Journal of Flood Risk Management
SPECIAL ISSUE Climate Change–Driven Debris Flow and Cascades
WILEY | CIWEM Chartered Institution of Water and Environmental Management

SPECIAL ISSUE PAPER OPEN ACCESS

Adaptive Prototype-Based Few-Shot Segmentation for Debris Flow Early Warning

Mingzhi Zhang[1,2,3] | Yueping Yin[1] | Chun Bao[4,5] | Rui Li[4] | Cuiqiong Zhou[6,7,8] | Zhikun Hu[4] | Frank Mugagga[9] | Yazidhi Bamutaze[9]

[1]China Institute of Geo-Environment Monitoring, Beijing, China | [2]School of Safety Science, Institute of Public Safety Research, Tsinghua University, Beijing, China | [3]China-Uganda Belt and Road Joint Laboratory on Natural Disaster Monitoring and Early Warning, Tsinghua University (Hefei), Hefei, China | [4]Zhiyang Innovation Technology Co., Ltd., Zibo, China | [5]Key Laboratory of Mountain Hazards and Earth Surface Process/Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China | [6]Yunnan Key Laboratory of Geohazard Forecast and Geoecological Restoration in Plateau Mountainous Area, Kunming, China | [7]Yunnan Institute of Geological Environment Monitoring, Kunming, China | [8]Key Laboratory of Geohazard Forecast and Geoecological Restoration in Plateau Mountainous Area, MNR, Kunming, China | [9]Department of Geography, Geoinformatics and Climatic Sciences, Makerere University, Kampala, Uganda

Correspondence: Yueping Yin (yinypcgs@hotmail.com)

Received: 1 July 2026 | **Revised:** 21 August 2026 | **Accepted:** 14 September 2026

Keywords: adaptive prototype construction | cross-modal feature enhancement | debris flow | early warning | few-shot semantic segmentation

ABSTRACT

Debris flows are highly destructive geohazards requiring rapid, accurate perception for effective early warning and emergency response. However, deep learning methods depend heavily on large datasets, whereas debris flow imagery is inherently scarce. Although few-shot semantic segmentation (FSS) addresses data limitations, existing methods suffer from semantic ambiguity and weak adaptability to complex target morphologies. To address these challenges, this study proposes a cross-modal few-shot segmentation framework and constructs a visible-light benchmark dataset named DebrisFlow-VisSeg (DFVS). The framework incorporates a Cross-Modal Feature Enhancement Unit (CFEU) to improve visual-semantic discrimination, an Adaptive Prototype Construction Mechanism (APCM) to generate robust prototypes for heterogeneous structures, and a Query Feature Disentanglement and Alignment Module (QF-DAM) to suppress background interference. Extensive experiments on DFVS, PASCAL-5^i, and COCO-20^i demonstrate that the method outperforms state-of-the-art approaches. Specifically, it achieves 76.31% mIoU under the 1-shot setting and 80.97% mIoU under the 5-shot setting on DFVS. Field deployment further verifies its practical applicability for real-world debris flow monitoring under data-scarce conditions.

1 | Introduction

Debris flows are sudden, catastrophic geohazards imposing critical threats to socio-ecological infrastructure in mountainous catchments (Cui et al. 2025; Peng et al. 2025). Conventional field monitoring predominantly relies on point-based contact sensors (e.g., rainfall gauges and geophones) (Yin et al. 2025; Yin and Gao 2024; Zhang et al. 2023). While mechanically principled, these methods suffer from limited spatial perception, substantial response delays, and high false-alarm rates under complex alpine conditions, failing to capture the continuous, multiscale, and heterogeneous spatial evolution of dynamic mudflows (Chmiel et al. 2021; Zhu et al. 2023; Pham and Kim 2022).

