Physicochem. Probl. Miner. Process., 62(4), 2026, 230977 | Physicochemical Problems of Mineral Processing

http://www.journalssystem.com/ppmp | ISSN 1643-1049

Multi-scale feature fusion and adaptive regularization for dynamic feature analysis of coal flotation froth

Cuicui Zheng [1], Xianwu Huang [1], Zhao Cao [2], Haili Shang [2], Dahua Yu [3]

[1] School of Digital and Intelligence Industry, Inner Mongolia University of Science &Technology, Baotou, China
[2] School of Mines and Coal, Inner Mongolia University of Science & Technology, Baotou, China
[3] School of Automation and Electrical Engineering, Inner Mongolia University of Science & Technology, Baotou, China

Corresponding author: huangxianwu1983@163.com (Xianwu Huang)

Abstract: Fine coal flotation is one of the most widely used separation techniques in coal processing and utilization, and the dynamic characteristics of flotation froth serve as key indicators for characterizing variations in flotation operating conditions. To address the challenges of feature extraction from fine coal flotation froth images under complex conditions, this paper proposes, for the first time, a feature encoding network that integrates a feature pyramid with shallow feature enhancement and incorporates it into a keypoint detection framework. Through multi-scale feature fusion and low-level feature enhancement, the proposed network improves the representation capability for froth structures at different scales, significantly enhancing the detectability and stability of feature points in froth images. In addition, a Spatially Adaptive Keypoint Regularization method is designed to address the issue of large local deviations in existing velocity estimation approaches. Furthermore, by integrating the proposed method with LightGlue, high-precision matching and dynamic feature extraction in complex froth scenarios are achieved. Experimental results demonstrate that, on the HPatches dataset, the proposed method achieves a repeatability of 0.605 and a homography estimation accuracy of 0.433. On a self-constructed coal slurry froth image dataset, the repeatability is improved to 58.4%, while the velocity entropy is reduced to 3.441. Moreover, the method exhibits low inference latency and low GPU memory consumption, meeting the requirements for industrial real-time monitoring, and providing more reliable technical support for real-time perception of flotation conditions and performance evaluation of the separation process.

Keywords: fine coal flotation, foam images, feature extraction, multi-scale features, regularization

1. Introduction

Fine coal flotation is a critical process in coal processing and clean utilization and also constitutes a core link in mineral processing operations (Zocche et al., 2023; He et al., 2024). The primary objective of this process is to achieve efficient separation of coal from gangue via froth flotation, where process stability and separation efficiency directly determine the final product quality and economic benefits. During flotation, the dynamic visual characteristics of the froth directly reflect the production conditions and are correlated with separation performance, serving as key parameters for predicting production indices and monitoring process optimization. Therefore, accurate perception and quantitative characterization of the dynamic surface features of the froth are not only essential for understanding the flotation mechanism and enhancing separation performance but also represent the fundamental starting point for analyzing the dynamic characteristics of fine coal flotation froth (Aldrich et al., 2022; Lu et al., 2024; Ma et al., 2025).

In actual flotation production, the assessment of froth conditions and the adjustment of process parameters still primarily rely on operators' empirical observation of froth appearance features, such as froth density, color changes, and flow velocity, based on which they manually adjust reagent dosages or operational conditions (Liu et al., 2011; Nayak et al., 2022; Wang et al., 2024). This approach is highly subjective and unstable, making it difficult to adapt to complex and variable working conditions, often

DOI: 10.37190/ppmp/230977

leading to fluctuations in production indices. Furthermore, the froth images generated during fine coal flotation exhibit visual characteristics such as low texture, strong reflections, and variable morphology. These factors pose significant challenges for the robust extraction and motion modeling of froth features, thereby further constraining the advancement of refined monitoring and intelligent control of the flotation process.

To address these challenges, feature-point-based image matching and motion analysis methods offer an efficient and feasible technical pathway for extracting froth motion parameters and characterizing spatial distribution properties (Chen et al., 2016).

Traditional feature extraction methods are based on manually constructed mathematical transformation theories, relying on predefined feature discrimination criteria to locate and extract local image features. The SIFT algorithm proposed by Lowe (2004) constructs feature point descriptors with scale and rotation invariance through scale-space extreme value detection and gradient histogram description. Combined with the RANSAC strategy, it can effectively eliminate mismatched points. However, the high-dimensional descriptors and complex pyramid construction process of this method lead to high computational complexity, making it difficult to meet the real-time requirements of industrial online monitoring. The ORB algorithm proposed by Rublee et al. (2011) employs FAST feature detection and BRIEF descriptors, improving rotation invariance through rBRIEF, thereby enhancing extraction speed and matching efficiency. However, it lacks scale invariance, resulting in insufficient adaptability to scenarios with drastic froth scale changes, and it is also sensitive to image rotation. Lu et al. (2020) proposed the R-K algorithm, which constructs a dynamic feature matching model for flotation froth by integrating Kalman filtering and the RANSAC algorithm. Although this method can suppress noise and optimize trajectory continuity, it still relies on manually designed feature operators and has limitations in real-time performance and adaptability to weak-texture scenes. Zhang and Zhao (2023) proposed an improved SIFT algorithm that constructs a scale space using nonlinear diffusion filtering, calculates uniform gradient information using the multi-scale Sobel operator and the multi-scale exponentially weighted mean ratio operator, and adopts an image blocking strategy to extract Harris feature points, thereby obtaining stable and uniformly distributed point features. Kaartinen et al. (2006) estimated froth displacement using image sub-block cross-correlation, but this method suffers from high computational load and low processing efficiency.

In recent years, deep learning techniques, with their powerful feature learning and adaptive capabilities, have been widely applied and have achieved remarkable results in the field of image feature extraction, offering a new technical pathway for froth image feature extraction. SuperPoint, proposed by Detone D et al. Detone et al. (2018), serves as an end-to-end feature point detection and description method based on convolutional neural networks. It synchronously learns the spatial distribution of feature points and local description information through a shared feature encoder, overcoming the limitations of traditional methods that rely on manually designed operators (Shi et al., 2022). Building on this, KeyNet (Barroso-Laguna and Mikolajczyk, 2023) enhances feature extraction capability under weak texture and repetitive structures through a learnable detection module, employing a hierarchical response selection mechanism to improve feature point repeatability. D2-Net (Dusmanu et al., 2019) innovatively uses descriptors for both detection and description, enhancing stability under illumination and viewpoint changes. R2D2 (Revaud et al., 2019) reduces computational load while maintaining high recall through multi-scale response aggregation and repeatability prediction, making it suitable for real-time scenarios. However, these deep learning methods are all designed for natural images; their network structures and training strategies do not account for the typical characteristics of industrial scenes such as low texture, high noise, and strong dynamic changes present in flotation froth, leading to persistent shortcomings in froth image analysis.

Nevertheless, both traditional methods and existing deep learning approaches are designed for natural images, and their network structures do not consider the specific characteristics of fine coal flotation froth. In the complex industrial production scenario of fine coal flotation, froth images typically exhibit low contrast, weak texture structures, blurred boundaries, and strong noise interference. The transient changes in froth morphology, high heterogeneity of surface texture, and feature degradation under strong interference significantly limit the applicability of the aforementioned methods.

Addressing the complexity of the above environment, this paper proposes a novel method for feature extraction and dynamic feature analysis of fine coal flotation froth images. After preprocessing froth images from industrial sites, effective analysis of froth dynamic features is achieved. The main innovations are as follows:

(1) To address the issues of feature point loss and unstable extraction in high-interference, non-steady-state environments, a multi-scale feature point extraction network, MFNet, is constructed. By integrating feature pyramid networks and enhancing low-level features, the representation of froth features is strengthened, effectively improving the detectability and stability of froth feature points.

(2) To tackle the problem of trajectory inconsistency in froth motion estimation caused by local noise and scale variations, a method combining detection confidence with spatially adaptive keypoint regularization is designed. This dynamically constrains the spatial distribution of feature points, effectively improving the global consistency of froth motion velocity estimation, significantly reducing motion estimation errors, and enhancing the reliability of trajectory tracking.

(3) To address the challenges of matching complex froth group motions and accurately characterizing their dynamic features, a froth motion feature analysis framework integrating MFNet and LightGlue is constructed. This framework achieves high-precision feature matching and dynamic feature extraction for complex froth motions, filling the gap in precise dynamic analysis for complex flotation scenarios.

2. Material and methods

As a representative self-supervised keypoint detection network, SuperPoint (Zhang et al., 2023; Liu et al., 2024) has significantly improved feature representation in complex scenarios, yet its original architecture has several limitations. Its single-scale encoder fails to capture the multi-scale structures in fine coal flotation froth images, where small bubbles and large froth clusters coexist, potentially leading to weak responses in small-scale regions and inaccurate boundary localization in large-scale ones. Moreover, extensive use of standard convolutions introduces high computational overhead, limiting its suitability for industrial real-time monitoring. The lack of an explicit attention mechanism further hinders the network from focusing on discriminative, stable regions under strong reflections, shadows, and noise, reducing keypoint repeatability.

To address these issues, this study designs a novel feature encoder and integrates it with the SuperPoint decoder to construct a feature detection and description network, termed MFNet. As shown in Fig. 1, the overall framework comprises three components: a shared feature extraction backbone, a dual-path feature enhancement and fusion module, and keypoint detection and descriptor prediction heads. A Feature Pyramid Network (FPN) (Luo et al., 2020) fuses multi-scale semantic information to enhance scale perception in froth structures, while a Feature Pyramid + Consistency (FPC) branch locally enhances shallow features to better capture fine details such as froth films and edges. The dual-path features are fused at a unified scale, providing richer and complementary representations for keypoint detection and descriptor learning.

