UDC 004.81

ESTIMATION OF DOWNSAMPLING INFLUENCE ON ACCELEROMETER SIGNAL QUALITY

HAO LI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 08, 2024

Abstract. Downsampling plays a crucial role in the field of accelerometer data analysis, especially in applications that require optimizing computing resources while ensuring minimal loss of signal integrity. This study examines the influence of downsampling on accelerometer signal quality, critical for optimizing activity recognition and health monitoring applications. Utilizing the Sisfall dataset, which captures diverse physical activities through triaxial accelerometers, the research evaluates the influence of downsampling on signal quality using metrics such as Mean Squared Error (*MSE*), energy, and Cumulative Explained Variance (*CEV*). The findings reveal that a specific downsampling factor, effectively balances data volume reduction with signal integrity preservation. This optimal factor enhances computational efficiency without significantly compromising signal quality, pivotal for the accurate analysis and recognition of physical activities. The study underscores the importance of selecting an appropriate downsampling rate to maintain the fidelity of accelerometer data in wearable technologies and health monitoring systems, providing a guideline for signal processing optimizations in such applications.

Keywords: downsampling, accelerometer data, signal quality, metrics, physical activity.

Introduction

The advent of wearable technology and the proliferation of devices equipped with accelerometers have ushered in a new era of data driven insights into human motion and activity [1]. Accelerometers, sensors capable of detecting and recording movements in dimensional space, are now ubiquitous in smartphones, fitness trackers, and medical monitoring devices [2]. They generate vast amounts of data, enabling the detailed analysis of physical activities, from daily routines to specialized athletic performance. However, the sheer volume of data produced poses significant challenges in terms of storage, processing, and analysis, necessitating efficient strategies to manage this deluge of information without compromising the quality and integrity of the signal [3].

Downsampling [4], the process of reducing the sampling rate of a signal by removing some of its components, emerges as a critical technique in addressing these challenges. It holds the promise of significantly reducing the data volume, thereby easing the demands on computational and storage resources. Nevertheless, downsampling is not without its pitfalls; it risks degrading the signal quality by omitting potentially crucial information, a consequence that must be meticulously managed, especially in applications where data integrity is paramount [5].

The Sisfall dataset, a comprehensive collection of accelerometer data focused on physical activities and falls, provides a rich foundation for exploring these challenges. It encompasses data meticulously recorded to capture the nuances of human motion, offering a unique opportunity to study the impact of downsampling on signal integrity within the context of activity recognition. Given the critical role of accelerometer data in applications ranging from healthcare monitoring to emergency response systems, understanding the effects of downsampling is paramount. Notably, in reference [9], researchers have utilized this dataset for fall detection, achieving commendable results, which underscores the dataset's value and applicability in practical scenarios.

This paper delves into the optimization of accelerometer data processing, with a specific focus on the influence of downsampling on signal quality. Through a systematic analysis using various metrics such as Mean Squared Error (*MSE*) [6], energy, Cumulative Explained Variance (*CEV*) [7], and a novel ratio measure, the study aims to identify an optimal downsampling strategy. This strategy seeks to balance the compromise between reducing data volume and preserving the fidelity of the original signal, thereby enhancing the computational efficiency of data processing without sacrificing the accuracy of activity recognition.

The implications of this research extend beyond academic interest, offering practical insights for the design and optimization of wearable technologies and health monitoring systems. By establishing a guideline for selecting appropriate downsampling rates, this study contributes to the advancement of signal processing workflows, ensuring that wearable devices and health monitoring applications can provide reliable and accurate data with minimized computational demands.

Dataset analysis

The module raw accelerometer data serves as a pivotal foundation for our research, encapsulating a comprehensive accelerometer sensor dataset tailored for the analysis of physical activities, the module is shown as Figure 1. This dataset is distinguished by its focus on five classes of physical activities, each meticulously recorded to capture the nuances of human movement and potential falls. The essence of this dataset is summarized in two essential tables, in the Table 1, which outlines the key characteristics of the Sisfall dataset [8], and in the Table 2, which categorizes the activities and falls under investigation.