However, the operational translation of fully supervised deep networks remains restricted by the severe temporal

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Journal of Flood Risk Management* published by Chartered Institution of Water and Environmental Management and John Wiley & Sons Ltd.

and geographical scarcity of documented debris flow events, which imposes a severe data-dependency bottleneck during catastrophic geohazard mitigation tasks (Dissanayake et al. 2026; Tian et al. 2022). Deploying existing frameworks directly into erratic geohazard channels reveals major geo-domain bottlenecks:

The active source zones of geomorphic catchments are often composed of poorly sorted scoria/rocks, varying vegetation canopies, and complex terrain shadows that share high spectral and textural similarities with the hazardous mudflow fluid. Conventional models lack query feature purification.

Debris flows present continuous rheological changes (viscous, dilute, or blocky phases) with irregular boundaries and rapid geometric scaling across tiered geomorphic zones. Holistic prototype averaging is inherently incapable of capturing this immense multiphase heterogeneity.

To effectively bridge this geo-domain operational gap, this study proposes an integrated few-shot framework engineering explicitly for complex debris flow analysis. We construct and release the DebrisFlow-VisSeg (DFVS) dataset, comprised of 963 multiscenario, high-resolution visible-light images systematically captured across diverse catchments.

The main contributions of this work are summarized as follows:

1. The DFVS Dataset: We construct and publicly release a standardized, domain-specific visible-light benchmark dataset, filling a critical resource gap in intelligence-driven geohazard perception.
2. Cross-Modal Feature Enhancement Unit (CFEU): We introduce a visual-semantic fusion strategy utilizing textual embeddings as high-level semantic anchors to cleanly isolate dynamic foreground hazard bodies from chaotic catchment backgrounds.
3. Adaptive Prototype Construction Mechanism (APCM): We develop a multilevel prototype framework driven by an inverse-segmentation auxiliary task to model the immense physical heterogeneities and boundary uncertainties of dynamic mudflows.
4. Query Feature Disentanglement and Alignment Module (QF-DAM): We design a query purification architecture utilizing target-aware disentanglement and bidirectional attention (BiAttn) to ensure robust cross-branch interaction under erratic wildcard conditions.

Extensive experiments on DFVS, PASCAL-5^i (Everingham et al. 2010), and COCO-20^i (Lin et al. 2015) demonstrate that our method consistently outperforms state-of-the-art frameworks. Furthermore, successful operational deployment in Dabaini Gully (Yunnan Province, China) validates its practical engineering capacity for real-world geohazard field monitoring and multistage early warning under stringent data constraints. Unlike conventional prototype-based few-shot segmentation methods that mainly focus on support-query matching, the proposed framework is specifically designed according to the visual characteristics of debris flow scenes. CFEU enhances the semantic discriminability of support features and suppresses irrelevant background information, APCM constructs adaptive multilevel prototypes to represent the morphological heterogeneity and boundary uncertainty of debris flows, and QF-DAM further decouples query features and aligns them with support representations to reduce query-side background interference. These modules form a progressive feature-refinement framework that jointly addresses semantic ambiguity, morphological heterogeneity, and support-query discrepancy in data-scarce debris flow segmentation.

2 | Background and Related Work

Although mechanically principled, these methodologies are computationally intensive and sensitive to subjective heuristics or data variations (Chang et al. 2023; Long et al. 2015; Chen et al. 2018). Over the past decade, deep convolutional neural networks (e.g., FCN (Long et al. 2015), U-Net, DeepLab, and PSPNet [Zhao et al. 2017]) have advanced automated geohazard mapping via fully supervised pixel-level segmentation from optical or synthetic aperture radar (SAR) data. However, their operational deployment in real-time debris flow monitoring is severely restricted by two bottlenecks: the prohibitive annotation cost of dynamic, rapidly eroding flow boundaries and the catastrophic overfitting induced by the high geographic scarcity of documented disaster samples across unobserved catchments (Wang et al. 2020; Snell et al. 2017; Finn et al. 2017; Goodfellow et al. 2020).