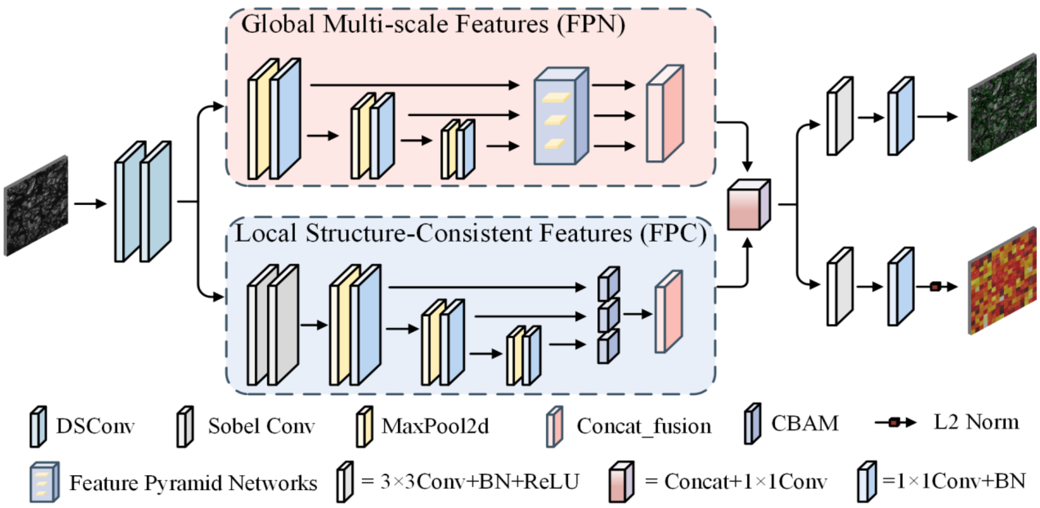


Fig. 1. MFNet architecture diagram

2.1. Multi-scale feature fusion branch (FPN)

Fine coal flotation froth images exhibit pronounced multi-scale characteristics. Large-scale regions correspond to the overall froth structure, while thin films, boundaries, and locally ruptured areas present fine-grained and scale-variant textures. Single-scale feature representations are insufficient to simultaneously capture global contours and local details, often resulting in weak keypoint responses in fine structures or uneven spatial distribution in large-scale regions.

To enhance the network's ability to perceive froth structures at different scales, a multi-scale feature fusion mechanism is introduced into the encoder. This branch is built upon the Feature Pyramid paradigm, where shallow, intermediate, and deep features are integrated to enable cross-level information propagation and joint representation learning, thereby improving the network's unified perception of both local details and global structures. Let the shallow, intermediate, and deep features be denoted as X_2, X_3 and X_4, respectively; the cross-layer feature fusion can be formulated as follows:

$$\begin{cases} P_4 = Conv_{1x1}(X_4) \\ P_3 = Conv_{1x1}(X_3) + Upsample(P_4) \\ P_2 = Conv_{1x1}(X_2) + Upsample(P_3) \\ F_{FPN} = Conv_{3x3}(Concat[P_4, Upsample(P_3), Upsample(P_2)]) \end{cases} \quad (1)$$

where *Upsample()* denotes the bilinear interpolation upsampling operation, and *Concat* represents concatenation along the channel dimension. Unlike the element-wise summation employed in the conventional Feature Pyramid Network (FPN), the concatenation-based fusion preserves the diversity of multi-scale feature representations and subsequently leverages convolutional operations for unified feature modeling, thereby enhancing the overall feature representation capability. The corresponding structure is illustrated in Fig. 2.

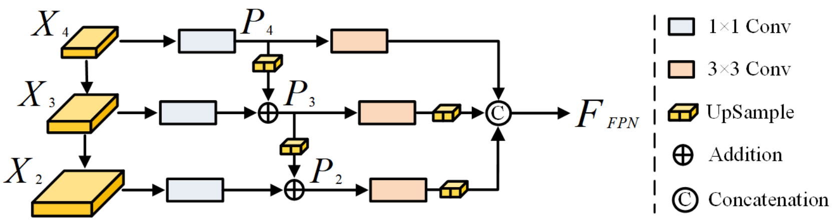


Fig. 2. FPN architecture diagram

2.2. Feature consistency enhancement branch (FPC)

Although multi-scale feature fusion improves the scale adaptability of feature representations, fine coal flotation froth images still suffer from low local texture contrast, blurred boundaries, and strong noise interference. These factors can lead to unstable keypoint responses and insufficient discriminability of descriptors. Such issues are particularly prominent in froth film regions, directly affecting the spatial distribution quality of feature points and the reliability of subsequent matching. To address these challenges, building upon multi-scale feature fusion, this study further designs a custom Feature Pyramid + Consistency (FPC) branch to enhance feature consistency. The FPC branch performs multi-scale contextual modeling on shallow features and jointly encodes them with fused FPN features during the keypoint detection stage, thereby improving the stability of feature points in low-contrast froth regions.

The FPC branch consists of an Edge Enhancement module and a Multi-scale Pyramid Fusion module arranged in series. The core design is tailored to the weak-texture and blurred-boundary characteristics of froth images:

2.2.1. Edge enhancement module

A 2D Sobel convolution operator is employed to accurately capture gradient information along froth boundaries and thin film textures, thereby strengthening locally meaningful structural features while suppressing responses in noisy and strong reflection regions. Specifically, the Sobel operator in the X-direction is used to detect vertical edges, while the Sobel operator in the Y-direction is used to detect horizontal edges. The convolution kernels are defined as follows:

- X-direction (vertical edge detection):

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

- Y-direction (horizontal edge detection):

$$K_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

In the computation process, the shallow input feature X_1 is first convolved with Sobel operators in both the X and Y directions to obtain the corresponding edge gradient features F_x and F_y. Subsequently, the two components are combined through element-wise squaring, summation, and square-root operations to obtain the complete edge-enhanced feature F_{Edge}. The formulation is given as follows:

$$F_{Edge} = \sqrt{Conv(X_1, K_x)^2 + Conv}(X_1, K_y)^2 \tag{2}$$

where *Conv()* denotes the convolution operation. This fusion strategy preserves edge details in both directions simultaneously, thereby avoiding feature loss that may occur when gradient information is extracted along a single direction.

2.2.2. Multi-scale pyramid fusion module

Different from the aforementioned FPN branch, the $F_{Pyramid}$ module is constructed solely on the shallow edge-enhanced feature F_{Edge}. It adopts a lightweight pipeline that first captures medium- and large-scale froth edge information via strided convolutional downsampling and then restores and refines details through bilinear interpolation upsampling and feature fusion. This design focuses on enhancing the multi-scale consistency of local edge features while avoiding the loss of fine details caused by global feature fusion.

Specifically, the edge-enhanced feature F_{Edge} is first downsampled using strided convolutions to obtain two-scale feature maps, denoted as $F_{Edge}{}^{2x}$ and $F_{Edge}{}^{4x}$. These downsampled features are then upsampled back to the original resolution via bilinear interpolation and fused with the base-scale feature F_{Edge} through element-wise addition, enabling complementary integration of multi-scale edge information. Finally, batch normalization followed by a ReLU activation function is applied to refine the fused representation, yielding the multi-scale context-enhanced feature $F_{Pyramid}$. The overall computation process is formulated as follows:

$$F_{Pyramid} = ReLU\left(BN\left(F_{Edge} + Upsample\left(F_{Edge}^{2x}\right) + Upsample\left(F_{Edge}^{4x}\right)\right)\right) \tag{3}$$

Let the shallow input feature be denoted as X_1. After passing through the FPC module, the enhanced feature F_{FPC} is obtained as:

$$F_{FPC} = F_{Ptramid}\left(F_{Edge}(X_1)\right) \tag{4}$$

To further enhance the discriminative capability and robustness of feature representations, a Convolutional Block Attention Module (CBAM) is introduced into the FPC branch to perform channel-wise and spatial attention weighting on the fused features:

$$F_{CBAM} = F_{FPC} \odot \sigma\left(CA(F_{FPC}) \odot \sigma(SA(F_{FPC})\right) \tag{5}$$

where *CA* and *SA* denote the channel attention and spatial attention modules, respectively, σ represents the Sigmoid activation function, and $\odot$ denotes element-wise multiplication. This mechanism adaptively enhances discriminative channels and spatial regions. The CBAM module can dynamically recalibrate feature saliency, enabling the network to more accurately extract discriminative keypoints under complex textures and noisy environments. The overall architecture is illustrated in Fig. 3.

Finally, the output of the FPC branch is fused with the output of the FPN branch as follows:

$$F_{Fuse} = Conv_{1x1}(Concat[F_{FPN}, F_{FPC}]) \tag{6}$$

By incorporating local structural enhancement and feature reweighting mechanisms, the FPC branch is able to emphasize physically meaningful regions for motion analysis, such as froth boundaries and film textures. As a result, the extracted keypoints exhibit more uniform spatial distribution and improved semantic consistency. Compared with approaches that rely solely on detection response strength, this branch enhances the representativeness and discriminative power of keypoints at the feature representation level.

The proposed FPC branch and FPN branch form a complementary relationship: the former focuses on the robust enhancement of local features, while the latter emphasizes the unified modeling of global multi-scale information. Their collaborative interaction enables the network to achieve stronger robustness in complex fine coal flotation froth scenarios, thereby providing a solid foundation for subsequent feature matching and dynamic froth feature analysis.