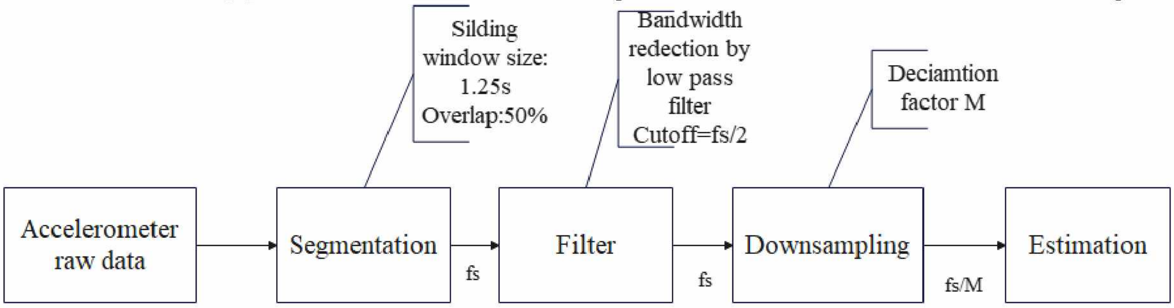

Figure 1. Block diagram of accelerometer data downsampling algorithm

Table 1. **The key characteristics of the Sisfall dataset**

Parameter	Sisfall dataset
Sampling frequency	200*Hz*
Number of activities of daily living	19
Number of fall activities	15
Sensors used	Accelerometer
Position of sensor	Waist
Window size	1.25 seconds
Window overlap rate	50%

Table 2. **Five classes of activities and falls selected for this work**

Code	Activity	Trials	Duration
*F*01	Fall forward while walking caused by a slip	5	15 seconds
*F*05	Fall forward while jogging caused by a trip	5	15 seconds
*F*13	Fall forward while sitting, caused by fainting or falling asleep	5	15 seconds
*D*08	Quickly sit in a half height chair, wait a moment, and up quickly	5	12 seconds
*D*13	Sitting a moment, lying quickly, wait a moment, and sit again	5	12 seconds

Boxplots are a standardized way of displaying the distribution of data based on a five metrics: minimum, first quartile, median, third quartile, and maximum. In the context of accelerometer data for physical activities, boxplots can reveal a lot about the nature of movements associated with each activity class. The boxplot of fall activity is shown as Figure 2.

From the boxplot, we can observe that fall activities display outliers, reflecting the sudden nature of falls. The range of acceleration values (indicated by the whiskers) might be wider due to the abrupt start and end points of a fall. Dynamic activities show a larger interquartile range (*IQR*), indicating variability in movement speed and style among different trials or subjects. Therefore, using boxplots allows us to more clearly observe the distribution and characteristics of the data, which helps us better analyze the data of different activities.

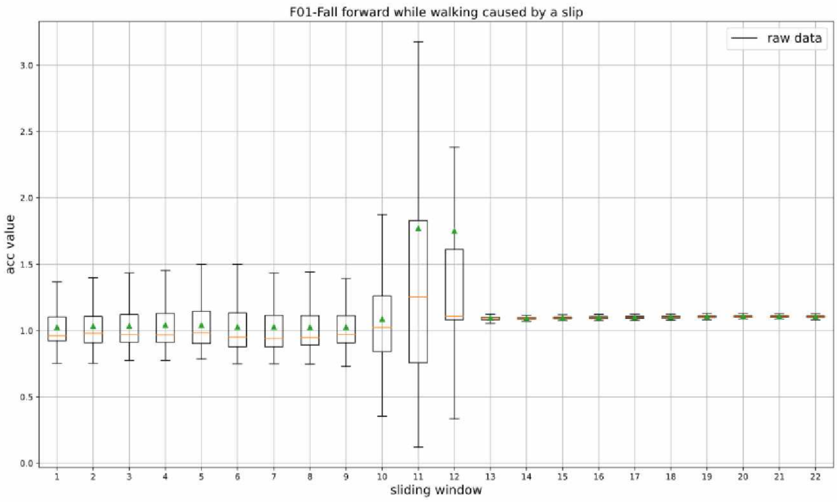

Figure 2. The boxplot of fall activity

By comparing these aspects across the classes of activities, we can glean insights into the characteristics of each activity type, such as the consistency of movement, the presence of abrupt starts or stops, and the overall intensity.

