International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 3, June, 2025, pp. 1643-1651
Journal homepage: http://iieta.org/journals/ts

A Multimodal Image Encoding-Driven Visual Analytics Model for Financial Risk Assessment

Yanhua Li

School of Management, Hankou University, Hankou 430212, China

Corresponding Author Email: 2051040091@hkxy.edu.cn

Copyright: ©2025 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ts.420334

Received: 8 November 2024
Revised: 20 April 2025
Accepted: 12 May 2025
Available online: 30 June 2025

Keywords:
multimodal image encoding, financial risk, visual analytics, WLD, directional encoding strategy

ABSTRACT

In the context of digital transformation, enterprise financial data increasingly exhibit multimodal characteristics, with image data containing crucial risk-related information for financial decision-making. However, traditional financial risk analysis methods often underutilize multimodal image data. Existing studies either focus on single-modality encoding strategies—overlooking the heterogeneity and complementarity of multimodal data—or solely extract global image features while neglecting critical local details. Moreover, there is a lack of systematic integration between directional and non-directional feature encoding strategies for financial imagery, limiting their applicability in complex financial scenarios. To address these challenges, this paper proposes a multimodal image encoding fusion framework tailored for financial risk visual analytics. A local encoding scheme is constructed based on the Weber Local Descriptor (WLD), integrating both directional and non-directional encoding strategies. This approach enables efficient encoding and feature fusion of multimodal financial image data. The experimental results demonstrate that the proposed model provides more precise visual representations for financial risk analysis, significantly enhancing risk identification accuracy and decision support capabilities. This work promotes the cross-disciplinary integration of image encoding techniques and financial risk management.

1. INTRODUCTION

With the deep advancement of digital transformation, enterprises are facing increasingly complex financial environments [1-4], and financial data present significant multimodal characteristics, covering various forms such as text, numerical data, and images. Among them, image data such as charts in financial statements and invoice images related to corporate transactions [5-8] contain rich financial risk information. However, traditional financial risk analysis methods mainly rely on structured numerical and textual data [9-11], with insufficient utilization of multimodal image data, making it difficult to comprehensively and intuitively present the complex situation of financial risks. With the rapid development of big data and visualization technologies, how to present complex financial risk information in a visualized way through multimodal image encoding technology has become a key issue to improve the efficiency and accuracy of financial risk analysis.

Conducting research on a multimodal image encoding-driven visual analytics model for financial risk has important practical significance. On the one hand, multimodal image encoding can effectively integrate various image data in the financial field, mine the hidden risk features, and provide richer information sources for financial risk analysis. Through visual analytics [12, 13], abstract financial risks can be transformed into intuitive visual information, helping financial decision-makers quickly grasp the key elements and development trends of risks, and improve the scientific and timeliness of decision-making. On the other hand, this research helps promote the cross-integration of the financial field with image encoding [14] and visualization technologies [15], providing new methods and ideas for financial risk analysis, enriching the theoretical system of financial risk management, and enhancing enterprises' ability to cope with risks in complex financial environments.

At present, some scholars have carried out research on image encoding and visual analytics of financial data. For example, Wang et al. [16, 17] proposed a financial image encoding method based on traditional image feature extraction, but this method only focuses on global features of images and ignores the importance of local detail information for financial risk analysis, resulting in insufficient encoding accuracy in complex financial scenarios. Balbás et al. [18] attempted to apply single-modality encoding strategies to financial risk visual analytics. However, they did not fully consider the differences and complementarity among multimodal image data, leading to poor results when integrating different types of financial image data, and failing to comprehensively capture the multidimensional characteristics of financial risks. In addition, existing studies lack systematic integration of directional and non-directional feature encoding strategies for financial images, making it difficult to meet the diversified feature representation requirements of financial risk visual analytics.

This paper focuses on the research of a multimodal image encoding fusion scheme for financial risk visual analytics. Specifically, a local encoding scheme based on WLD is proposed, which combines two encoding strategies: One is the directional encoding strategy, which targets directional features in financial images such as trend lines, coordinate axes, etc., for targeted encoding to accurately capture the impact of directional information on financial risks; the other is the non-directional encoding strategy, used to process non-directional features in financial images such as the distribution density of data points, regional textures, etc., to ensure comprehensive retention of image detail information. Through the organic combination of these two strategies, efficient encoding and fusion of multimodal financial image data is achieved, providing more accurate and richer feature representations for financial risk visual analytics. The value of this research lies in filling the gap in current research on multimodal financial image encoding and fusion. Through innovative encoding strategies, it effectively enhances the processing ability of financial risk visual analytics models for complex multimodal image data. The proposed scheme can more accurately mine risk features in financial images, provide more intuitive and comprehensive visualized results for financial decision-makers, and help improve the efficiency and accuracy of financial risk analysis, having important theoretical significance and practical application value.

2. MULTIMODAL IMAGE ENCODING FUSION SCHEME ORIENTED TO FINANCIAL RISK VISUAL ANALYTICS

The multimodal image encoding fusion scheme oriented to financial risk visual analytics is based on multimodal images that mainly cover various types of image data in the financial field with value for risk analysis: First, visualized financial statement charts, including line charts reflecting the dynamic changes of financial indicators, bar charts showing data distribution characteristics, heatmaps revealing risk aggregation areas, and scatter plots indicating correlations between indicators. These images carry directional and structural information of financial risks through visual elements such as axis scales, trend directions, and data point densities. Second, financial business bill images, such as scanned invoices, screenshots of bank statements, reimbursement voucher images, and transaction contract bills. These contain texture and pixel features such as stamp clarity, standardization of handwritten digits, and format completeness, which can identify data authenticity and compliance risks through non-directional encoding strategies. Third, hybrid modality images related to enterprise finance, such as heatmaps of electronic financial system interfaces and scatter plots of paper reports. These images often integrate multiple visual elements like text, symbols, and graphics, requiring directional encoding strategies to analyze directional risk features such as abnormal index trends and breakages in interlocking relationships.