To mitigate these data dependencies, few-shot semantic segmentation (FSS) frameworks (e.g., PANet (Ding et al. 2024), ASGNet (Li et al. 2021), and PFENet (Tian et al. 2020)) utilize meta-learning and dense support-query metric matching to segment novel categories using minimal annotated templates. First, standard architectures overlook query branch purification, leading to severe semantic confusion when confronted with massive environmental clutters (e.g., poorly sorted scoria/rocks, dense alpine canopies, and mountain shadows) that texturally mimic the hazardous fluid body (Chang et al. 2026; Nie et al. 2024; Zheng et al. 2025; Li et al. 2025; Pambala et al. 2021). Second, traditional pipelines rely on a singular, static prototype generated via holistic global pooling, which is structurally incapable of modeling the immense physical heterogeneities of debris flow surges, such as multiphase rheological transitions (viscous, dilute, or blocky) and wide-bandwidth structural dispersions across multitier geomorphic zones (Chen et al. 2024).

To bridge this operational geo-domain gap, this study introduces a specialized few-shot framework driven by cross-modal feature enhancement, adaptive multilevel prototype parsing, and target-aware query disentanglement.

3 | Problem Definition

In the few-shot debris flow semantic segmentation task, we are given a support set $S = \{s_k^I, s_k^M\}$ and a query image I_q. Where, $s_k^I \in R^{H\times W\times 3}$ represents the k-th support image, and $s_k^M \in \{0,1\}^{H\times W}$ is corresponding binary mask, indicating the

pixel area of the debris flow target in the image. K Is the number of support samples, usually 1 or 5. Our goal is to predict the pixel-level mask $M_q \in \{0,1\}^{H\times W}$ of the debris flow target in the query image I_q. The debris flow categories in the support set and query set are unseen during the training phase but known during the testing phase.

4 | Data and Method

4.1 | Datasets

Evaluation is performed on three repositories: our self-built DFVS dataset and two generic FSS benchmarks, PASCAL-5^i and COCO-20^i.

4.1.1 | The DFVS Dataset

DFVS is constructed to fill the vacancy of debris flow segmentation benchmarks, containing 963 annotated RGB images captured under diverse terrain, lighting, and debris rheological conditions. We follow standard 1-shot and 5-shot few-shot protocols for evaluation. The DFVS dataset represents debris flow scenes acquired in complex mountainous environments, where heterogeneous terrain and surface materials surround the monitored flow channels. The observed scenes contain debris flow bodies together with rocks, sediment deposits, vegetation, water, and terrain shadows. Variations in flow materials, surface structure, target morphology, illumination, and surrounding terrain produce substantial visual differences among samples. These characteristics result in irregular target boundaries, strong target-background similarity, and considerable background interference, making DFVS a challenging benchmark for few-shot debris flow segmentation.

4.1.2 | Generic FSS Benchmarks

PASCAL-5^i and COCO-20^i contain 20 and 80 semantic categories, respectively.

4.1.3 | Evaluation Metrics

Segmentation performance is quantitatively evaluated via Mean Intersection over Union (mIoU) and Foreground-Background IoU (FB-IoU).

4.2 | Few-Shot Semantic Segmentation Framework

We propose a debris-flow-oriented few-shot segmentation framework. The full architecture in Figure 1 consists of three core modules: CFEU, APCM, and QF-DAM, which cooperate to achieve reliable segmentation under limited annotated geohazard data. APCM produces multiscale prototypes aided by an inverse segmentation auxiliary task to model complex debris flow morphologies. QF-DAM extracts clean target features from queries and aligns them with support prototypes to counteract background interference. Detailed module implementations are elaborated below. The three modules form a sequential feature-refinement framework with complementary functions. CFEU first enhances support representations