Assessing the influence of downsampling on signal quality

In this study, we aimed to evaluate the impact of downsampling on the quality of accelerometer signals across various physical activities, including walking, jogging, sitting, and dynamic postural transitions. Given the critical role of accurate accelerometer data in applications ranging from healthcare monitoring to activity recognition, understanding the effects of downsampling is paramount for optimizing signal processing workflows. The formula of downsampling by using the expression

$$y[n] = x[nM], \quad (1)$$

where $y[n]$ is the accelerometer signal after downsampling. $x[n]$ is the accelerometer original signal, with n is the sample number in the original signal. M is the downsampling factor, determining the interval at which samples are selected from the original signal. For example, if $M = 2$, every other sample is taken from the original signal for downsampling, if $M = 3$, every third sample is taken.

By selecting every Mth data point from the original signal, the downsampling process lowers the signal's sampling rate. This means if the original signal has a sampling rate of fs, the sampling rate of the downsampled signal will be fs / M. Downsampling is a common operation in signal processing, especially when there's a need to reduce the volume of data to simplify analysis or decrease processing load. However, downsampling can lead to the loss of some high frequency information in the signal, known as aliasing. Therefore, appropriate antialiasing filtering, according to the Nyquist theorem, is usually performed before downsampling to prevent this loss of information.

1. Metrics for evaluating the impact of downsampling on signal quality.

Evaluating the impact of downsampling on signal quality involves assessing how well the downsampled signal retains the essential characteristics of the original signal, considering the inherent compromise between reducing data size and maintaining fidelity. Several metrics can be employed to quantify this impact, each offering insight into different aspects of signal quality.

1.1. *MSE*. This metric measures the average error between the downsampled data and the original data, serving as a direct metric for assessing data loss during downsampling. By measuring the average error between the downsampled data and the original data, it provides a direct quantification of the extent of information loss during the downsampling process.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \quad (2)$$

where N is the number of samples, y_i is the signal after downsampling, $\hat{y}_i$ is the original signal.

1.2. Energy. this metric is used to calculate the energy ratio of the downsampled data. This helps to assess the impact of downsampling on signal energy and the degree of information loss in the signal during the downsampling process.

$$Energy = \frac{\sum_{i=0}^{k} x[i]^2}{\sum_{i=0}^{N-1} x[i]^2}, \quad (3)$$

where $x[i]$ is the signal after downsampling, N is the number of samples, k is the maximum frequency $f_{\max}/\Delta f$.

1.3. *CEV*. This metric is applicable for assessing the percentage of the total variance explained by the first K principal components in *PCA* analysis of downsampling data. It helps us understand how much of the original data's variability is retained during the downsampling process.

$$CEV = \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{NPC} \lambda_i}, \quad (4)$$

where λ_i is the eigenvalue of the ith principal component, NPC is the number of principal component, K is the number of principal components chosen for the calculation.

1.4. Ratio (*Ratiodown*). By computing and comparing metric, we can quantify the impact of downsampling on the signal structure. A downsampled ratio value close to 1 indicates that the downsampling has preserved the signal's recursive similarity structure, while a value far from 1 indicates a significant impact on the signal structure. This approach provides a basis for choosing the downsampling rate, specifically selecting those that minimize the impact on the signal's intrinsic similarity while reducing the amount of data.

$$Ratiodown = \frac{S_{j,d}}{S_j}, \quad (5)$$

$$S_{j,d} = \frac{1}{N_d}\sum_{j=0}^{N_d - i - i} | X_{d,j+i} - X_{d,j} |, \quad (6)$$

where X_d is the signal obtained by downsampling the original S, $X_{d,j}$ is the the value of the downsampled signal at time j, N_d is the length of the downsampled signal, i is the lag step size, S_i is the recursive similarity measure of the original signal at the same delay i.