This paper proposes a multimodal image encoding fusion scheme oriented to financial risk visual analytics, constructing the core encoding framework based on WLD, focusing on the refined extraction and quantitative expression of local features in images. This scheme first calculates the optimal enhanced direction to determine the key influencing direction of the pixel neighborhood in financial images. Only in this direction are the neighborhood values compared with the central pixel value in multimodal images: when the neighborhood value is greater than the central pixel value, it is encoded as "1", otherwise encoded as "0". This binary encoding method can efficiently capture intensity difference features in local regions of financial images, such as the slope change of trend lines and abnormal fluctuations of data points in financial statement charts, or ink intensity differences of digits and clarity of stamp edges in bill images. Compared with traditional encoding strategies, this method, by limiting to the optimal enhanced direction, avoids redundant computations across all directions in the neighborhood, thereby reducing time cost and significantly improving the recognition accuracy of key features related to financial risks, providing lightweight yet discriminative feature representation for subsequent visual analytics.

The scheme innovatively divides the encoding strategy into two major modules: directional and non-directional, forming a collaborative processing mechanism targeting diverse features of financial images. In the directional encoding strategy, it mainly processes financial image features with clear directional attributes, such as the upward/downward trends of financial indicators, abnormal shifts in axis scales, and flow breakages in interlocking relationships of report data. Through directional comparison based on the optimal enhanced direction, it accurately captures the evolutionary patterns of risk signals in the directional dimension, for example, identifying market risks implied by sharp turns in the revenue trend line of the income statement. The non-directional encoding strategy focuses on local detail features without clear directional attributes, such as abnormal distribution density of chart data points, pixel texture defects in bill images, and grayscale uniformity in heatmap regions. It retains the hidden risk detail information in the image through non-directional comparison between neighborhood values and the central pixel value. Both strategies implement feature encoding based on the same WLD framework and are dimensionally integrated through a multimodal feature fusion algorithm, ultimately generating composite feature vectors containing both directional trends and non-directional details. This provides multi-dimensional feature inputs for financial risk visual modeling, enabling the analysis model to simultaneously present the macro evolution trend and micro abnormal signals of risks in the visualized graph.

2.1 Directional encoding

The directional encoding strategy oriented to financial risk visual analytics first constructs a spatial directional framework of the local neighborhood by defining the main direction, achieving a structured description of directional features in financial images. The schematic diagram of the principle is shown in Figure 1. In this paper, direction 5 is taken as the reference to explain the specific encoding calculation process. In the 5×5 pixel neighborhood of the WLD, eight discrete directions $\phi = 1$ to $\phi = 8$ cover the main spatial orientations of the image local region, and each direction contains two symmetrically distributed pixels, forming a bidirectional constraint on the directional features. The setting of the main direction provides a unified spatial reference system for encoding. In specific cases, such as line chart analysis of financial statements, direction 5 may correspond to the positive direction of the horizontal axis, while directions $\phi =$

3, ϕ = 6 correspond to the ascending or descending slope of trend lines. By associating the optimal enhanced direction representing the most significant change of local features with the main direction, the strategy can accurately locate the dominant direction of financial indicator fluctuations, such as identifying the abnormal steepening of the revenue trend in the income statement or the sudden break of cash flow, providing a spatial coordinate basis for directional quantification of risk. Specifically, suppose the enhanced direction of the image is represented by ϕ_j (= 1, 2, ..., 8), the enhanced current pixel value is denoted as Q_{az}^-, the two different pixels in the same direction in the 5×5 neighborhood of the WLD are represented by $Q_{\phi j}$ and $Q_{\phi j+8}$, and the differences between the values of the two adjacent pixels on both sides of the main direction and the central pixel are denoted as SW and SO. The encoding of the current pixel is denoted as $z_{\phi j}$ and z_z. The two different encodings of the optimal enhanced direction ϕ_j in the neighborhood of the WLD are denoted as z_{jm} and z_{je}. When $1<\phi_j<8$, then:

$$\begin{cases} z_z = z_{\varphi_j} = 1 \\ z_{je} = z_{\varphi_{j+1}} = 1, SW > 0 \\ z_{jm} = z_{\varphi_{j-1}} = 1, SO > 0 \\ 0, otherwise \end{cases} \tag{1}$$

where,

$$\begin{cases} SO = Q_{\varphi_j} - Q_{\overline{a_z}} \\ SW = Q_{\varphi_{j+8}} - Q_{\overline{a_z}} \end{cases}, \varphi_j < 5 \tag{2}$$

$$\begin{cases} SW = Q_{\varphi_j} - Q_{\overline{a_z}} \\ SO = Q_{\varphi_{j+8}} - Q_{\overline{a_z}} \end{cases}, \varphi_j > 5 \tag{3}$$