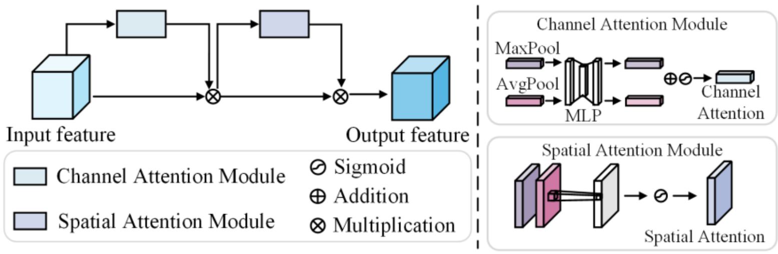


Fig. 3. CBAM attention mechanism

2.3. Spatially Adaptive Keypoint Regularization (SAKR)

Fine coal flotation froth images exhibit inherently highly non-uniform texture distributions and strong local structural similarities. Under such conditions, keypoint detection based solely on feature response strength often leads to excessive keypoint clustering in certain regions while sparse or missing detections occur in weak-texture areas (Prokopov et al., 2025). This imbalance in spatial distribution not only introduces many redundant keypoints, significantly increasing the computational burden of subsequent feature matching and motion estimation, but also degrades the global consistency of froth motion velocity field estimation due to either local information overload or the absence of effective features. Consequently, it becomes a critical bottleneck that limits the accurate analysis of dynamic froth characteristics.

To address this issue, this study proposes a Spatially Adaptive Keypoint Regularization (SAKR) method. By imposing constraints on the local spatial density of keypoints, SAKR regularizes the structural complexity of the keypoint set and promotes spatially uniform distribution, thereby laying a solid foundation for high-precision feature matching and reliable velocity estimation.

The design of SAKR is based on the local continuity of froth motion. Specifically, the motion of fine coal flotation froth exhibits continuity and smoothness within local spatial regions, where the dominant motion patterns in each area can be effectively represented by a small number of high-confidence keypoints. Based on this assumption, SAKR achieves spatial uniformity of keypoints through three core steps: equal-area grid partitioning, adaptive grid size computation, and high-confidence keypoint selection. The mathematical modeling and theoretical formulation of each step are presented as follows:

2.3.1. Equal-area grid partitioning

The froth image with a size of $W \times H$ is divided into N_{grid} non-overlapping rectangular grid cells of equal area, where each grid cell has a size of $w_g \times h_g$, satisfying:

$$N_{grid} \times w_g \times h_g = W \times H \tag{7}$$

where W and H denote the width and height of the image, respectively, w_g and h_g represent the width and height of each grid cell, and N_{grid} is the total number of grid cells, which is adaptively determined based on the image size and the local keypoint density.

2.3.2. Adaptive grid size computation

The design objective of the grid size is to ensure that each grid cell contains K high-confidence keypoints, where K is an empirically optimal value. Based on sensitivity analysis, the method achieves optimal performance when $K \in [2,4]$; in this study, K is set to 3. This setting ensures the integrity of local froth structural features while avoiding feature redundancy. Given the total number of detected global keypoints N_k, the total number of grid cells N_{grid} is adaptively computed as follows:

$$N_{grid} = \left\lfloor \frac{N_k}{K} \right\rfloor \tag{8}$$

where $\lfloor \quad \rfloor$ denotes the floor (round-down) operation, ensuring that the number of keypoints within each grid does not exceed *K*. Substituting Eq. (8) into Eq. (7), the size of each grid cell can be derived as:

$$w_g = \left\lfloor \frac{W}{\sqrt{N_{grid}}} \right\rfloor, \ \ h_g = \left\lfloor \frac{H}{\sqrt{N_{grid}}} \right\rfloor \tag{9}$$

Through this adaptive formulation, SAKR can dynamically adjust the grid size according to the distribution characteristics of feature points in different froth images, thereby accommodating the regularization requirements of feature points across varying froth morphologies. This overcomes the limitations of fixed-grid approaches in diverse application scenarios.

2.3.3. High-confidence keypoint selection

For the set of keypoints within each grid cell $\{P_1 | i = 1,2, \dots, n_g\}$ (where n_g denotes the number of keypoints in the g-th grid), keypoints are sorted and selected based on their detection confidence scores.

The confidence score of MFNet keypoints essentially quantifies the probability that a given pixel corresponds to a froth feature point. The raw response value $R(u,v)$ is determined by the feature representation at the corresponding location in the fused feature map and its semantic similarity to froth keypoints. A higher similarity leads to a larger $R(u,v)$, and after normalization and non-maximum suppression (NMS), the resulting confidence score c_i is obtained. The mathematical formulation of this process can be expressed as follows:

$$\begin{cases} R = Conv_{1x1}(F_{Fuse}) \\ C_{init}(u,v) = \sigma(R(u,v)) \\ c_i = NMS_{3x3}(C_{init}(u_i,v_i)) \end{cases} \tag{10}$$

NMS_{3x3}: Non-maximum suppression within a 3×3 neighborhood, ensuring local uniqueness; (u_i,v_i): The coordinates of the i-th candidate keypoint; c_i: The keypoint confidence score output by the MFNet detection head, where a higher value indicates greater reliability.

Subsequently, the keypoints within each grid are sorted in descending order according to their confidence scores c_i, and only the top-*K* keypoints are retained. The selection rule is defined as follows:

$$P_g^* = top_k(\{P_i | c_i \in [0,1], i = 1,2, \dots, n_g\}) \tag{11}$$

where P_g^* denotes the set of selected keypoints in the g-th grid. After applying the above selection procedure to all grid cells, the globally regularized keypoint set is obtained as $P^* = U_{g=1}^{N_{grid}} P_g^*$, achieving spatial uniformization of feature points across the entire image.

It should be noted that the proposed SAKR is designed to regularize the spatial distribution of keypoints within each individual frame rather than explicitly enforcing temporal consistency across consecutive frames. By encouraging a more uniform distribution of reliable keypoints, SAKR provides stable geometric support for subsequent feature matching. Temporal correspondences between consecutive frames are established by LightGlue, which performs descriptor-based feature matching under geometric consistency constraints. Therefore, keypoints located within the same grid cell at different time instants are not assumed to correspond directly. Instead, their correspondences are determined adaptively according to descriptor similarity and matching confidence.

Furthermore, no explicit temporal filtering technique (e.g., Kalman filtering) is employed in the proposed framework. Instead, the stability of the estimated motion field relies on the combination of the spatially regularized keypoint distribution provided by SAKR and the robust feature matching capability of LightGlue. Together, these components effectively reduce mismatches and suppress abnormal fluctuations in local motion vectors, enabling stable and continuous froth motion estimation without introducing additional temporal constraints.

3. Experiments

3.1. Datasets

To enhance the generalization capability and robustness of the feature point extraction model in complex scenarios, both real-world coal flotation froth datasets and three types of public datasets are employed for training and evaluation. A multi-stage, multi-source collaborative training and evaluation strategy is adopted. Model parameter learning is primarily conducted on publicly available datasets with rich structural diversity and geometric variations, including synthetic datasets and the COCO dataset (Lin et al., 2014), enabling the model to effectively learn multi-scale structural representations and complex texture patterns.

During the evaluation stage, a coal flotation froth dataset is introduced as a representative industrial vision scenario to assess the model's keypoint detection stability, matching reliability, and dynamic feature perception under real flotation conditions. In addition, the HPatches dataset (Balntas et al., 2017) is utilized to further evaluate the model's robustness under cross-viewpoint changes, illumination variations, and geometric transformations.

3.1.1. Coal slime flotation froth dataset

The coal flotation froth dataset was collected from the No. 1 and No. 2 flotation cells of a coal preparation plant in Inner Mongolia Autonomous Region. The acquisition system consists of an industrial camera, a storage device, and auxiliary lighting. During data collection, the on-site environment was shielded from ambient light to minimize illumination interference. The experimental setup is illustrated in Fig. 4. The captured data has a resolution of 3312×2496 pixels. The video frame rate is 16.39 FPS. During model training and testing, frames are extracted from the recorded videos to construct an image dataset covering various flotation conditions and froth morphologies. The dataset includes typical scenarios such as sparse small bubbles and dense froth clusters, ensuring the diversity and representativeness of the training data.

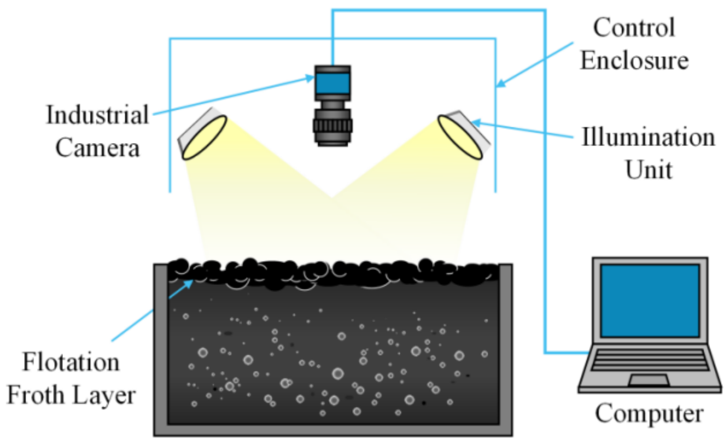


Fig. 4 Coal flotation froth data acquisition setup diagram

3.1.2. Public datasets

To comprehensively evaluate the model's robustness, generalization capability, and feature matching performance, three types of public datasets were used in the experiments:

The Synthetic Shapes dataset is a commonly used benchmark in the field of feature point detection. It contains various simple geometric shapes and texture patterns, and generates image pairs by applying transformations such as scaling, rotation, and blurring. This dataset is suitable for evaluating a model's ability to handle basic geometric transformations and serves as a controlled environment for assessing fundamental feature detection performance.