1.5. The difference between downsampled signal and original signal (*Diff*). The metric serves as a quantitative measure of the effectiveness of the signal reconstruction process, specifically after a signal has undergone downsampling and subsequent upsampling with interpolation. It calculates the average absolute difference between the original signal $x[n]$ and its reconstructed version $z[n]$, effectively capturing the fidelity of reconstruction. A lower metric value indicates a closer match between the original and reconstructed signals, suggesting minimal loss of information and high preservation of signal quality through the downsampling and upsampling process. Therefore, this metric is very important in the evaluation of signal downsampling and upsampling data quality. It can more intuitively show the changes after data sampling.

$$Diff = \frac{\sum_{i=1}^{N} |z[i] - x[i]|}{N}, \quad (7)$$

$$z[n] = \sum_{k=0}^{L-1} y[k] \cdot \sin c\left(\frac{n-kM}{M}\right), \tag{8}$$

where $z[n]$ is the signal obtained after upsampling $y[n]$ back to the original sampling rate and applying interpolation to fill in the gaps, L is the length of the downsampled signal $y[n]$, $\sin c(n-kM/M)$ is the *sinc* function used for interpolation. It acts as an ideal low pass filter kernel to reconstruct the signal at intermediate points.

Experiment and analysis

This chapter is founded on a comprehensive evaluation utilizing a range of metrics, including *MSE*, Energy, *CEV*, and ratio. Each metric is tailored to scrutinize distinct aspects of signal integrity following downsampling. Through systematic experimentation with these metrics, the study aims to discern the most effective downsampling strategy, ultimately minimizing information loss.

MSE offers insights into the average discrepancy between the downsampled and original signals, highlighting the overall fidelity of the downsampling process. Energy metrics delve into the distribution of signal energy, shedding light on potential alterations in signal dynamics downsampling. *CEV* analysis contributes to understanding the variability of signal characteristics, aiding in identifying regions where downsampling may have a more pronounced impact.

The utilization of normalized values of *MSE*, energy, and *CEV* activities and bandwidths, as depicted in Figure 3, further enhances the depth of analysis. This visualization allows for a comparative examination of downsampling outcomes across different signal types and processing configurations. By scrutinizing the impact of downsampling under varying conditions, researchers can glean valuable insights into the robustness and versatility of different downsampling strategies.

Ultimately, the amalgamation of diverse evaluation metrics and systematic experimentation enables the identification of optimal downsampling approaches tailored to specific application requirements.

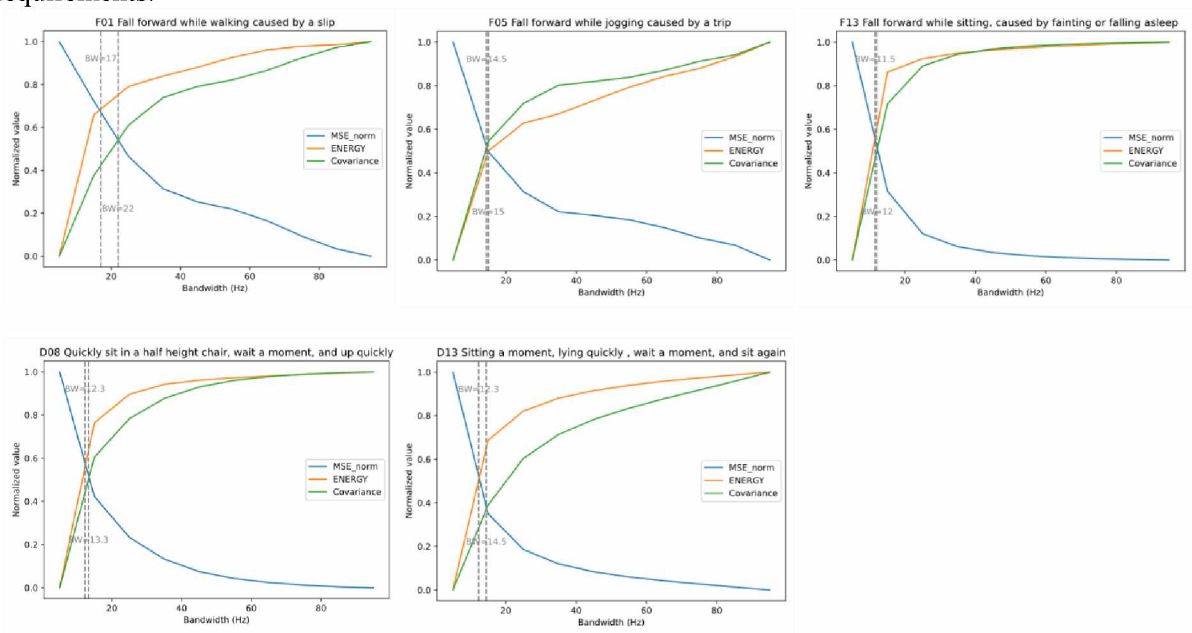

Figure 3. Comparative analysis three metrics of different bandwidths for activities