The core rule of directional encoding revolves around the comparison of neighborhood pixel differences in the optimal enhanced direction ϕ_j, and realizes quantitative expression of directional features through binary encoding. When $1<\phi_j<8$, the strategy generates encoding values by comparing the adjacent pixel values SO and SW on both sides of the main direction: if SO>0, the left encoding bit ckl=1; if SW>0, the right encoding bit z_{je}=1; in other cases, it is 0. This rule can effectively capture the intensity variation of directional features in financial images. For example, in the axis area of a balance sheet, a high encoding value in direction ϕ_j=3 may indicate abnormal expansion of asset items in that direction, while a low encoding value may reflect a contraction trend of liability items. For special directions ϕ_j=1 and ϕ_j=8 at the boundary of the neighborhood, the strategy maintains spatial symmetry of the encoding by shifting high and low bits, e.g., shifting to higher bits for ϕ_j=1 and to lower bits for ϕ_j=8, to avoid feature loss in boundary directions. Finally, the grayscale feature value $D(\phi_j)$ of the central pixel is obtained by summing the 8-bit encoding, converting the directional risk signal into a computable and comparable quantitative indicator, providing a numerical basis for subsequent visual analysis. The specific summation formula is:

$$D\left(\varphi_j\right) = \begin{cases} z_{jm} * 2^{j-2} + z_z * 2^{j-1} + z_{je} * 2^{j2}, 1 < \varphi_j < 8 \\ z_{8l} * 2^5 + z_z * 2^6 + z_{8e} * 2^7, \varphi_j = 8 \\ z_{1m} * 2^0 + z_z * 2^1 + z_{1e} * 2^2, \varphi_j = 1 \end{cases} \tag{4}$$

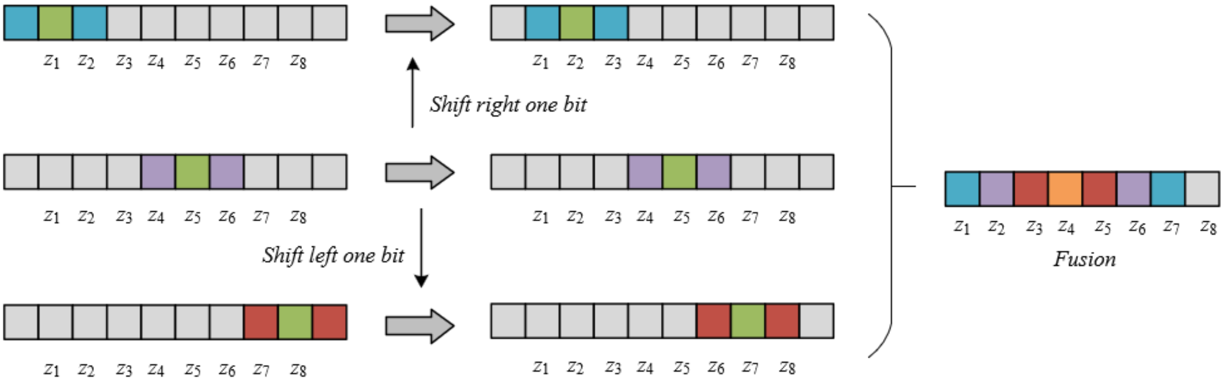


Figure 1. Schematic diagram of multimodal image directional encoding principle

The directional encoding strategy, through collaboration with the multimodal image fusion framework, realizes a three-dimensional visual presentation of financial risks. In the fusion process, the optimal enhanced direction of each single modality is first aligned with the main direction, ensuring that the directional features of different modalities are encoded and integrated under a unified spatial reference. In specific cases, when integrating the income statement line chart with the cash flow arrow diagram, aligning the main direction allows spatial superposition of revenue growth trends and cash outflow trends in the visual graph, intuitively showing the risk transmission relationship between the two. The encoding results are further mapped to parameters of visual elements, such as the color depth of different directions in a risk heatmap, the thickness and orientation of arrows in a flow diagram, enabling financial decision-makers to quickly capture the dominant direction and evolution path of risks through visual perception. Figure 2 shows an example of directional encoding.

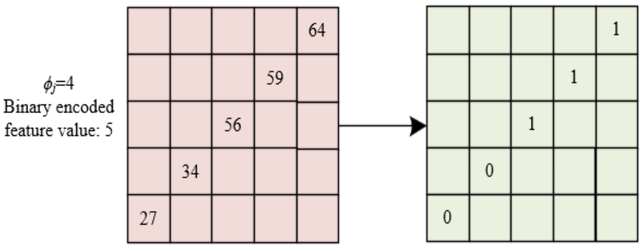


Figure 2. Example of directional encoding

2.2 Non-directional encoding

The non-directional encoding strategy oriented to financial risk visual analytics constructs the encoding reference based on a fixed main direction, solving the quantification problem of features without clear spatial directionality in financial images. The schematic diagram of the principle is shown in Figure 3. This paper still takes direction 5 as the reference, and the specific encoding calculation process is as follows:

$$\begin{cases} z_5 = 1 \\ z_6 = 1, SW > 0 \\ z_4 = 1, SO > 0 \\ 0, otherwise \end{cases} \tag{5}$$

where,

$$\begin{cases} SO = Q_{\varphi_j} - Q_{\overline{a_z}} \\ SW = Q_{\varphi_{j+8}} - Q_{\overline{z_z}} \end{cases}, \varphi_j < 5 \tag{6}$$

$$\begin{cases} SW = Q_{\varphi_j} - Q_{\overline{a_z}} \\ SO = Q_{\varphi_{j+8}} - Q_{\overline{a_z}} \end{cases}, \varphi_j > 5 \tag{7}$$

Unlike the dynamic optimal enhanced direction in the directional strategy, this strategy forces the encoding value of the main direction bit to be set as 1, forming a stable local feature reference point, focusing on capturing risk signals that do not rely on directional changes. In the encoding rule design, only the adjacent directions ϕ=4 and ϕ=6 on both sides of the main direction are activated. By comparing the pixel values on both sides with the central pixel value, binary quantification of non-directional features is achieved. For example, if *SO*>0, then z_4 is encoded as 1; if *SW*>0, then z_6 is encoded as 1; all other directions are set to 0 by default. In specific cases, in financial bill images, edge blurring of invoice stamps or uneven ink intensity of digits in bank statements can be accurately identified by changes in encoding values of z_4 or z_6, without focusing on the specific direction of the anomaly, providing stable detail feature input for compliance inspection.