The COCO dataset is a large-scale, widely used dataset for object detection and segmentation. It includes a diverse range of natural scene images covering multiple object categories such as people, vehicles, and everyday items. The images exhibit complex backgrounds and significant scale variations, making it suitable for evaluating a model's generalization ability in challenging real-world scenarios.

The HPatches dataset is specifically designed for image feature matching tasks. It consists of image sequences captured under real-world conditions and is divided into two subsets: sequences with homography transformations and sequences with real viewpoint and illumination variations. This dataset provides a comprehensive benchmark for evaluating the robustness of feature matching algorithms under geometric distortions and environmental changes.

3.2. Experimental environment

All experiments in this study are conducted on a unified hardware platform. The experimental environment consists of an Intel® Core™ i7-14700KF 20-core processor, 32 GB DDR5 4800 MHz RAM, and an NVIDIA GeForce RTX 5070 GPU with 12 GB memory, running on a 64-bit Microsoft Windows 10 operating system. The development environment is PyCharm 2024, with the primary software stack including Python 3.10 and PyTorch 2.8.0+cu128.

The proposed MFNet and deep learning-based comparison methods (SuperPoint, D2-Net, and LoFTR) are implemented in PyTorch, while traditional feature-based methods (SIFT, ORB, and AKAZE) are implemented using OpenCV.

To ensure fairness and reproducibility in efficiency evaluation, all models adopt identical inference settings: a batch size of 1 to align with the real-time single-frame inference scenario in industrial deployment, and a unified inference precision of FP32.

3.3. Evaluation metrics

In the SuperPoint training and public dataset evaluation stage, the proposed MFNet is comprehensively evaluated using key metrics including Repeatability (Rep), Localization Error (MLE), Nearest Neighbor mean Average Precision (M.AP), Matching Score (M.score), and Homography Estimation Accuracy. These metrics assess the model from multiple perspectives, including keypoint detection stability, localization accuracy, descriptor discriminability, and geometric consistency.

In the coal flotation froth feature extraction and matching stage, owing to the deformable and dynamic nature of froth images and the absence of ground-truth velocity measurements in industrial flotation systems, quantitative validation of the estimated velocity field against reference values is not feasible. Therefore, the evaluation focuses on the number of detected keypoints, high-confidence keypoints, matched points, Matching Confidence, and the entropy of the estimated velocity and motion direction. Together, these metrics provide an indirect assessment of feature extraction robustness, feature matching reliability, and the stability and consistency of the estimated froth motion field.

For efficiency evaluation, the following metrics are adopted: the number of parameters (Params), computational complexity (FLOPs), inference latency (Latency), and GPU memory usage (GPU Memory), where lower values are preferred. In contrast, inference speed (FPS) is expected to be higher, reflecting better real-time performance.

3.4. Comparison of keypoint extraction methods

3.4.1. Keypoint detection comparison

To comprehensively evaluate the performance of the proposed MFNet in keypoint detection and description, quantitative comparison experiments are first conducted on the public HPatches dataset. The comparison is performed from two aspects: homography estimation accuracy and the quality of feature detection and descriptors. The proposed method is compared with traditional methods (SIFT, ORB, and AKAZE), deep learning-based methods (SuperPoint, LoFTR, and D2-Net), as well as the method proposed by Zhou et al. (2024). The experimental results are reported in Table 1 and Table 2.

3.4.1.1. Quantitative evaluation on public datasets

Table 1 presents a quantitative comparison of different methods on the HPatches dataset from two aspects: keypoint detection and descriptor performance. Specifically, Rep and MLE are used to evaluate the repeatability and localization accuracy of keypoints, respectively, while M.AP and M.score are adopted to measure the discriminative capability of descriptors in feature matching. From the detector perspective, the method of Zhou et al. achieves the highest repeatability, with MFNet performing comparably and ranking closely behind. From the descriptor perspective, MFNet attains the best performance in both mean Average Precision (M.AP) and overall matching score (M.score), followed by SuperPoint. In contrast, traditional methods such as SIFT and ORB exhibit inferior performance across all metrics compared to these learning-based approaches, while D2-Net shows the weakest overall results. Overall, the proposed MFNet achieves a favorable balance between detection stability and descriptor discriminability, demonstrating its effectiveness and superiority in feature extraction tasks.

Table 2 presents the homography estimation performance of different methods on the HPatches dataset, evaluated under three error thresholds (ε = 1, 3, and 5). MFNet achieves the best or second-best results across all three metrics. Notably, under the strict constraint of ε = 1, MFNet attains an accuracy of 0.433, significantly outperforming other methods, demonstrating its stable and superior matching performance.

Table 1. Comparison of keypoint detection and descriptor performance

Methods	Detector metrics		Descriptor metrics	
	Rep↑	MLE↓	M.AP↑	M.score↑
SIFT	0.512	1.163	0.747	0.264
ORB	0.150	1.157	0.395	0.438
AKAZE	0.284	1.102	0.598	0.417
Zhou et al	0.612	1.082	0.802	0.491
LoFTR	0.601	2.110	0.672	0.399
D2-Net	0.245	3.530	0.245	0.127
SuperPoint	0.581	1.158	0.821	0.470
MFNet	0.605	1.074	0.838	0.499

Table 2. Comparison of homography estimation performance

Methods	Homography estimation↑		
	ε=1	ε=3	ε=5
SIFT	0.424	0.676	0.759
ORB	0.423	0.546	0.617
AKAZE	0.357	0.476	0.676
Zhou et al	0.353	0.728	0.832
LoFTR	0.399	0.672	0.763
D2-Net	0.077	0.245	0.360
SuperPoint	0.310	0.684	0.829
MFNet	0.433	0.736	0.811

3.4.1.2. Quantitative analysis on the coal flotation froth dataset

On the fine coal flotation froth image dataset, a further statistical analysis of feature extraction and matching performance for different methods is conducted, as summarized in Table 3. To ensure a fair comparison, the number of extracted keypoints for all methods is uniformly limited to 1024 in the experiments. Specifically, Keypoints is used to verify the consistency in the number of detected feature points, while Repeatability characterizes the ability to repeatedly detect keypoints under dynamic variations of the froth. Matched Points reflects the number of valid feature correspondences across frames. In addition, Entropy is employed to evaluate the uniformity of the distribution of froth motion in terms of velocity and direction; lower entropy values indicate a more uniform feature distribution and more stable dynamic representation.

As shown in Table 3, all methods generally satisfy the constraint on the number of extracted keypoints, ensuring a fair comparison. In terms of Repeatability, MFNet achieves the best performance with 58.4%, slightly outperforming SuperPoint and AKAZE, and significantly surpassing other methods such as ORB and LoFTR. This indicates that MFNet can stably and repeatedly detect feature points under transient froth variations and in weak-texture, high-interference environments, effectively addressing the issue of feature loss in existing approaches. Regarding Matched Points, LoFTR attains the highest value due to its dense matching strategy; however, when considered together with its Repeatability, its matching stability is relatively limited. Although MFNet yields fewer matched points than LoFTR, it significantly outperforms SIFT, ORB, and AKAZE, and far exceeds D2-Net, demonstrating its strong effectiveness in feature matching. For the Entropy metric, MFNet achieves the lowest values in both velocity and direction dimensions, indicating that the extracted motion features of froth—both in terms of speed and direction—are more uniformly distributed, leading to more stable dynamic representations and enabling more accurate characterization of froth motion patterns.

Table 3. Comparison of froth feature analysis

Methods	Keypoints↑	Repeatability↑	Matched points↑	Entropy↓	
				Velocity	Direction
SuperPoint	1024	57.3%	782	3.532	3.484
SIFT	1020	37.4%	488	3.637	3.956
ORB	1024	26.7%	334	3.544	3.688
AKAZE	1024	56.2%	569	3.637	3.956
LoFTR	1005	38.4%	1005	3.488	3.740
D2-Net	1024	40.9%	583	3.554	3.775
MFNet	1024	58.4%	693	3.441	3.686

Overall, MFNet exhibits superior performance across all evaluation metrics on the real-world fine coal flotation froth dataset. It not only ensures the stability of feature detection and the effectiveness of feature matching but also enables precise representation of froth dynamic characteristics. These results further validate the adaptability of the proposed method to the specific characteristics of fine coal flotation scenarios, demonstrating clear advantages over both traditional and existing deep learning-based methods, and better meeting the practical requirements of dynamic froth analysis in industrial environments.

3.4.1.3. Visual comparison of keypoint detection results

Based on the quantitative analysis, Fig. 5 presents the visualized keypoint detection results of different methods on coal flotation froth images, including traditional methods (SIFT, ORB, AKAZE), SuperPoint, LoFTR, D2-Net, and the proposed MFNet method.As observed from the figure, traditional methods are capable of detecting a large number of feature points; however, they generally suffer from pronounced local clustering during the extraction process. SuperPoint exhibits limited responsiveness in weak-texture and low-illumination regions, and its feature point distribution still shows a certain degree of non-uniformity. In contrast, LoFTR and D2-Net tend to exhibit missing or sparsely distributed feature points on fine coal flotation froth images, making it difficult to fully cover froth structures and boundary regions. Moreover, the distribution of extracted feature points varies across different frames, indicating relatively poor robustness.

By comparison, MFNet presents a more dispersed and uniform distribution of feature points, enabling more comprehensive coverage of froth structural boundaries and critical regions. Combined with the aforementioned quantitative results, it can be concluded that although some traditional methods have advantages in terms of the number of detected feature points, they often suffer from feature redundancy and limited discriminative capability. In contrast, MFNet achieves a better balance among the number of feature points, spatial distribution uniformity, and matching stability.