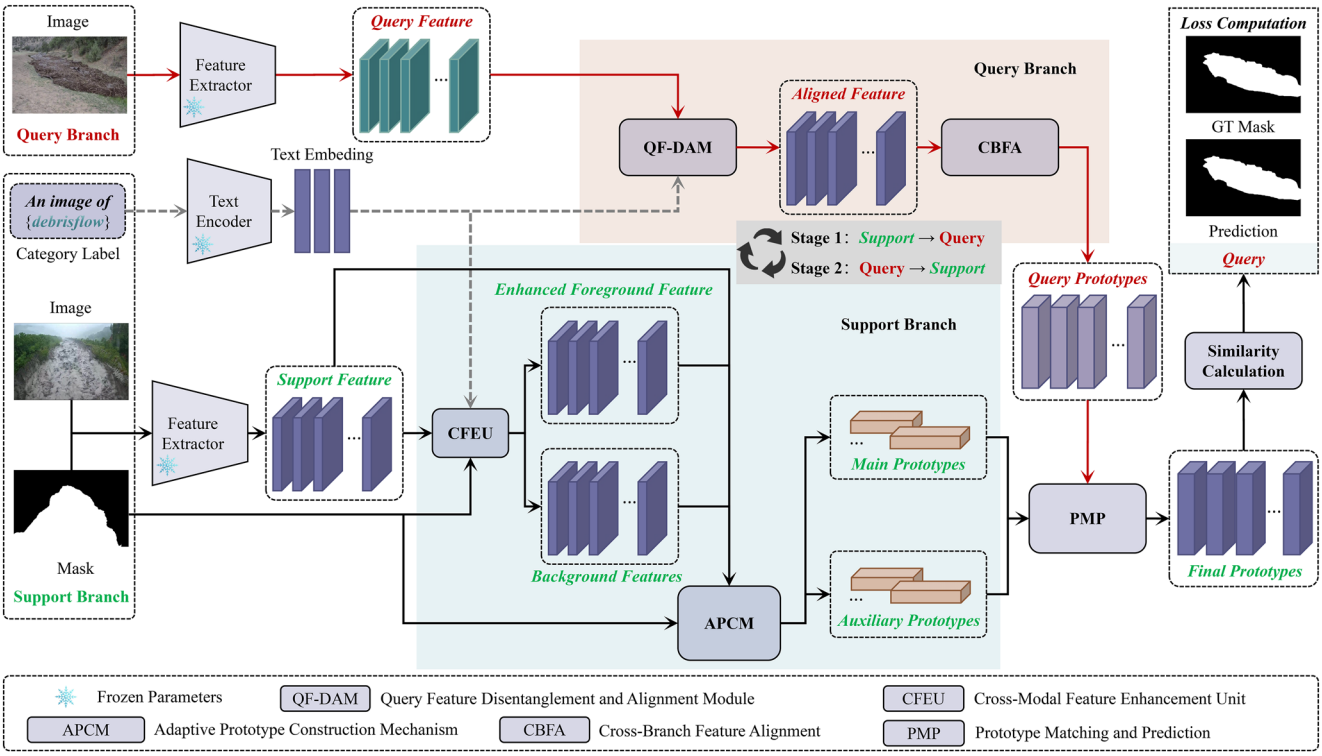

FIGURE 1 | Overview of the proposed few-shot semantic segmentation framework for debris flow. Its main components include input, backbone network, text encoder, support branch, query branch, prototype matching and prediction, and loss function.

by suppressing irrelevant background information and incorporating semantic guidance. The resulting support features are then used by APCM to construct adaptive multilevel prototypes that preserve complementary global and boundary characteristics. Based on these prototypes, QF-DAM further decouples query features and performs support-query alignment to reduce background interference and representation discrepancy. This design establishes a coherent relationship among support feature enhancement, prototype construction, and query feature refinement.

4.2.1 | CFEU

Standard few-shot segmentation models only extract features via backbone networks, suffering from feature ambiguity and detail loss due to high visual resemblance between debris flows and complex backgrounds.

The architecture of CFEU is shown in Figure 2. Given a support image I_s and its binary mask M_s, deep features $X_s \in R^{H_f \times W_f \times C}$ are extracted via an ImageNet pretrained ResNet-50 backbone $B(\cdot)$, where H_f, W_f, and C denote feature map dimensions. Concurrently, a pretrained Contrastive Language-Image Pretraining (CLIP) text encoder $E_T(\cdot)$ encodes the category prompt "debrisflow" into a text embedding $E_c \in R^D$. CFEU generates enhanced background features $X_{s,BG}$ and foreground features $X_{s,FG}$ via cross-modal fusion, preserving target feature integrity while suppressing background distractions.