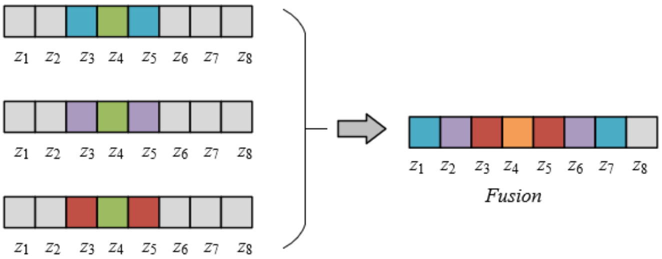


Figure 3. Schematic diagram of multimodal image non-directional encoding principle

The non-directional encoding strategy calculates the grayscale feature value $D(\phi_j)$ by summing the 8-bit binary encoding, achieving efficient aggregation of non-directional risk details in financial images. Since the main direction bit is fixed as 1, and only z_4 and z_6 are dynamically encoded based on pixel differences, while all other directions are set to 0, the value range of $D(\phi_j)$ is limited to 0–3, forming a lightweight feature expression. This design is particularly suitable for scenarios such as abnormal distribution density of data points and abrupt changes in grayscale values in heatmaps. In specific cases, in heatmap analysis of income statements, a sudden increase in $D(\phi_j)$ in a certain region reflects abnormal coupling of multi-dimensional financial indicators without the need to distinguish the specific direction of risk evolution. In reimbursement voucher images, local fluctuations in $D(\phi_j)$ can locate format misalignment or pixel tampering areas, providing quantitative basis for rapid annotation of audit risk points. The specific calculation method is given as:

$$D(\varphi_j) = z_4 \times 2^3 + z_5 \times 2^4 + z_6 \times 2^5 \tag{8}$$

The core difference between the non-directional encoding strategy and the directional strategy lies in the mechanism of enhanced direction selection: the former eliminates directional dependence by fixing the main direction, while the latter captures directional features based on dynamic optimal direction. The two form a complementary feature encoding system. In multimodal fusion, the significant advantage of the non-directional strategy lies in avoiding the "high-low bit imbalance" problem that may be caused by directional encoding. In the directional strategy, some directions may correspond to high/low encoding bits for a long time, easily leading to feature bias or detail blurring in the visualized image. The non-directional strategy, by forcibly setting the main direction bit to 1 and only activating the two side directions, ensures spatial symmetry and balance of the encoding values. In specific cases, when integrating scanned paper reports and financial indicator trend charts, the non-directional encoding can clearly retain texture defects of stamps in scanned documents, while not interfering with the directional strategy's detection of trend line slope anomalies. Ultimately, a layered display of "texture detail – trend direction" can be realized in the visualized graph, enhancing financial decision-makers' multi-dimensional insight into complex risk scenarios.

2.3 Encoding fusion framework

The proposed multimodal image encoding fusion scheme for financial risk visual analysis is based on the WLD's 5×5 neighborhood 8-direction division, constructing an 8-bit binary coding system [z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8] for integrating feature information of multimodal financial images. The framework structure is shown in Figure 4. This framework maps the local feature encoding of each modality into binary sequences, where (z_4, z_6), (z_3, z_7), and (z_2, z_8) correspond respectively to the feature encoding results of the target pixel under different modalities. Through permutation and combination, six sorting methods are formed, aiming to extract commonality and differential features with risk analysis value from multisource data such as financial report charts, bill images, and hybrid modal interface screenshots. In a specific case, when integrating income statement line charts with invoice scans, the binary coding system can simultaneously retain trend line directions and stamp textures, providing multidimensional feature input for subsequent risk visualization.

The core of the scheme lies in determining the fusion sorting rules of multimodal features through optimal encoding bit filtering. For the directional encoding strategy, during fusion, the optimal enhanced direction of each single modality needs to be aligned with the main direction of the fusion framework, ensuring that directional features of different modalities are compared and integrated under a unified spatial reference. In a specific case, when fusing balance sheet line charts with cash flow scatter plots, the two modalities' optimal enhanced directions are first adjusted to the main direction, then the encoding bits are reorganized according to sorting rules such as (z_3, z_7), enabling explicit expression of cross-modal directional risk signals. For the non-directional encoding strategy, since the main direction is fixed as direction 5, the fusion only needs to reorder the main direction's two adjacent encoding bits (z_4, z_6) of each single modality according to preset rules to achieve cross-modal aggregation of non-directional features such as texture anomalies in bill images and density differences in chart data points, avoiding coding imbalance problems caused by directional changes.

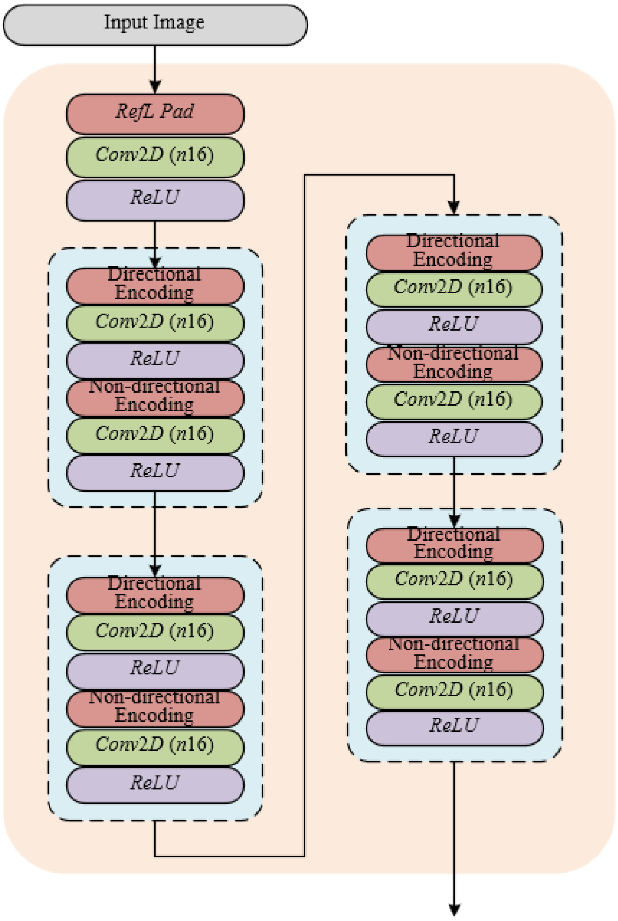


Figure 4. Multimodal image encoding fusion framework for financial risk visual analysis

