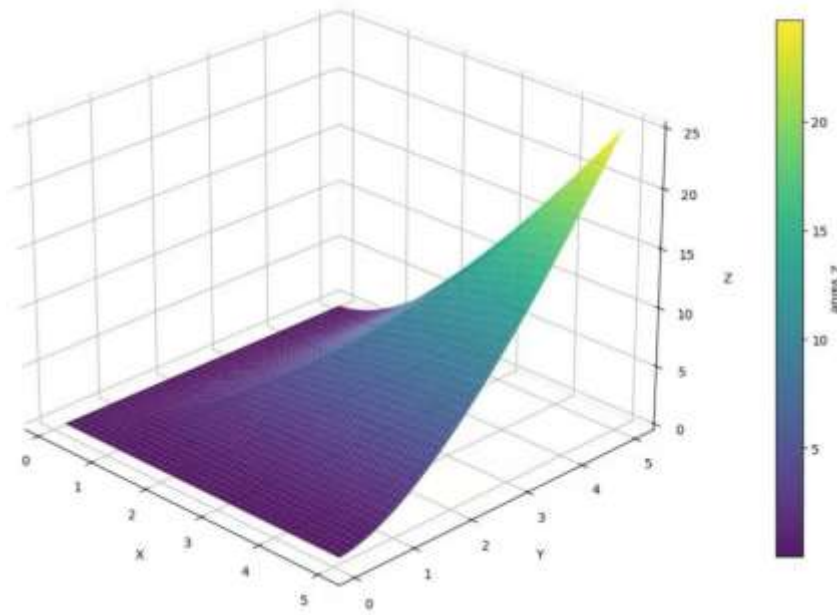


Figure 1 Schematic diagram of comprehensive evaluation function

The specific meanings of each parameter in the function are as follows:

(1) X and y represent the fidelity of sound quality and storage efficiency, respectively, and are the core components of the balance relationship.

(2) A and b: Weight coefficients representing fidelity of sound quality and storage efficiency, respectively. In practical applications, they should be flexibly adjusted according to specific applicable scenarios to better meet the requirements of balanced relationship quantification results.

This article provides an in-depth analysis of the various parts of the function: the function can be divided into product terms x_a , y_b and balance terms $1 - \frac{|ax-by|}{ax-by}$. Next, this article will analyze these two parts separately.

This article adopts an exponential form design, which can better reflect the nonlinear impact of indicator changes on overall evaluation. By adjusting indices a and b, this article can precisely control the importance of each indicator. In addition, the exponential form also ensures that the evaluation results increase monotonically as the indicators improve.

This study draws inspiration from the Gini coefficient in economics and is used to measure the equilibrium between two indicators[3]. When two indicators reach equilibrium, a value of 1 indicates optimal balance; When there is a significant difference between indicators, this value approaches 0, which means that the effect of punishing imbalance is enhanced, thereby ensuring the rationality of the evaluation.

2.2. Three level indicator system

In order to achieve a systematic evaluation of audio format performance, a complete three-level indicator system was constructed for the audio storage balance quantification model. On the basis of in-depth analysis of storage efficiency

and sound fidelity, this article comprehensively considers multiple key factors, including file size, sound quality loss, encoding and decoding complexity, and compression algorithm efficiency, and finally constructs this three-level indicator system.

By establishing a three-level indicator system, this article can more accurately evaluate the performance of different audio formats in different application scenarios. Among them, the first level is the basic technical indicator layer, which covers parameters that directly affect sound quality and storage efficiency; The second level is the performance evaluation layer, which further analyzes the actual impact of these technical parameters on user experience; The third level is the comprehensive evaluation layer, which makes the final judgment on the balance relationship as a whole.

2.3. Solution of Quantitative Model for Sound Storage Balance

To comprehensively evaluate the balance between storage efficiency and sound quality fidelity of different audio formats, this paper selects four representative audio formats [2]: WAV, MP3, AAC, and FLAC formats, and conducts in-depth quantitative analysis on them.

2.3.1. Determine the evaluation index system and raw data matrix

Firstly, based on the three-level indicator system established by the sound storage balance quantification model, this article collected data on various indicators for different audio formats. These data include sound quality loss, file size, encoding and decoding complexity, compression algorithm efficiency, etc. Due to hardware limitations on bit depth and sampling rate, the same values are uniformly used in all formats. The indicator values for each audio format are shown in Table 1.