As shown in the Figure 3, from the experiments conducted, it is apparent that the optimal bandwidth hovers around 15 *Hz*. The experiments lead us to conclude that at this bandwidth, data volume is effectively reduced, facilitating more efficient storage and faster processing, without considerably compromising the signal's overall quality and usability. The bandwidth range of 12 *Hz* to 17 *Hz* emerges as the most favorable, crucial for diminishing data volume whilst maintaining the signal's core attributes. This balance ensures that computational efficiency and the retention of premium signal are optimized, enabling precise activity analysis and recognition.

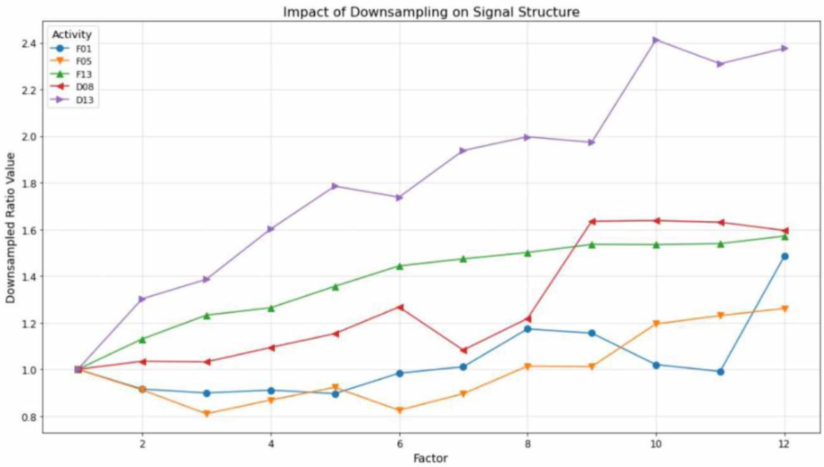

Figure 4. Comparative analysis of downsampling factors using downsampling ratio

As shown in the Figure 4, we observe that although different activities have their optimal downsampling factors at different time intervals, a common trend emerges. For most activities, the ratio is still closest to the ideal value of 1 at the initial interval, but grows to a certain trend and is relatively flat or declines. Choosing a downsampling factor in this range may improve data reduction and maintain signals in different types of activities. Produces an optimal balance between structural integrity.

Table 3. **Difference between downsampled signal and original signal**

Factor	$F01$	$F05$	$F13$	$D08$	$D13$
2	0.122873562	0.243020296	0.036218504	0.041545743	0.01433779
4	0.104106957	0.147924838	0.024735706	0.027431345	0.012322775
6	0.077419495	0.114917775	0.01832282	0.02039018	0.009007972
8	0.053764864	0.091222865	0.011434971	0.016155145	0.007636401
10	0.050401664	0.087729274	0.01005775	0.016219336	0.00729715
12	0.078672189	0.099855768	0.017619462	0.024762008	0.009341281
Best factor	10	10	10	8	10

Table 3 indicates that a downsampling factor of 10 show the smallest difference between downsampled and original signals for most activities, suggesting it is the optimal choice for preserving signal quality while reducing data. The exception is activity $D08$, which has an optimal factor of 8, indicating that this activity retains signal integrity better at a slightly higher sampling rate. This information is essential for efficiently processing accelerometer data without compromising on quality.