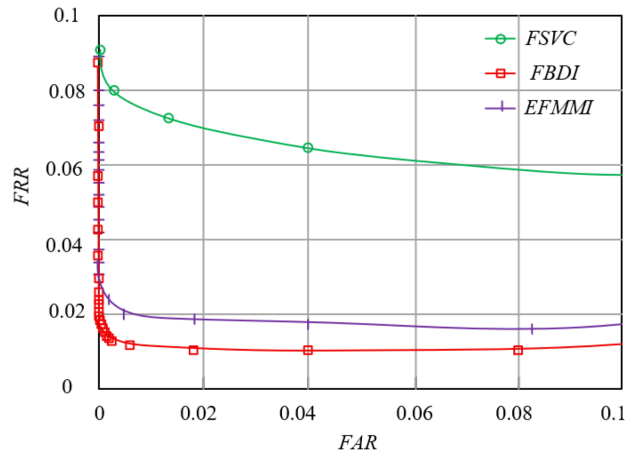
The fusion scheme designs a standardized two-stage processing flow: firstly, for single-modality feature encoding, main direction alignment is performed according to modality type. Directional modalities are dynamically adjusted to the main direction based on the optimal enhanced direction, while non-directional modalities directly use the fixed main direction as the reference; secondly, the 8-bit encoding is reorganized according to the selected sorting method to generate a composite feature vector containing multimodal information. In financial risk analysis scenarios, this process effectively handles feature heterogeneity among different modalities. In a specific case, when integrating electronic financial system interface screenshots with scanned paper reports, directional encoding aligns the main direction to accurately capture abnormal directions of indicator warning arrows in the interface, while non-directional encoding retains pixel defects of stamps in scanned documents, finally forming a "risk direction - detail anomaly" associative display in the visualized graph, assisting decision-makers in rapidly locating risk resonance points in multisource data.

The scheme chooses the non-directional encoding strategy as the basis for single-modality encoding, due to its significant advantages in detail retention and coding balance in financial images. Compared with the directional encoding which may cause high-low bit imbalance issues, the non-directional strategy ensures spatial symmetry of encoding values by fixing the main direction and activating only the two adjacent encoding bits of the main direction, making texture features of bill images clearer and more distinguishable in visualization results. In a specific case, during audit compliance checks, bill image features fused based on non-directional encoding can quickly identify pixel tampering or format errors in reimbursement vouchers by precise localization of grayscale abnormal regions, significantly improving risk identification accuracy. Moreover, the fused composite feature vector can effectively integrate non-directional details of multimodal data with trend features from directional strategy, forming layered visualization of "macro trend layer - micro detail layer" in risk heatmaps, providing financial decision-makers with a stereoscopic analysis perspective from overall risk trends to local abnormal signals, meeting the demand for multidimensional risk insights in complex financial scenarios.

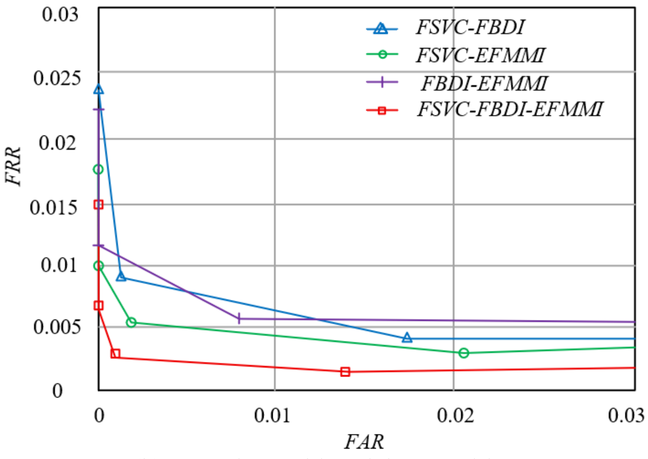
In enterprise financial risk management practice, the constructed multimodal image encoding-driven financial risk visual analysis model can be widely applied in various core scenarios: In internal comprehensive financial analysis scenarios, the model can integrate multimodal financial chart images such as balance sheet line charts, income statement heatmaps, cash flow scatter plots, capturing directional features such as abnormal slope changes of trend lines and axis scale shifts through directional encoding strategy, combined with non-directional encoding strategy parsing nondirectional details such as data point distribution density and regional texture uniformity, generating visualized risk distribution maps that assist financial staff in locating risk areas like revenue fluctuations and cash flow anomalies; In financial institution credit evaluation scenarios, the model can fuse multimodal data such as scanned paper financial statements submitted by enterprises, electronic bill images, transaction contract screenshots, encoding non-directional features like stamp clarity and numeric writing norms on bills by WLD, while analyzing deviation degrees of financial indicator trend lines through directional encoding strategy, constructing a visual enterprise credit risk matrix providing intuitive risk reference for credit decisions; In audit compliance inspection scenarios, the model can handle large amounts of invoice images, reimbursement voucher scans, bank statement screenshots, recognizing abnormal data reconciliation directions between reports via directional encoding, and detecting pixel anomalies, format misalignment and other detail features in voucher images via non-directional encoding, presenting risk point distribution during audits as visual graphs to improve efficiency and accuracy of compliance checks; In securities investment analysis scenarios, the model integrates multimodal data such as listed company financial report chart images, transaction data visualization interface screenshots, industry trend heatmaps, capturing features like stock price fluctuation trends and financial indicator correlation directions via directional encoding strategy, combined with analysis of data point abnormal clustering and interface display anomalies via non-directional encoding strategy, generating investment risk visual analysis reports, providing investors with multidimensional risk judgment bases. The above application scenarios all rely on the efficient integration and feature extraction ability of multimodal image encoding technology for financial image data, transforming abstract financial risks into visually interpretable information through visual analysis, offering systematic risk insight support for financial decisions of different stakeholders.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The unimodal ROC curves in Figure 5(a) show that the FSVC financial statement visualization charts have an FRR of about 0.08 at low FAR, which gradually decreases as FAR increases, reflecting its capability to capture directional risk in trend-type charts; the FBDI financial business bill images have an FRR of 0.04 at FAR=0.01 with a smooth curve, indicating stability in compliance risk identification of bill details; the EFMMI mixed-modal images have the lowest initial FRR but performance degrades at high FAR, reflecting the initial filtering advantage of mixed features. Among the three, FBDI exhibits the best unimodal performance, while FSVC and EFMMI complement each other in different FAR intervals.