Table 1 Audio Format Indicator Fixed Values Table (WAV, MP3, AAC, FLAC)

Indicator category	Indicator Name	unit	WAV	MP3	AAC	FLAC
Sound quality fidelity	Signal to Noise Ratio(SNR)	dB	96	60	70	96
	Distortion rate	percentage(0~ 1)	0.0050	0.1500	0.1000	0.0050
	bit depth	bits/sample	16	16	16	16
Storage Efficiency	sampling rate	Hz	44100	44100	44100	44100
	file size	MB/minute(Dual channel)	10.1000	1.0000	0.9000	5.5000
	Metadata compression ratio	percentage(0~ 1)	0.0100	0.3000	0.2500	0.0200
	Quantitative accuracy adjustment	fraction(1~5)	1	3	4	2
	Whole frame structure design	fraction(1~5)	1	3	4	2

Then, organize these data into a raw decision matrix for subsequent standardization processing.

2.3.2. Standardized raw data

To ensure comparability of data for different indicators, it is first necessary to standardize the original data matrix. The commonly used standardization methods include range standardization and Z-score standardization. Standardized data can eliminate dimensional differences between different indicators, ensuring consistency in the contribution of each indicator to the final evaluation results.

For each indicator i , its standardized value X_{ij} can be calculated using the following formula:

$$X'_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (2)$$

Among them, X_{ij} represents the value of the i -th indicator in the j th sample, and $\min(X_i)$ and $\max(X_i)$ are the minimum and maximum values of the indicator, respectively.

Specifically, certain indicators (such as signal-to-noise ratio, bit depth, etc.) belong to profitability indicators, meaning that the larger the value, the better, indicating higher sound quality and storage efficiency; And other indicators (such as distortion rate, file size, etc.) belong to the cost type indicators, that is, the smaller the value, the better, indicating lower storage occupancy and information loss[4]. By standardizing calculations, it ensures that each indicator data can be analyzed at the same scale, thereby making it consistent and comparable in comprehensive evaluation.

2.3.3. Calculate the score of the third level indicator

Due to the entropy weight method determining the importance of each indicator in the overall evaluation based on its information entropy[5][6], this paper uses the entropy weight method to calculate the weights of each indicator in the three-level indicator system, and calculates the score by weighted summation.

2.3.4. Calculate the proportion of standardized values for each indicator

Firstly, this article calculates the proportion of standardized values for each indicator. The formula for calculating the normalized weight P_{ij} of the i -th indicator and j -th sample

data X'_{ij} is:

$$P_{ij} = \frac{x'_{ij}}{\sum_{j=1}^m x'_{ij}} \quad (3)$$

Among them, X'_{ij} is the standardized value of the i -th indicator in the j -th sample, and m is the number of samples.

2.3.5. Calculate the entropy value of each indicator

Next, this article calculates the entropy value H_i of each indicator, which reflects the information uncertainty of that indicator. The formula for calculating information entropy is as follows:

$$H_i = -k \sum_{j=1}^m P_{ij} \ln P_{ij} \quad (4)$$

$$k = \frac{1}{\ln m}$$

Among them, $\frac{1}{\ln m}$ is a constant, and P_{ij} is the standardized proportion of the i -th indicator in the j -th sample.

2.3.6. Calculate the weight of each indicator

According to the size of information entropy, this article can calculate the weight w_i of each indicator. The smaller the entropy value, the greater the weight of the indicator. The formula for calculating weights is:

$$w_i = \frac{1 - H_i}{\sum_{i=1}^n (1 - H_i)} \quad (5)$$

Among them, H_i is the entropy value of the i -th indicator, n is the total number of indicators, and w_i is the weight of the i -th indicator.

This article notes that when the values of certain indicators are completely consistent, the variability of these indicators is zero, resulting in their entropy reaching its maximum value (i.e., the larger the entropy value, the smaller the weight). According to the principle of entropy weight method, the greater the degree of variation of an indicator, the greater the amount of information it carries, so its weight should be higher. Due to the fact that bit depth and sampling rate are completely consistent in different audio formats, their variability is zero and they do not have sufficient discrimination. Therefore, this article decides to exclude these two indicators from the calculation of entropy weight method to avoid adverse effects on the final results.

2.4. TOPSIS method evaluation

After obtaining the comprehensive scores of each option, this article uses the TOPSIS method to further evaluate these scores. The TOPSIS method is a multi criteria decision analysis tool aimed at evaluating the relative advantages of each option by calculating the distance between each option and the ideal and negative ideal solutions.