3.4.2. Comparison of feature matching before and after SAKR

In coal flotation froth images, due to the highly non-uniform spatial distribution of froth structures, directly detecting keypoints based on feature response intensity often leads to dense clustering in regions with rich local textures, while obvious gaps appear in thin-film rupture areas or low-contrast regions. This imbalance in spatial distribution not only introduces many redundant keypoints, increasing the computational cost of subsequent feature matching and motion estimation, but also reduces matching accuracy and motion field estimation precision due to local information overload or absence.

To address this issue, the SAKR method is introduced in the feature detection stage. This approach first divides the input image into regular grid cells and then performs adaptive selection within each cell based on keypoint confidence and spatial distribution constraints, ensuring that only the most representative keypoints are retained in each local region. Experimental results show that, while ensuring sufficient coverage of key structural regions, SAKR effectively reduces redundant keypoints in densely populated areas, resulting in a more uniform and reasonable overall distribution. Quantitatively, after applying SAKR, the total number of keypoints is moderately reduced while preserving information completeness, and the proportion of high-confidence matched points is significantly improved. Compared with the non-regularized model, the feature point accuracy improves by 4.7%, indicating that SAKR does not weaken feature representation but instead enhances the representativeness of the overall feature set by suppressing redundant points.

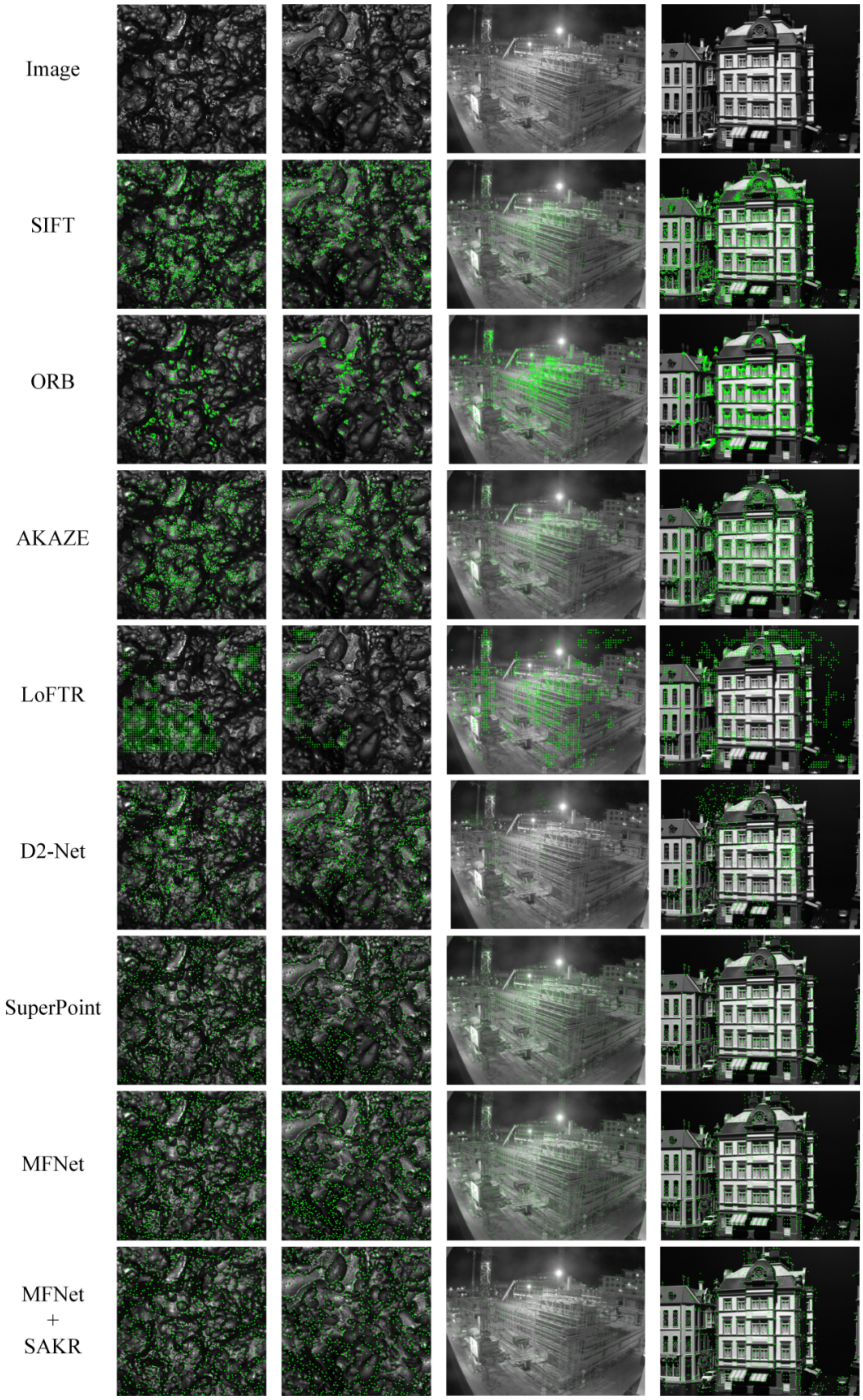
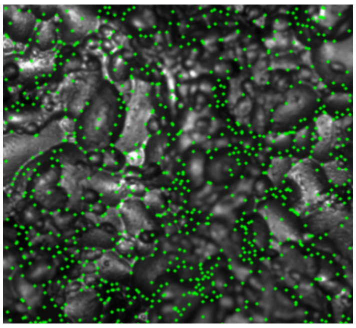


Fig. 5. Comparison of keypoint detection results extracted by different methods

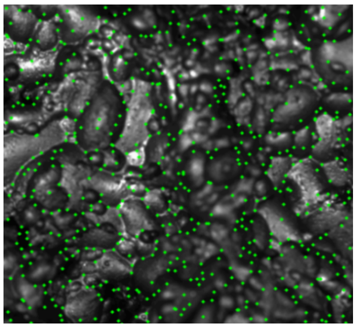
The visualization of keypoint distribution is shown in Fig. 6. Fig. 6(a) illustrates the keypoints detected by MFNet, where some regions contain excessive points while others lack detections, resulting in an uneven distribution. In contrast, Fig. 6(b) shows the keypoints after regularization, where the points are more evenly distributed across the entire image, significantly alleviating the issue of local clustering.

The proposed MFNet model, when combined with SAKR, achieves significant performance improvements in froth dynamic analysis. As shown in Table 4, the MFNet+SAKR configuration increases the matching confidence from 88.3% in the baseline model to 93.4%, while the high-confidence matching rate improves from 81.9% to 86.6%. This improvement is of critical importance for subsequent froth motion velocity computation, as the increase in the proportion of high-confidence matches directly enhances the reliability of feature trajectory tracking, thereby improving the accuracy and stability of velocity field estimation.

The core contribution of this work lies in the synergistic design of a multi-scale feature pyramid and a distribution optimization strategy. This design reduces redundant feature point computations by 22% while significantly improving matching quality, providing a more reliable quantitative basis for froth dynamics analysis.



(a) Keypoint distribution before SAKR



(b) Keypoint distribution after SAKR

Fig. 6. Comparison of Results Before and After SAKR

Table 4. Comparison of Froth Dynamic Analysis Before and After SAKR

Experimental setting	Keypoints	Matches	High-confidence matches	Match confidence	High-confidence Ratio
SP+LG	102,224	75622	65614	88.3%	86.8%
SP+SAKR+LG	86351	63641	55738	88.4%	87.6%
MFNet+LG	103424	65274	53454	91.7%	81.9%
MFNet+SAKR+LG	80628	48530	42020	93.4%	86.6%

3.4.3. Efficiency evaluation and comparison of different models

To evaluate the computational efficiency differences among various feature extraction and matching methods, this study compares SIFT, ORB, AKAZE, SuperPoint, D2-Net, LoFTR, and the proposed MFNet across multiple dimensions, including model parameters, computational complexity, frame rate, inference latency, and GPU memory usage. The results are summarized in Table 5.

Table 5 compares the computational efficiency of the proposed MFNet with several mainstream feature matching methods. All deep learning-based approaches are evaluated under identical experimental conditions. The results show that SuperPoint achieves the highest inference speed of 223.71 FPS with only 1.31M parameters and 26.05 GFLOPs, demonstrating clear advantages in lightweight real-time application scenarios. In contrast, although LoFTR has the largest number of parameters and the highest computational cost, it is primarily designed to maximize matching accuracy, at the expense of runtime efficiency. D2-Net lies between the two, achieving 43.80 FPS with a parameter scale comparable to MFNet.

Table 5 Efficiency Evaluation Results

Method	Params(M)↓	FLOPs(G)↓	FPS↑	Latency(ms)↓	GPU memory (MB)↓
SIFT	—	4.72	1.87	533.4	—
ORB	—	0.94	5.52	181.29	—
AKAZE	—	3.77	2.63	379.72	—
SuperPoint	1.31	26.05	223.71	4.47	313.15
D2-Net	7.64	168.76	43.80	22.83	532.41
LoFTR	11.56	348.95	16.99	58.84	1079.01
MFNet	7.79	114.26	75.19	13.30	296.34

Note: The symbol "—" indicates that SIFT, ORB, and AKAZE are traditional hand-crafted feature extraction algorithms without trainable network parameters

The proposed MFNet provides a balanced solution between lightweight design and high accuracy. It achieves an inference speed of 75.19 FPS with a latency of only 13.30 ms, meeting the requirements of most real-time applications. In terms of computational complexity, MFNet requires only 114.26 GFLOPs, representing a reduction of 32.3% compared to D2-Net and 67.3% compared to LoFTR. More importantly, MFNet demonstrates excellent memory efficiency, consuming only 296.34 MB of GPU memory—44.3% less than D2-Net and 72.5% less than LoFTR and even slightly outperforming the lightweight SuperPoint in this regard.