(a) ROC for unimodal



(b) ROC for multimodal recognition

Figure 5. Comparison of financial risk identification results using different modality combinations

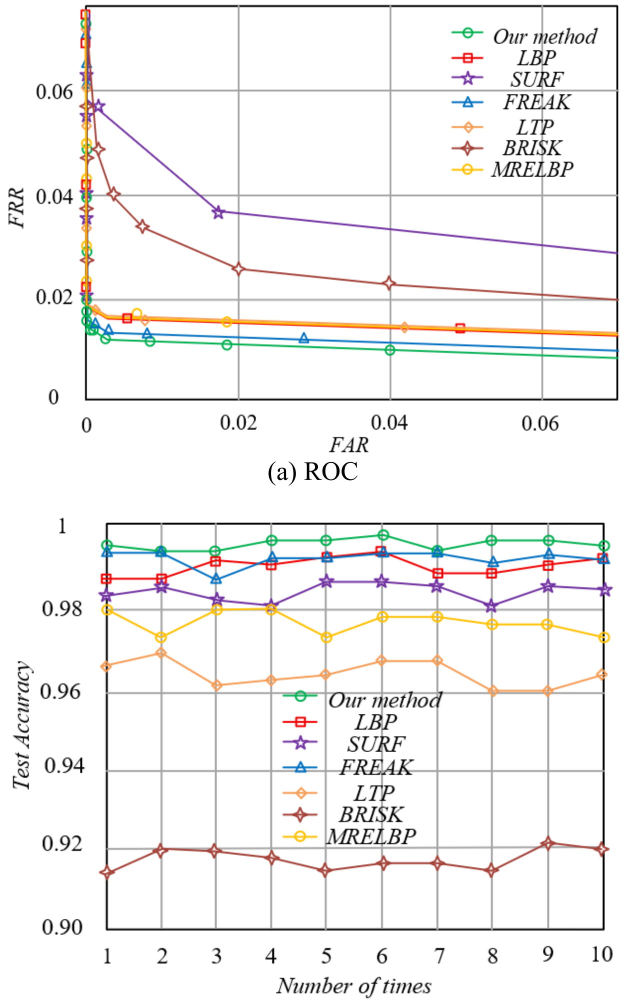
In the multimodal combinations shown in Figure 5(b), the three-modal fusion FSVC-FBDI-EFMMI achieves an FRR of only 0.005 at FAR=0.01, significantly lower than bimodal combinations, and maintains below 0.005 at FAR=0.03, much better than bimodal fusions. This indicates that multimodal encoding fusion, by integrating directional and non-directional strategies, achieves dual improvements in risk identification accuracy and robustness: the directional strategy enhances spatial localization of trend risks, while the non-directional strategy ensures balanced encoding of bill details and mixed features. Their synergy makes the three-modal fusion perform excellently across the entire FAR range, verifying the scheme's advantage in multidimensional feature complementarity.

Table 1. Comparison of financial risk identification performance with different fusion methods

Modality Combination	Equal Error Rate ($\times 10^{-2}$)
FSVC	1.25
FBDI	6.27
EFMMI	1.78
FSVC-FBDI	0.73
FSVC-EFMMI	0.51
FBDI-EFMMI	0.64
FSVC-FBDI-EFMMI	0.27

The equal error rate data in Table 1 clearly shows the impact of modality combinations on financial risk identification accuracy. Among unimodal modalities, FSVC, leveraging directional encoding to capture report trend directions, performs excellently in trend-type risk identification; FBDI, focusing on bill details via non-directional encoding, has relatively lower unimodal performance, but its fusion with FSVC reduces the equal error rate by 48.8%, reflecting preliminary complementarity between directional and non-directional features. EFMMI, as a mixed modality integrating multidimensional financial features, further improves performance when fused with FSVC, indicating the enhancing effect of mixed features on trend directions. The three-modal fusion FSVC-FBDI-EFMMI has an equal error rate only 4.3% of unimodal FBDI, and reduces 47.1% compared with the best bimodal combination FSVC-EFMMI. This data shows that directional encoding precisely locates trend risks, non-directional encoding comprehensively preserves details, and their combination enables deep fusion of risk features at the encoding layer, avoiding blind spots of unimodal features. The gradual decrease of equal error rate from unimodal to bimodal to trimodal verifies effective utilization of feature redundancy across modalities. FSVC's trend directions, FBDI's bill details, and EFMMI's mixed features form a "direction-detail-multidimensional" stereoscopic coverage in risk identification, significantly improving model adaptability to complex financial scenarios.