2.4.1. Build decision matrix

Construct a decision matrix based on the comprehensive scores of each option under each indicator. Each column represents an indicator, and each row represents an option.

2.4.2. Standardized decision matrix

Standardize the decision matrix to eliminate dimensional differences between different indicators. The standardized formula is as follows:

$$X'_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}} \quad (6)$$

Among them, X_{ij} is the score of the i -th option under the j th indicator, and X'_{ij} is the standardized value.

2.4.3. Determine the ideal solution and negative ideal solution

The ideal solution and negative ideal solution are the

maximum and minimum values of each indicator, respectively (distinguishing between benefit based and cost based indicators).

(i) The ideal solution for benefit oriented indicators is the maximum value of each indicator, while the negative ideal solution is the minimum value of each indicator;

(ii) The ideal solution for cost indicators is the minimum value for each indicator, and the negative ideal solution is the maximum value for each indicator.

2.4.4. Calculate Euclidean distance

Calculate the Euclidean distance between each option and the ideal and negative ideal solutions:

$$\sqrt{\sum_{i=1}^n (X'_{ij} - A_j^+)^2} \quad (7)$$

$$\sqrt{\sum_{i=1}^n (X'_{ij} - A_j^-)^2} \quad (8)$$

Among them, A_j^+ and A_j^- are the values of the ideal solution and the negative ideal solution under the j th index,

respectively[7].

2.4.5. Calculate relative proximity

Finally, calculate the relative proximity C_i of each option to determine its superiority or inferiority:

$$C_i = \frac{D_i^+}{D_i^+ + D_i^-} \quad (9)$$

The larger the value, the closer the option is to the ideal solution and the higher the priority.

When performing range normalization, some indicators may have zero values, which may affect subsequent calculations. To avoid this problem, this article adopts the translation method, which ensures that all values are positive by translating the data as a whole, i.e. $x_{ij} + \epsilon$ [8][9]. After sensitivity analysis of the translation constant, this article sets ϵ to 10⁻⁶, ensuring that it is sufficiently small without affecting the reasonable allocation of weights and maintaining dimensional consistency. The sensitivity analysis results are shown in Figure 2:

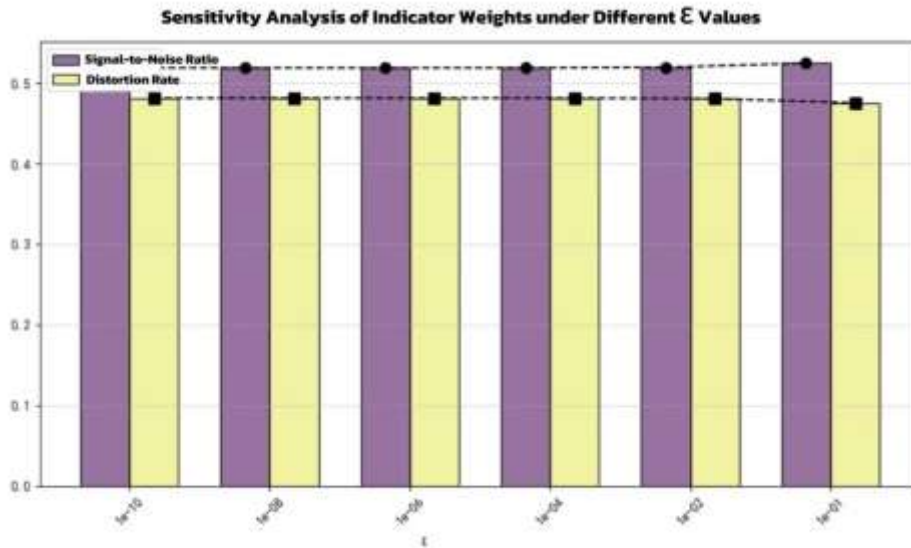


Figure 2 Sensitivity analysis of ϵ value

2.5. Comprehensive score calculation

By using the entropy weight TOPSIS method, this article can obtain the comprehensive score of audio fidelity and storage efficiency for each format through weighted summation[10]. The formula is as follows:

$$S_j = \sum_{i=1}^n X'_{ij} \times w_i \quad (10)$$

Among them, S_j represents the comprehensive score of the j th sample, X'_{ij} is the standardized value of the i -th indicator in the j th sample, and w_i is the weight of the i -th indicator. The results are shown in Table 2:

Table 2 Comprehensive Score Table for Audio Format

Audio	Audio fidelity score	Storage efficiency score
WAV	0.9500	0.2610
MP3	0.1000	0.5210
AAC	0.3200	0.7320
FLAC	0.9500	0.4870

2.6. Scene selection and function weight determination

Based on the comprehensive evaluation function in the constructed sound storage balance quantification model, considering the applicability of different application scenarios, this paper progressively assigns weights according to the requirements of sound quality and storage efficiency for each scenario. The progressive relationship of weights for each scene is shown in Figure 3:

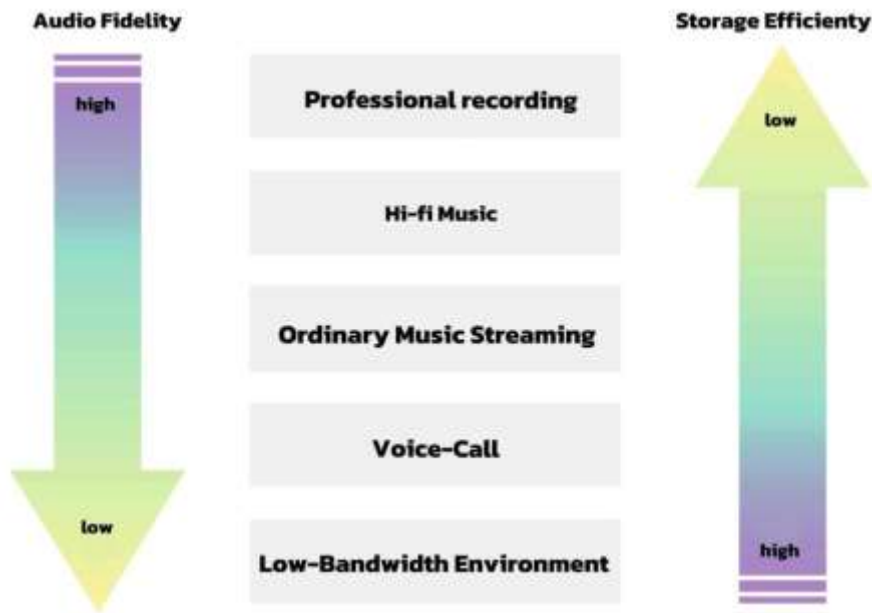


Figure 3 Progressive Relationship Diagram for Scene Selection

2.7. Quantitative results of balancing relationships

Based on the comprehensive score and the calculation of function weights for each evaluation dimension, this article can quantify the performance of different audio formats in

various scenarios and give them a comprehensive rating. These ratings reflect the overall performance of each format in key factors such as sound quality, file size, and compression efficiency. The specific scoring results are shown in Table 3:

Table 3 Comprehensive Evaluation Score of Audio Format (Different Scenarios)

Scenario Type	FLAC	WAV	AAC	MP3
Professional Recording and Music Production	0.1850	0.1710	0.0670	0.0440
High fidelity music scene	0.5420	0.3000	0.1800	0.0480
Ordinary music streaming scene	0.4610	0.2150	0.2940	0.0740
Voice call scenario	0.2800	0.1360	0.4060	0.1020
Low bandwidth environment	0.0960	0.0830	0.1410	0.0860

The radar images of each audio format in different scenarios are shown in Figure 4. In the radar chart, each angle

represents an application scenario, and each line in the chart represents the applicability of the audio format.