These efficiency advantages make MFNet particularly suitable for deployment in resource-constrained environments such as embedded systems and mobile robots, providing a practical solution that effectively balances accuracy and efficiency for real-world applications.

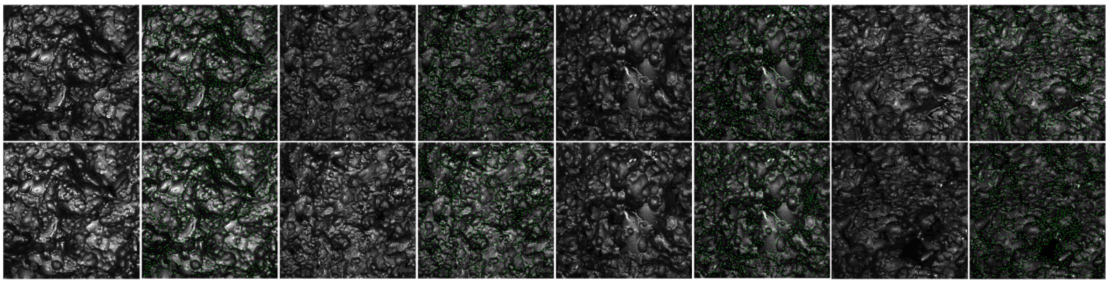
3.4.4 Performance evaluation under different froth conditions

Different flotation operating conditions produce distinct froth morphologies, and these morphological variations directly affect image texture distribution, local structural features, as well as the performance of keypoint detection and feature matching (Polat et al., 2003; Wen and Xia, 2017). In industrial flotation, variations in operating conditions and slurry characteristics, particularly the presence of fine slime particles, can alter froth stability through changes in bubble coalescence and liquid-film drainage, resulting in the evolution from fine, stable froth to coarse and fragile froth (Ma et al., 2021; Ge et al., 2025; Zhao and Zhang, 2025; Huang et al., 2026). Therefore, to evaluate the adaptability of the proposed method under different flotation states, representative froth images captured at different stages of the same flotation process were selected for analysis. Fig. 7 illustrates the typical evolution of the froth from a fine and stable state to a coarse and fragile state. The images are arranged chronologically according to the flotation process, where (a) represents Fine stable froth, (b) Transitional froth, (c) Coarse froth, and (d) Fragile froth. These four representative froth conditions cover the typical morphological evolution observed during industrial flotation and provide progressively decreasing texture richness for evaluating the robustness of the proposed method.

Different flotation operating conditions produce distinct froth morphologies, and these morphological variations directly affect image texture distribution, local structural features, as well as the performance of keypoint detection and feature matching (Polat et al., 2003; Wen and Xia, 2017). Therefore, to evaluate the adaptability of the proposed method under different flotation states, representative froth images captured at different stages of the same flotation process were selected for analysis. Fig. 7 illustrates the typical evolution of the froth from a fine and stable state to a coarse and fragile state. The images are arranged chronologically according to the flotation process, where (a) represents Fine Stable Froth, (b) Transitional Froth, (c) Coarse Froth, and (d) Fragile Froth. These four representative froth conditions cover the typical morphological evolution observed during industrial flotation and provide progressively decreasing texture richness for evaluating the robustness of the proposed method.

(a) Fine stable froth

This stage generally corresponds to stable flotation conditions. The bubbles are small and uniformly distributed, with intact liquid films and continuous boundaries. Rich texture details and pronounced grayscale variations provide abundant stable local structures for keypoint detection. Consequently, a large number of uniformly distributed keypoints can be detected, providing a reliable basis for subsequent feature matching and motion estimation.



(a) Fine stable froth (b) Transitional froth (c) Coarse froth (d) Fragile froth

Fig. 7. Representative flotation froth images under different froth conditions. (a) fine stable froth; (b) transitional froth; (c) coarse froth; (d) fragile froth

(b) Transitional froth

As flotation progresses, small bubbles gradually coalesce into medium-sized bubbles, increasing the bubble size distribution. Smooth liquid-film regions begin to appear, and the density of bubble boundaries decreases. Although local texture information is slightly reduced, sufficient structural information remains, allowing the proposed method to maintain a high number of Keypoints, as well as good Repeatability, Matched Points, and Matching Confidence.

(c) Coarse froth

With further bubble coalescence, large bubbles become dominant, resulting in larger continuous liquid-film regions and reduced texture and edge information, indicating a decline in froth stability. Nevertheless, MFNet effectively extracts discriminative local features through multi-scale feature fusion and shallow feature enhancement, leading to only a slight decrease in Repeatability, Matched Points, and Matching Confidence.

(d) Fragile froth

This stage is generally associated with unstable flotation conditions. The froth becomes increasingly coarse, accompanied by local bubble rupture and liquid-film thinning. As texture information decreases and bubble boundaries become discontinuous, the number of stable local features is significantly reduced, making feature detection and matching more challenging. Consequently, Keypoints, Repeatability, Matched Points, and Matching Confidence all decrease to some extent, making this the most challenging froth condition for evaluating the robustness of the proposed method.

The experimental results under different froth conditions are summarized in Table 6. The proposed method consistently detects many keypoints while maintaining high Repeatability, Matched Points, and Matching Confidence across different froth morphologies. As the froth evolves from a fine and stable state to a coarse and fragile state, the number of detected keypoints, matched points, and matching confidence gradually decrease owing to reduced texture richness and increased smooth regions. Nevertheless, the proposed method still achieves reliable feature extraction and matching performance, demonstrating the effectiveness of the proposed multi-scale feature fusion and shallow feature enhancement mechanisms under varying froth textures. Meanwhile, the velocity and direction entropy values vary with the froth condition, reflecting differences in froth motion characteristics under different operating states rather than a monotonic degradation in image quality, indicating that the proposed method can effectively characterize motion patterns across diverse froth morphologies.

Since ground-truth velocity measurements are unavailable in industrial flotation systems, quantitative validation of the estimated motion field against reference values is not feasible. Therefore, the reliability of feature matching is evaluated using Matching Confidence rather than the ground-truth-based M.Score. Furthermore, reagent dosage and other process variables are difficult to measure and synchronize accurately with image acquisition in industrial flotation systems. Therefore, this study adopts froth morphology as the evaluation criterion, providing a practical assessment of the generalization capability of the proposed method under real operating conditions. Future work will integrate synchronized process measurements, such as reagent dosage and air flow rate, with visual information to establish a quantitative relationship between operating parameters and feature extraction performance.

Table 6. Performance of the proposed method under different froth conditions

Froth Condition	Keypoints↑	Repeatability↑	Matched points↑	Matching confidence↑	Entropy↓	
					Velocity	Direction
Fine Stable Froth	1024	59.2%	685	0.925	3.068	1.647
Transitional Froth	1024	55.8%	643	0.885	3.153	3.546
Coarse Froth	1024	55.1%	585	0.874	2.566	2.251
Fragile Froth	938	51.4%	538	0.708	3.624	1.063

3.5. Feature matching and dynamic feature analysis

After completing keypoint extraction and spatial regularization, the optimized keypoints and their descriptors are fed into the LightGlue(Lindenberger et al., 2023) network for feature matching. Compared with conventional local similarity-based matching methods, LightGlue performs adaptive inference of correspondences through global context modeling, enabling more stable and consistent matching results in foam scenes with complex textures and significant illumination variations.

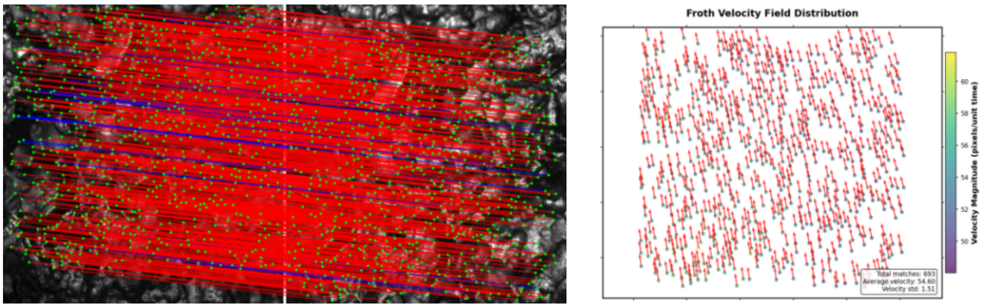
Under the proposed MFNet and the SAKR keypoint regularization strategy, the keypoints entering the matching stage exhibit significantly improved spatial distribution and semantic consistency. This effectively reduces high similarity among redundant keypoints and directs the matching process toward foam regions with genuine physical significance, thereby enhancing overall matching quality and reliability.

Fig. 8 presents the visualization of feature matching on the HPatches dataset. It can be observed that, even under complex illumination changes and viewpoint perturbations, the proposed method maintains relatively dense and accurate correspondences. The matched points are more evenly distributed, and the number of mismatches is significantly reduced, demonstrating the effectiveness of the proposed feature extraction network and the SAKR strategy in general scenarios.

Fig. 8. Matching results on the Hpatches dataset

Furthermore, the proposed method is applied to real-world fine coal flotation foam video sequences. Feature matching is performed between consecutive frames, and the resulting correspondences are used to compute the local motion velocity field of the foam. Fig. 9 illustrates the feature matching results between the K-th and (K+1)-th frames, along with the corresponding velocity field. It can be observed that the correspondences are clear and stable, and the velocity vectors exhibit good directional consistency.