The ROC curve in Figure 6(a) indicates that the proposed method achieves extremely low miss rates at low false alarm rates. When FAR is 0.01, our method's FRR is about 0.02, significantly lower than traditional methods such as LBP and SURF, and remains about 0.01 as FAR extends to 0.06, far outperforming others. This demonstrates the high sensitivity of our method in identifying financial risks, effectively capturing risk signals under extremely low false alarm conditions. The ACC curve in Figure 6(b) shows that the testing accuracy of our method is consistently above 98%, with the smallest fluctuations across multiple experiments, while LBP, SURF, and others have lower accuracy and worse stability. This result validates that the combination of directional and non-directional encoding enables our method to comprehensively capture multidimensional features of financial images. In processing FSVC, directional encoding precisely locates abnormal directions in revenue trends, while non-directional encoding balances pixel differences in bill textures in FBDI, jointly improving feature discriminability and yielding superior ROC curves. Based on the WLD's 8-bit binary coding, our method achieves much lower computational complexity during multimodal data fusion than traditional methods, ensuring high stability of ACC and meeting real-time requirements of mobile VR/AR scenarios.



(a) ROC

(b) ACC

Figure 6. Comparison of financial risk identification performance using different multimodal image feature expression methods

Table 2 comprehensively demonstrates the performance advantages of our method from the three dimensions of equal error rate, average accuracy, and processing time. In terms of equal error rate, our method's rate is 1.23×10^{-2}, significantly lower than traditional methods such as MRELBP and SURF, and only slightly higher than LTP; however, through the synergy of directional and non-directional encoding strategies, it achieves better feature discrimination in multimodal fusion, effectively reducing risk recognition false positives and false negatives. Regarding average accuracy, our method reaches 98.57% ± 0.12%, higher than LTP, BRISK, and other methods, with the smallest standard deviation, reflecting very high stability in complex financial scenarios and ensuring consistency of risk recognition results. In terms of processing time, our method and methods such as FREAK and LTP all operate between 0.22-0.26 seconds, outperforming MRELBP. This advantage benefits from the lightweight design of binary encoding based on the WLD, meeting the stringent real-time requirements of mobile VR/AR scenarios.

Table 2. Comparison of financial recognition results with different multimodal image feature expression methods

Methods	**Equal Error Rate ($\times10^{-2}$)**	**Average Accuracy (%)**	**Time (s)**
MRELBP	3.35	97.52±0.22	0.26
BRISK	1.34	95.23±0.35	0.22
LTP	1.28	98.41±0.22	0.25
FREAK	1.47	97.27±0.15	0.22
SURF	2.52	92.89±0.21	0.23
LBP	1.69	96.37±0.31	0.26
Our method	1.23	98.57±0.12	0.22

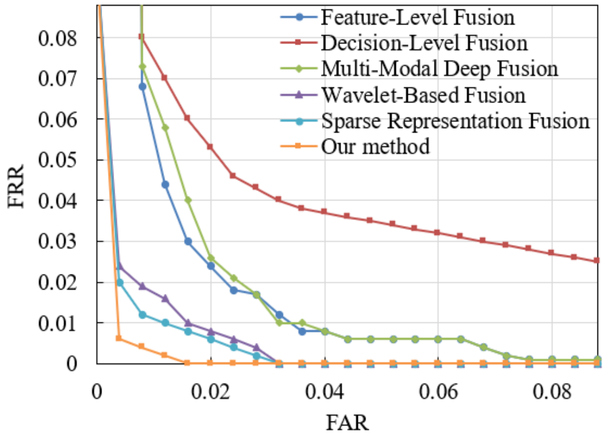


Figure 7. ROC curves of different multimodal image fusion methods

The ROC curves in Figure 7 visually present the risk recognition performance of each fusion method. Our method maintains the lowest FRR across the full FAR range. For example, when FAR is 0.01, our method's FRR is about 0.01, far lower than traditional methods such as feature-level fusion and decision-level fusion; as FAR increases to 0.08, our method's FRR approaches 0, while other methods still show obvious misses. This result shows that the combination of directional and non-directional encoding allows our method to comprehensively capture risk features in multimodal financial images. In FSVC, directional encoding dynamically encodes ϕ_j to precisely locate trend anomalies; in FBDI, non-directional encoding activates z_4/z_6 to retain bill details; in EFMMI, both synergistically integrate multidimensional features, forming a "direction-detail-multimodal" three-dimensional risk description, significantly improving the concavity of the ROC curve and enhancing sensitivity and specificity of risk recognition. Based on the WLD's 8-bit binary encoding, our method avoids complex feature-level or decision-level calculations, has low computational complexity, and ensures FRR remains close to 0 at very low FAR, meeting the stringent real-time and high-precision requirements of mobile VR/AR scenarios.

4. CONCLUSION

For financial risk visual analysis, this paper proposed a multimodal image encoding scheme based on the WLD, integrating directional and non-directional strategies: the former captured report trends, and the latter preserved bill details, achieving efficient fusion of FSVC, FBDI, and EFMMI. Experimental data show that the three-modal fusion's equal error rate was as low as 0.27×10^{-2}, average accuracy reached 98.57%, and the ROC curve outperformed traditional methods, verifying the scheme's advantages in risk identification accuracy and computational efficiency. The scheme, through "direction-detail-multimodal" feature complementarity, provided technical support for immersive financial analysis, promoting financial risk visualization toward stereoscopic and real-time development, and improving decision timeliness. By dual encoding strategies and multimodal fusion, this paper provided an innovative technical solution for financial risk visual analysis, achieving breakthroughs in accuracy, efficiency, and scenario adaptability, laying a foundation for immersive financial analysis in mobile VR/AR and similar scenarios. Future work needs to continuously explore modality expansion, strategy optimization, and scenario deepening to further improve the model's generalization ability and application value, advancing financial risk analysis toward more intelligent and immersive directions.