4.2.1.1 | Background Feature Reconstruction. To eliminate complex background noise (e.g., vegetation, rocks, and water bodies) with textures similar to mudflows, an inverse mask $(1 - M_s)$ is used to locate background regions. We integrate text embedding E_c into a gating module $G_{BG}(\cdot)$ and a self-attention module $\mathrm{Attn}_{\mathrm{BG}}(\cdot)$ for adaptive background reconstruction, formulated as:

$$X_{s,BG} = G_{BG}\left(X_s \odot \left(1 - M_s\right), E_c\right) \oplus \left(X_s \odot \mathrm{Attn}_{\mathrm{BG}}\left(1 - M_s, X_s, E_c\right)\right) \tag{1}$$

where $\odot$ and $\oplus$ denote element-wise multiplication and feature fusion (concatenation followed by 1×1 convolution), respectively. The binary mask M_s is resized to match the feature map resolution. $G_{BG}(\cdot)$ Is a lightweight network with convolution and Sigmoid activation that outputs pixel-wise weight maps. In $\mathrm{Attn}_{\mathrm{BG}}(\cdot)$, masked background features serve as Query, while linear-transformed E_c or X_s act as Key and Value to suppress semantically irrelevant background pixels, minimizing noise for subsequent prototype generation.

4.2.1.2 | Foreground Feature Reconstruction. To ensure the purity and accuracy of prototype learning, M_s is applied to isolate target regions. Features are refined via text guidance, a foreground gating module $G_{FG}(\cdot)$, and an attention module $\mathrm{Attn}_{\mathrm{FG}}(\cdot)$ to obtain enhanced foreground features $X_{s,FG}$:

$$X_{s,FG} = G_{FG}\left(X_s \odot M_s, E_c\right) \oplus \left(X_s \odot \mathrm{Attn}_{\mathrm{FG}}\left(M_s, X_s, E_c\right)\right) \tag{2}$$

$G_{FG}(\cdot)$ And $\mathrm{Attn}_{\mathrm{FG}}(\cdot)$ share architectures with the background counterparts but focus on debris flow regions to filter foreground noise and align visual patterns with the textual concept "debrisflow." A cross-modal interaction module $I_{CM}(\cdot)$ is further introduced to resolve visual ambiguity by deep fusion of foreground features and text embeddings, yielding the refined feature $X'_{s,FG}$:

$$X'_{s,FG} = \mathrm{Softmax}\left(\frac{\left(W_K X_{s,FG}\right) \cdot \left(W_Q E_c\right)^T}{\sqrt{d_k}}\right)\left(W_V X_{s,FG}\right) \tag{3}$$

where W_Q, W_K, and W_V are learnable linear projection matrices, d_k is the projection dimension (set to C), and T denotes matrix transpose. This attention mechanism dynamically aligns visual features onto text-matched target semantics. Unlike simple feature recombination, CFEU achieves fine-grained foreground-background separation and cross-modal fusion, producing highly discriminative support features.

4.2.2 | APCM

Prototypes dominate few-shot segmentation performance. Conventional methods generate a single prototype via global average pooling, failing to capture debris flow heterogeneity induced by material, texture, terrain, and illumination variations.