Fig. 10 illustrates the motion velocity trends obtained from the continuous analysis of the mid-layer froth in the flotation cell. The results show that, across a sequence of 100 consecutive froth image pairs, the feature points matched by the proposed method maintain strong directional consistency with the overall froth motion. The estimated velocity vector field remains continuous and spatially coherent, effectively reducing local velocity discontinuities and abnormal fluctuations caused by feature mismatches. Owing to the absence of ground-truth velocity measurements in industrial flotation systems, the quality of the estimated motion field is assessed indirectly through feature matching consistency and the statistical properties of the velocity distribution. These observations indicate that the proposed method can provide robust and reliable relative motion estimation under real industrial conditions, thereby offering an effective basis for dynamic froth feature analysis and flotation condition monitoring.



(a) Matching results (b) Velocity feature field

Fig. 9. Matching results of adjacent froth frames

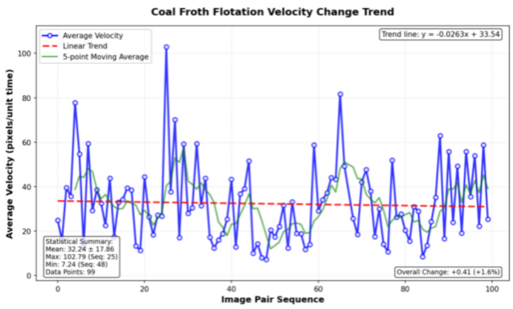


Fig. 10. Trend of froth motion velocity variation

The above results indicate that, through the joint design of "feature enhancement—keypoint regularization—high-quality matching," the proposed method provides a more stable and reliable basis for motion estimation in foam dynamic feature analysis, thereby laying a solid data foundation for subsequent modeling of foam flow behavior and process condition analysis.

3.6. Ablation study

To evaluate the effectiveness and generalization capability of the core modules in the network for general feature extraction tasks, a systematic ablation study was conducted on the widely adopted HPatches benchmark dataset in the computer vision community. Utilizing this publicly available dataset allows for the exclusion of complex disturbances specific to industrial scenarios, thereby enabling a purer assessment of the performance gains brought by structural improvements in the network, such as feature pyramid fusion and shallow-layer enhancement mechanisms. Additionally, it facilitates fair comparison with existing representative methods under a recognized evaluation framework, ensuring objectivity and reproducibility of the experimental results, and laying a foundation for subsequent evaluation of the method's scene adaptability on dedicated fine coal flotation foam datasets.

3.6.1 Single-module ablation

To quantitatively assess the contribution of each functional module to the overall network performance, single-module ablation experiments were conducted on the MFNet network. Each module was individually introduced while keeping the output resolution, number of channels, and loss function unchanged. The experimental results are summarized in Tables 7 and 8.

Table 7. Results of single-module ablation

Model	FPN	FPC	DSC	Edge	CBAM	Rep↑	MLE↓	M.AP↑	M.score↑
Baseline	×	×	×	×	×	0.581	1.158	0.821	0.470
a	×	×	√	×	×	0.595	1.167	0.830	0.485
b	×	×	×	√	×	0.713	1.968	0.398	0.278
c	×	×	×	×	√	0.616	1.054	0.824	0.498
d	√	×	×	×	×	0.598	1.107	0.841	0.501
e	×	√	×	×	×	0.604	1.093	0.828	0.478

Table 8. Results of single-module ablation

Model	FPN	FPC	DSC	Edge	CBAM	Homography Estimation↑		
						ε=1	ε=3	ε=5
Baseline	×	×	×	×	×	0.310	0.684	0.829
a	×	×	√	×	×	0.385	0.709	0.785
b	×	×	×	√	×	0.466	0.469	0.495
c	×	×	×	×	√	0.467	0.739	0.793
d	√	×	×	×	×	0.405	0.734	0.810
e	×	√	×	×	×	0.417	0.739	0.795

From the perspective of keypoint detection and matching performance, each module exhibits varying degrees of improvement across different evaluation metrics, reflecting their functional focus and complementarity. Specifically, the introduction of the DSC module increases the keypoint repeatability from 0.581 to 0.595, the mean matching accuracy (MMA) from 0.821 to 0.830, and the matching score to 0.485, indicating that this module enhances the discriminative capability of descriptors. The CBAM module reduces the localization error from 1.158 to 1.054 and increases the matching score from 0.470 to 0.498. Meanwhile, the homography estimation accuracy under ε=1and ε=3 improves to 0.467 and 0.739, respectively, demonstrating the effectiveness of attention mechanisms in improving feature selection quality and geometric consistency. The FPN module, through multi-scale feature fusion, increases the MMA and matching score to 0.841 and 0.501, and boosts the homography estimation accuracy under ε=1 and ε=3 to 0.405 and 0.734, indicating the positive role of multi-scale information in enhancing matching robustness. The FPC module achieves relatively balanced improvements across all metrics, with localization error reduced to 1.093, keypoint repeatability increased to 0.604, and homography estimation accuracy under ε=1 and ε=3 reaching 0.417 and 0.739, highlighting its advantage in keypoint localization precision and maintaining geometric consistency. Overall, the homography estimation results in Table 6 further indicate that the complete MFNet network achieves balanced and stable performance across different error thresholds, demonstrating that the collaborative multi-module design can simultaneously enhance keypoint detection and matching performance across multiple evaluation dimensions. These experimental findings provide strong evidence of the proposed method's effectiveness in geometric consistency modeling and overall robustness.

3.6.2 Multi-module ablation

To systematically analyze the impact of each module on model performance, single-module ablation experiments were first conducted to evaluate the individual contributions of FPN, FPC, DSC, Edge, and CBAM without introducing other enhancement modules. On this basis, multi-module combination ablation experiments were further carried out (Tables 9 and 10) to investigate the synergistic relationships among different modules.

The multi-module ablation results indicate that the core structure composed of FPN and FPC provides a fundamental basis for multi-scale feature modeling and feature calibration. However, when relying solely on this core structure, the model still exhibits noticeable deficiencies in matching accuracy and geometric consistency. With the introduction of DSC, the localization error is significantly reduced from 1.363 to 1.184, while the mean Average Precision recovers to 0.792, demonstrating that DSC plays a crucial role in alleviating feature localization bias caused by multi-scale fusion. Furthermore, incorporating the CBAM attention mechanism improves the matching mAP and Matching Score to 0.854 and 0.553, respectively, and also enhances homography estimation performance. This indicates that the attention mechanism can effectively strengthen feature discriminability and matching robustness. However, introducing the Edge module alone or in an improper manner may degrade matching performance or stability to some extent. This suggests that edge constraints can only play a positive role within a multi-module collaborative framework, where they help suppress geometric distortions caused by excessive attention focus. Overall, the complete MFNet achieves the best balance between keypoint localization accuracy and overall matching performance, while maintaining stable performance under different homography error thresholds. These results fully validate the rationality, effectiveness, and necessity of the proposed MFNet architecture for feature extraction and matching tasks.

Table 9. Results of multi-module ablation

Model	FPN	FPC	DSC	Edge	CBAM	Rep↑	MLE↓	M.AP↑	M.score↑
Baseline	×	×	×	×	×	0.581	1.158	0.821	0.470
a	√	√	×	×	×	0.589	1.363	0.538	0.461
b	√	√	√	×	×	0.597	1.184	0.792	0.472
c	√	√	×	√	×	0.585	1.245	0.635	0.443
d	√	√	×	×	√	0.696	1.669	0.629	0.428
e	√	√	√	√	×	0.563	1.038	0.648	0.338
f	√	√	√	×	√	0.551	1.212	0.854	0.553
g	√	√	×	√	√	0.632	1.407	0.715	0.481
MFNet	√	√	√	√	√	0.605	1.074	0.838	0.499

Table 10. Results of multi-module ablation

Model	FPN	FPC	DSC	Edge	CBAM	Homography Estimation↑		
						ε=1	ε=3	ε=5
Baseline	×	×	×	×	×	0.310	0.684	0.829
a	√	√	×	×	×	0.441	0.485	0.586
b	√	√	√	×	×	0.398	0.711	0.803
c	√	√	×	√	×	0.412	0.598	0.705
d	√	√	×	×	√	0.416	0.513	0.687
e	√	√	√	√	×	0.426	0.689	0.795
f	√	√	√	×	√	0.436	0.789	0.893
g	√	√	×	√	√	0.428	0.652	0.768
MFNet	√	√	√	√	√	0.433	0.736	0.811

3.7. Discussion on robustness under harsh industrial conditions

With the continuous development of intelligent monitoring and control technologies for flotation processes, vision-based online monitoring methods have been increasingly integrated into industrial flotation equipment (Zhang et al., 2026).

However, the coal slime flotation process is conducted in a typical complex industrial visual environment, where steam, water mist, coal dust, and camera lens contamination are commonly encountered during actual production. These factors may degrade image quality, resulting in reduced image contrast, local blurring, weakened edge information, and uneven illumination, which can consequently affect the performance of keypoint detection, feature matching, and subsequent motion analysis.

In this study, representative froth images corresponding to different froth states, including fine stable froth, transitional froth, coarse froth, and fragile froth, were analyzed to evaluate the adaptability of the proposed method under varying texture distributions and froth morphologies. The experimental results demonstrate that the proposed MFNet combined with the SAKR strategy can maintain relatively stable feature detection and matching performance across different froth conditions, indicating a certain degree of robustness to texture variations commonly encountered in industrial flotation processes. Nevertheless, it should be noted that the present work did not include dedicated experiments under explicitly controlled harsh environmental conditions, such as dense steam occlusion, severe coal dust deposition, or significant camera lens contamination. Therefore, the discussion on robustness under such disturbances is mainly based on the characteristics of the proposed method and the observations obtained from different froth states, rather than direct experimental verification under these specific conditions.