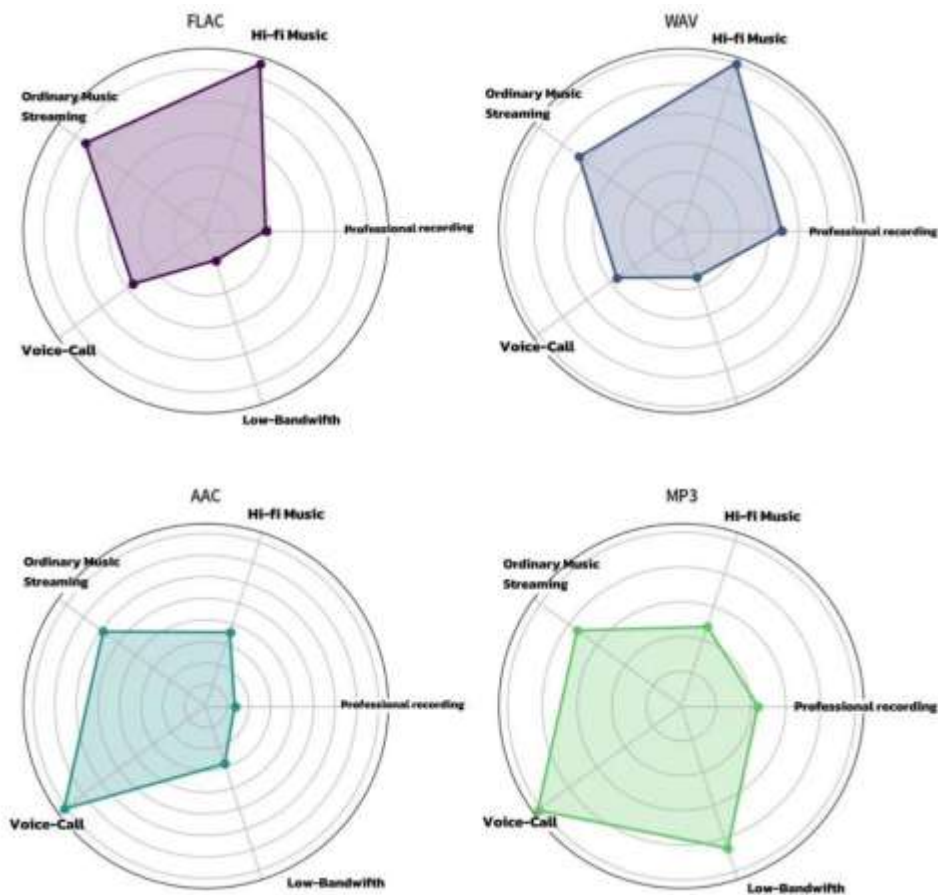


Figure 4 Rating Radar Chart of Various Audio Formats in Different Scenarios

3. Conclusions

Through the comprehensive evaluation indicators designed above, this article comprehensively evaluates the performance of different audio formats in various scenarios, and obtains corresponding comprehensive scores for each audio format in different scenarios. These ratings can reflect the balance between storage efficiency and sound fidelity for each format, providing a scientific basis for audio format selection in practical applications.

The results indicate that AAC format performs the best in terms of sound storage balance. It can effectively compress file size while maintaining good sound quality, balancing storage efficiency and sound quality. Although WAV and FLAC formats provide better sound quality, they may not be suitable for scenarios with limited storage space due to their lower storage efficiency and larger file size. In contrast, MP3 format focuses more on storage efficiency, especially in low bandwidth environments, which can reduce file size through higher compression ratios, thereby improving data transmission efficiency.

References

[1] Chen Qingyun, Gai Junmei Research on the Selection and Conversion Path of Audio File Archiving Format [J]. Lantai World, (06):50-53, 2022.

[2] Hang, T., Gu, S., Li, C., Bao, J., Chen, D., Hu, H., Geng, X., Guo, B. (2023). Efficient Diffusion Training via Min-SNR Weighting Strategy. CVF, 7441: 1-8, 2023.

[3] Geng Lele A Study on the Equity of Per Capita Expenditure on Basic Education in China: Calculation Based on Gini Coefficient and Theil Index [J]. Journal of East China Normal University (Education Science Edition), 40 (01): 60-73, 2022.

[4] Wang Lu Research on Anomaly Detection Algorithm Based on Multi Model Integration [D]. Shandong University, 2023.

[5] Shi Guigang, Liu Wanduan, Zuo Jingyao Research and Comparative Analysis of Index Weighting Techniques Based on Entropy Weight Method [J]. Journal of Jiamusi University (Natural Science Edition), 2024, 42 (11): 168-171, 2024.

[6] Yang Liang, Hou Wei Error analysis of multi-source 3D geological modeling data based on entropy weight method [A] Proceedings of the 12th National Symposium on Mathematical Geology and Geological Information [C]: 391, 2013.

[7] Sun Dajun, Ma Chao, Mei Jidan, etc Vector array deconvolution beamforming method based on non negative least squares [J]. Journal of Harbin Engineering University, 40 (07): 1217-1223, 2019.

[8] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017). Focal Loss for Dense Object Detection. IEEE, pp. 1-9, 2017.

[9] Li, F., Guo, S. (2024). Multi-Scale Dilated Convolution Network for Long-Term Time Series Forecasting. ICLR, pp. 1-9, 2024.

[10] Kang Ning, Mo Luyao, Jing Ke (2024). Semi parametric smooth transformation quantile autoregressive model and its application Systems Science and Mathematics: 844, 2024.