4.2.2.1 | Initial Prototype Generation. We compute initial background prototype P_{BG} and foreground prototype P_{FG} by average pooling over CFEU-enhanced features to establish the basis for subsequent refinement:

$$P_{BG} = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i \in \Omega} s, BG^X_{(k)}(i) \cdot I\left[s^M_{(k)}(i) = 0\right]}{\sum_{i \in \Omega} I\left[s^M_{(k)}(i) = 0\right]} \tag{4}$$

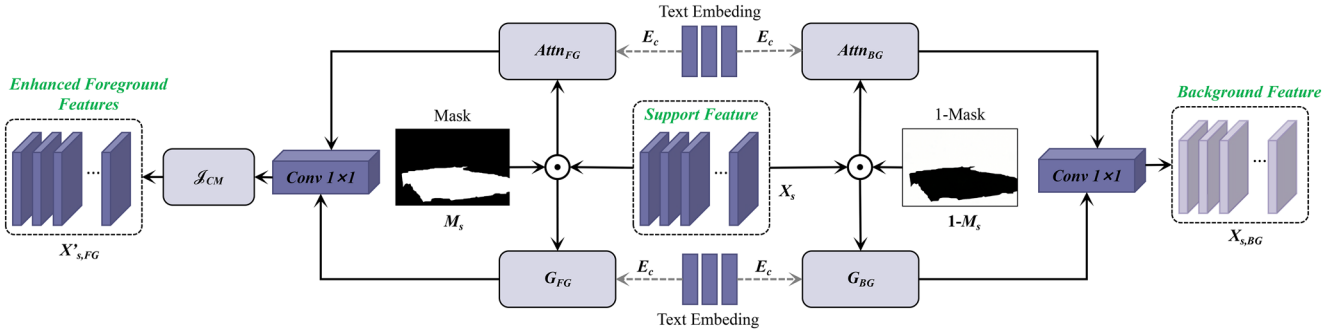

FIGURE 2 | Architecture of the cross-modal feature enhancement unit (CFEU).

$$P_{FG} = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i\in\Omega} X'_{s,FG}(i)\cdot I\left[s^{M}_{(k)}(i)=1\right]}{\sum_{i\in\Omega} I\left[s^{M}_{(k)}(i)=1\right]} \tag{5}$$

where K is the shot count, Ω denotes all pixel positions, s, $BG^{X}_{(k)}(i)$ and $X'_{s,FG}(i)$ represent regional features of the k-th sample at pixel i, and $I[\cdot]$ is an indicator function.

4.2.2.2 | Multilevel Prototype Refinement. Guided by pseudo-masks $\widetilde{M}_s$ from the inverse segmentation task, we construct main prototype P_{Main} and auxiliary prototype P_{Aux} to capture morphological diversity and boundary details. Unlike APENet's direct reliance on ground-truth masks, our auxiliary task outputs adaptive structural priors under limited support samples (Chang et al. 2026). P_{Main} Aggregates support features weighted by pseudo-mask confidence to represent global debris flow semantics:

$$P_{\text{Main}} = \frac{\sum_{i\in\Omega} X_s(i)\cdot \text{Weight}_{\text{Main}}\left(\widetilde{M}_s(i), X_s(i)\right)}{\sum_{i\in\Omega}\text{Weight}_{\text{Main}}\left(\widetilde{M}_s(i), X_s(i)\right)} \tag{6}$$

where $X_s(i)$ is the support feature at pixel i,$\widetilde{M}_s(i)\in[0,1]$ is supervised by ground-truth M_s, and the weighting function $\text{Weight}_{\text{Main}}(\cdot)$ is formulated as:

$$\text{Weight}_{\text{Main}} = \text{Sigmoid}\left(\alpha\cdot\widetilde{M}_s(i)+\beta\cdot Sim\left(X_s(i), P_{FG}\right)\right) \tag{7}$$

here α and β are learnable hyperparameters, and $Sim(\cdot)$ computes cosine similarity to prioritize high-confidence foreground pixels. Conversely, P_{Aux} compensates for missing edge details by extracting features from discrepancy regions between M_s and $\widetilde{M}_s$:

$$P_{Aux} = \frac{\sum_{i\in\Omega} X_s(i)\cdot \text{Weight}_{\text{Aux}}\left(\text{Diff}\left(M_s,\widetilde{M}_s\right)(i), X_s(i)\right)}{\sum_{i\in\Omega}\text{Weight}_{\text{Aux}}\left(\text{Diff}\left(M_s,\widetilde{M}_s\right)(i), X_s(i)\right)} \tag{8}$$