Future work will focus on establishing controlled experimental scenarios involving steam, water mist, coal dust, lens contamination, and illumination variations, and on integrating image enhancement, image dehazing, lens contamination restoration, and domain adaptation techniques to further improve and quantitatively evaluate the robustness of the proposed method under representative harsh industrial conditions.

4. Conclusions

To address the challenges of weak feature robustness, inaccurate motion estimation, and the difficulty of deploying algorithms in coal slime flotation industrial scenarios, this paper proposes a dynamic feature analysis method that integrates multi-scale feature enhancement with Spatially Adaptive Keypoint Regularization (SAKR). The proposed framework consists of the MFNet feature detection network and the SAKR strategy, enabling efficient and accurate froth feature extraction and motion representation. Experimental results on both public benchmark datasets and real industrial flotation data demonstrate that MFNet consistently outperforms conventional methods, including SIFT, ORB, SuperPoint, and LoFTR, in terms of keypoint detection and descriptor discriminability, effectively overcoming the limitations of poor robustness in traditional methods and the performance trade-offs of existing deep learning-based approaches. Moreover, the proposed regularization strategy optimizes the spatial distribution of keypoints, leading to improved matching stability and more motion feature extraction. Benefiting from its low inference latency, small memory footprint, and real-time processing capability, the proposed method satisfies the practical requirements of industrial online monitoring. By achieving an effective balance between accuracy, computational efficiency, and resource consumption, the proposed framework is well suited for complex industrial vision applications and provides a promising solution for intelligent coal slime flotation. Future work will focus on further optimizing the lightweight network architecture, enhancing the model's generalization capability under extreme operating conditions and its robustness in challenging industrial environments, thereby providing stronger technical support for the intelligent operation of the entire flotation process.

Acknowledgements

This work was supported by the Natural Science Foundation of Inner Mongolia Autonomous Region of China [grant number 2023QN05024]; and the Natural Science Foundation of Inner Mongolia Autonomous Region of China [grant number 2022MS05016].

References

ALDRICH, C., AVELAR, E., LIU, X., 2022. *Recent advances in flotation froth image analysis*. Minerals Engineering 188, 107823.

BALNTAS, V., LENC, K., VEDALDI, A., MIKOLAJCZYK, K., 2017. *HPatches: a benchmark and evaluation of handcrafted and learned local descriptors*. IEEE Conference on Computer Vision and Pattern Recognition 3852-3861.

BARROSO-LAGUNA, A., MIKOLAJCZYK, K., 2023. *Key.Net: keypoint detection by handcrafted and learned CNN filters revisited*. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 698-711.

CHEN, Y., JIANG, H., LI, C., JIA, X., GHAMISI, P., 2016. *Deep feature extraction and classification of hyperspectral images based on convolutional neural networks*. IEEE Transactions on Geoscience and Remote Sensing 54, 6232-6251.

DETONE, D., MALISIEWICZ, T., RABINOVICH, A., 2018. *SuperPoint: self-supervised interest point detection and description*. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 224-236.

DUSMANU, M., ROCCO, I., PAJDLA, T., POLLEFEYS, M., SIVIC, J., TORII, A., SATTLER, T., 2019. *D2-Net: a* *trainable CNN for joint description and detection of local features.* IEEE/CVF Conference on Computer Vision and Pattern Recognition 8084-8093.

GE, W.C., LIU, J., CHENG, Z.Y.Y., GAO, P., ZHU, Y.M., LI, H., LI, Y.J., ZHANG, J.H., YUAN, S., 2025. *New insights into the selective depression mechanism of galactomannan in flotation separation of chalcopyrite and talc.* Colloids and Surfaces A: Physicochemical and Engineering Aspects 721, 137263.

HE, M., WANG, J., NIU, J., 2024. *Investigation on the preparation of coal slime flotation microemulsion collector: A combination of experimental and simulation method.* Fuel 359, 130322.

HUANG, X.L., WANG, Z., DENG, B., CHEN, R.F., ZHOU, M.Y., HE, X., MI, S.L., LAN, C.S., WANG, J., 2026. *Understanding the influence of fine particles on froth stability in smithsonite flotation systems*. Minerals Engineering 244, 110291

KAARTINEN, J., HÄTÖNEN, J., HYÖTYNIEMI, H., MIETTUNEN, J., 2006. *Machine vision based control of zinc flotation – A case study.* Control Engineering Practice 14, 1455-1466.

LIN, T.Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., RAMANAN, D., DOLLÁR, P., ZITNICK, C.L., 2014. *Microsoft COCO: common objects in context.* Computer Vision – ECCV 2014 740-755.

LINDENBERGER, P., SARLIN, P.E., POLLEFEYS, M., 2023. *LightGlue: local feature matching at light speed.* IEEE/CVF International Conference on Computer Vision 17581-17592.

LIU, T., WAN, G., BAI, H., KONG, X., TANG, B., WANG, F., 2024. *Real-time video stabilization algorithm based on SuperPoint.* IEEE Transactions on Instrumentation and Measurement 73, 1-13.

LIU, Y.Y., LI, Y.P., ZHU, T.T., GE, Y., RUAN, X.D., 2011. *Imaging study of surfactant effect on bubble behaviors in froth flotation.* International Symposium on Water Resource and Environmental Protection 1199-1202.

LOWE, D. G., 2004. *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision 60, 91-110.

LU, F., LIU, H., LV, W., 2024. *Deep correlation and precise prediction between static features of froth images and clean coal ash content in coal flotation: an investigation based on deep learning and maximum likelihood estimation.* Measurement 224, 113843. https://doi.org/10.1016/j.measurement.2023.113843

LU, M., LIU, D., DENG, Y., WU, L., XIE, Y., CHEN, Z., 2020. *R-K algorithm: a novel dynamic feature matching method of flotation froth.* Measurement 156, 107581.

LUO, Z., ZHOU, L., BAI, X., CHEN, H., ZHANG, J., YAO, Y., LI, S., HUANG, K., LIU, Y., 2020. *ASLFeat: learning local features of accurate shape and localization.* IEEE/CVF Conference on Computer Vision and Pattern Recognition 6589-6598.

MA, X., WANG, H., XIE, W., 2025. *Research progress in froth flotation for recycling of spent lithium-ion batteries.* Minerals Engineering 227, 109275.

MA, Z., LIAO, Y., NI, X., LIANG, Z., ZHONG, S., 2021. *Zinc recovery from Wulagen sulphide flotation plant tail, Part II: Pilot tests using cyclonic static microbubble flotation column.* Canadian Metallurgical Quarterly 60, 118-129.

NAYAK, A., JENA, M.S., MANDRE, N.R., 2022. *Beneficiation of lead-zinc ores – a review.* Mineral Processing and Extractive Metallurgy Review 43, 564-583.

POLAT, M., POLAT, H., CHANDER, S., 2003. *Physical and chemical interactions in coal flotation.* International Journal of Mineral Processing 72, 199-213.

PROKOPOV, E., USACHEVA, D., RUMIANTCEVA, M., EFIMOVA, V., 2025. *Weak segmentation and unsupervised evaluation: application to froth flotation images.* VISIGRAPP: VISAPP.

REVAUD, J., WEINZAEPFEL, P., DE SOUZA, C., PION, N., CSURKA, G., CABON, Y., HUMENBERGER, M., 2019. *R2D2: repeatable and reliable detector and descriptor.* Advances in Neural Information Processing Systems 32, 12405-12415.

RUBLEE, E., RABAUD, V., KONOLIGE, K., BRADSKI, G., 2011. *ORB: an efficient alternative to SIFT or SURF*. IEEE International Conference on Computer Vision 2564-2571.

SHI, Y., CAI, J., SHAVIT, Y., MU, T., FENG, W., ZHANG, K., 2022. *ClusterGNN: cluster-based coarse-to-fine graph neural network for efficient feature matching.* IEEE/CVF Conference on Computer Vision and Pattern Recognition 12507-12516.

WANG, Y., HUANG, X., SHANG, H., 2024. *A multiscale network for deblurring coal slurry foam images.* Signal, Image and Video Processing 18, 4841-4849.

WEN, B., XIA, W., 2017. *Effect of particle shape on coal flotation.* Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 39, 1390-1394.

ZHANG, B.L., WANG, H.N., YAN, X.K., RAN, J.C., ZHANG, H.J., 2026. Research progress in flotation technology and equipment for fine coal. Coal Preparation Technology 54, 1-16.

ZHANG, R., YAO, T., YAN, L., 2023. *Feature point detection and description networks based on asymmetric convolution and the cross-resolution image-matching method.* International Journal of Intelligent Systems 2023, 5131440.

ZHANG, W., ZHAO, Y., 2023. *An improved SIFT algorithm for registration between SAR and optical images.* Scientific Reports 13, 6346.

ZHAO, L., ZHANG, Q., 2025. *A significant review of froth stability in mineral flotation.* Chemical Engineering Science 302, 120738.

ZHOU, Y., GAO, Z., BU, F., ZHAO, L., 2024. *A lightweight feature point detection model based on feature pyramid and attention.* IEEE International Conference on Electronic Technology, Communication and Information 1349-1357.

ZOCCHE, J. J., SEHN, L. M., PILLON, J. G., SCHNEIDER, C. H., OLIVO, E. F., RAUPP-PEREIRA, F., 2023. *Technosols in coal mining areas: Viability of combined use of agro-industry waste and synthetic gypsum in the restoration of areas degraded.* Cleaner Engineering and Technology 13, 100618.