Current research limitations include insufficient modality coverage and universality constraints of encoding strategies. Future efforts should expand multimodal data sources and construct datasets with stronger generalization; optimize encoding strategies by combining deep learning to enhance robustness of feature expression. Meanwhile, deepening VR/AR interaction design and strengthening immersive analysis experience will promote financial risk visual analysis toward intelligence and interactivity upgrades, further exploiting multimodal encoding's application value in complex scenarios, providing more accurate technical support for risk decision-making, and facilitating technological innovation in financial analysis scenarios.

REFERENCES

[1] Luo, S. (2022). Digital finance development and the digital transformation of enterprises: Based on the perspective of financing constraint and innovation drive. Journal of Mathematics, 2022(1): 1607020. https://doi.org/10.1155/2022/1607020

[2] Peng, Z., Huang, Y., Liu, L., Xu, W., Qian, X. (2024). How government digital attention alleviates enterprise financing constraints: An enterprise digitalization perspective. Finance Research Letters, 67: 105883. https://doi.org/10.1016/j.frl.2024.105883

[3] Li, Y., Zhang, Y., Geng, L. (2024). Digital finance, financing constraints and supply chain resilience. International Review of Economics & Finance, 96: 103545. https://doi.org/10.1016/j.iref.2024.103545

[4] Zhao, K., Shan, H., Chen, Z., Wu, W. (2024). Can the development of digital finance stimulate enterprise innovation? Empirical evidence from China. Economics of Innovation and New Technology, 33(7): 979-1001. https://doi.org/10.1080/10438599.2023.2266376

[5] Luo, S., Yu, J. (2024). SGFNet: A semantic graph-based multimodal network for financial invoice information extraction. Expert Systems with Applications, 258: 125156. https://doi.org/10.1016/j.eswa.2024.125156

[6] Farrell, K.T., Lute, J.E. (2005). Document-management technology and acquisitions workflow: A case study in invoice processing. Information Technology and Libraries, 24(3): 117-122. https://doi.org/10.6017/ital.v24i3.3372

[7] Kim, Y., Hwang, S., Park, J., Kim, J. (2023). Delivery invoice information classification system for joint courier logistics infrastructure. Computers, Materials & Continua, 75(2): 3027-3044. https://doi.org/10.32604/cmc.2023.027877

[8] Khandokar, I.A., Deshpande, P. (2024). Computer vision-based framework for data extraction from heterogeneous financial tables: A comprehensive approach to unlocking financial insights. IEEE Access., 13: 17706-17723. https://doi.org/10.1109/ACCESS.2024.3522141

[9] Nazarova, K., Bezverkhyi, K., Hordopolov, V., Melnyk, T., Poddubna, N. (2021). Risk analysis of companies' activities on the basis of non-financial and financial statements. Agricultural and Resource Economics: International Scientific E-Journal, 7(4): 180-199. https://doi.org/10.22004/ag.econ.316827

[10] Aljadani, A. (2024). Extreme PORT for Norwegian fire financial claims: Empirical assessment and financial VAR analysis. Alexandria Engineering Journal, 108: 852-862. https://doi.org/10.1016/j.aej.2024.09.035

[11] Zadorozhnyy, Z.M., Zhukevych, S., Portovaras, T., Rozelyuk, V., Zhuk, N., Nazarova, I. (2023). Analysis of risks in the financial security management system of business entities. Financial and Credit Activity-Problems of Theory and Practice, 6(53): 82-95. https://doi.org/10.55643/fcaptp.6.53.2023.4242

[12] Grzybowska, U., Karwański, M. (2022). Archetypal analysis and DEA model, their application on financial data and visualization with PHATE. Entropy, 24(1): 88. https://doi.org/10.3390/e24010088

[13] Didimo, W., Liotta, G., Montecchiani, F. (2014). Network visualization for financial crime detection. Journal of Visual Languages & Computing, 25(4): 433-451. https://doi.org/10.1016/j.jvlc.2014.01.002

[14] Barra, S., Carta, S.M., Corriga, A., Podda, A.S., Recupero, D.R. (2020). Deep learning and time series-to-image encoding for financial forecasting. IEEE/CAA Journal of Automatica Sinica, 7(3): 683-692. https://doi.org/10.1109/JAS.2020.1003132

[15] Lokanan, M.E. (2022). Financial fraud detection: the use of visualization techniques in credit card fraud and money laundering domains. Journal of Money Laundering Control, 26(3): 436-444. https://doi.org/10.1108/JMLC-04-2022-0058

[16] Wang, D., Wang, T., Florescu, I. (2020). Is image encoding beneficial for deep learning in finance? IEEE Internet of Things Journal, 9(8): 5617-5628. https://doi.org/10.1109/JIOT.2020.3030492

[17] Wang, W., Lu, M., Dai, X., Jiang, P. (2024). Financial digital images compression method based on discrete cosine transform. Automatic Control and Computer Sciences, 58(5): 592-601. https://doi.org/10.3103/S014641162470069X

[18] Balbás, A., Balbás, B., Galperin, I., Galperin, E. (2008). Deterministic regression model and visual basic code for

optimal forecasting of financial time series. Computers & Mathematics with Applications, 56(10): 2757-2771. https://doi.org/10.1016/j.camwa.2008.07.032