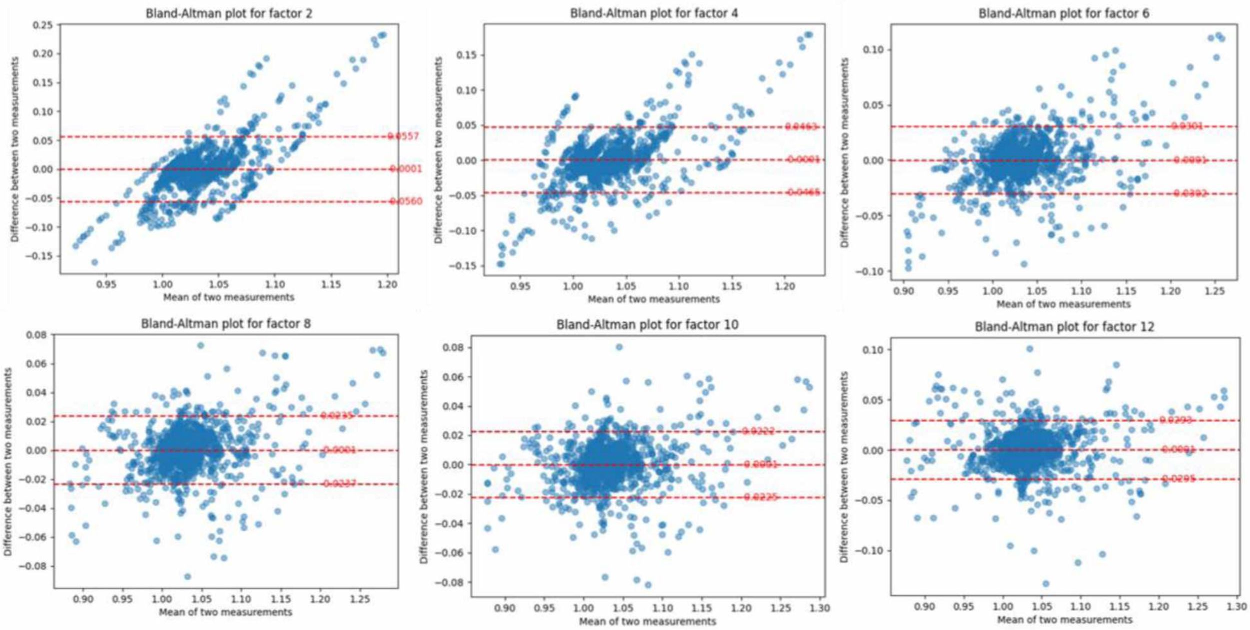

Figure 5. Bland altman plots for original and downsampled data across various downsampling factors

The use of bland altman plots, as shown in Figure 5, complements the findings presented in table 4, serving as a visual affirmation of the quantitative analysis. Bland altman plots are effective for assessing the agreement between two different measurements, in this case, between the downsampled and original data. If the plots show that the differences mostly lie within acceptable limits of agreement and the mean difference is close to zero, it implies a strong consistency between the two sets of data. This visual tool thus provides a clear and intuitive confirmation that a downsampling factor of 10 is indeed effective in retaining the integrity of the signal while significantly reducing the data size, consistent with the analytical results in Table 3.

Conclusion

The study successfully identified an optimal downsampling factor for accelerometer data, balancing data reduction with signal quality preservation. This finding is crucial for enhancing the processing efficiency of wearable technologies and health monitoring systems without sacrificing the accuracy of activity recognition. The results of this research serve as a metric for optimizing signal processing in applications where maintaining data integrity is as important as minimizing computational load.

References

1. Yang C. C., Hsu Y. L. // Sensors. 2010. Vol. 10(8). P. 7772–7788.
2. Picerno P., Iosa M., D'Souza C., et al. // Expert Review of Medical Devices. 2021. Vol. 18(sup1). P. 79–94.
3. Xiao Z. G., Menon C. // Sensors. 2019. Vol. 19(11). P. 2432.
4. Gopinath R. A., Burrus C. S. // IEEE Transactions on Signal Processing. 1994. Vol. 42(4). P. 812–824.
5. Díaz García J., Brunet Crosa P., Navazo Álvaro I., et al. // Proceedings of the International conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2017 and Big Data Analytics, Data Mining and Computational Intelligence 2017: Lisbon, Portugal, July 21–23, 2017. IADIS Press. 2017. P. 12–20.
6. Abdi H., Williams L. J. // Wiley Interdisciplinary Reviews: Computational Statistics. 2010. Vol. 2(4). P. 433–459.
7. Shcherbakov M. V., Brebels A., Shcherbakova N. L., et al. // World Applied Sciences Journal. 2013. Vol. 24(24). P. 171–176.
8. Sucerquia A., López J. D., Vargas–Bonilla J. F. // Sensors. 2017. Vol. 17(1). P. 198.
9. Li H., Ma J., Ren X., et al. // Pattern Recognition and Information Processing. 2023. P. 169–175.