where $\text{Diff}\left(M_s,\widetilde{M}_s\right)(i) = |M_s(i)-\widetilde{M}_s(i)|$ quantifies local prediction uncertainty. The auxiliary weighting function is defined as:

$$\text{Weight}_{\text{Aux}} = \text{Sigmoid}\left(\gamma\cdot\text{Diff}\left(M_s,\widetilde{M}_s\right)(i)+\delta\cdot Sim\left(X_s(i),P_{FG}\right)\right) \tag{9}$$

where γ and δ are learnable parameters to emphasize uncertain boundary zones, enabling fine-grained semantic capture.

4.2.3 | QF-DAM

Raw query features contain heavy background noise. QF-DAM purifies target representations and eliminates prototype bias via cross-branch bidirectional alignment. We extract query feature X_q from I_q. A lightweight predictor produces pseudo-mask $q^{\widetilde{M}}_{\text{pre}}$ using initial prototype similarity and text embedding guidance. Decoder $D(\cdot)$ recovers lost fine details to generate pure foreground feature $X_{q,FG}$ as shown in Equation (10).

$$X_{q,FG} = D\left(I_q\odot\text{Upsample}\left(q^{\widetilde{M}}_{\text{pre}}\right)\right)\oplus\left(X_q\odot\text{Attn}_{Q,FG}\left(q^{\widetilde{M}}_{\text{pre}},X_q,E_c\right)\right) \tag{10}$$

4.2.3.1 | Cross-Branch Feature Alignment (CBFA). Prototype bias arises when support prototypes fail to generalize to query target appearances. BiAttn enables mutual refinement between enhanced support feature $X'_{s,FG}$ and purified query feature $X_{q,FG}$. Aligned feature X_{Aligned} is defined in Equation (11), with detailed attention calculation in Equation (12).

$$X_{\text{Aligned}} = \text{BiAttn}\left(X'_{s,FG}, X_{q,FG}\right) \tag{11}$$

$$\begin{aligned} X_{\text{Aligned}} = {} & \text{Softmax}\left(\frac{\left(W_{S,K}X'_{s,FG}\right)\cdot\left(W_{Q,Q}X_{q,FG}\right)^{T}}{\sqrt{d_a}}\right)\left(W_{S,V}X'_{s,FG}\right) \\ & +\text{Softmax}\left(\frac{\left(W_{Q,K}X_{q,FG}\right)\cdot\left(W_{S,Q}X\prime_{s,FG}\right)^{T}}{\sqrt{d_a}}\right)\left(W_{Q,V}X_{q,FG}\right) \end{aligned} \tag{12}$$

$W_{S,K}$, $W_{Q,Q}$, $W_{S,V}$, $W_{Q,K}$, $W_{S,Q}$, $W_{Q,V}$ denote linear projections; $d_a=\sqrt{C}$ is the attention scaling factor.

4.3 | Prototype Matching and Prediction

We compute pixel-wise cosine similarity between raw query feature and initial prototypes to obtain coarse mask q^{M}_{init} as shown in Equation (13).

$$q^{M}_{\text{init}}(i) = \text{argmax}_{j\in\{FG,BG\}}\left(\frac{P_j\cdot X_q(i)}{\|P_j\|\cdot\|X_q(i)\|}\right) \tag{13}$$

Final segmentation integrates similarities between X_{Aligned} and dual prototypes P_{Main}, P_{Aux} then upsamples to original resolution as:

$$q^{M}_{\text{final}} = \text{Upsample}\left(\text{Sim}\left(X_{\text{Aligned}},P_{\text{Main}}\right)+\text{Sim}\left(X_{\text{Aligned}},P_{\text{Aux}}\right)\right) \tag{14}$$

4.4 | Loss Function

To train our model end-to-end, we design a comprehensive loss function, which includes the main segmentation task loss and the auxiliary inverse segmentation task loss.

Binary cross-entropy L_{Seg} supervises final prediction against ground truth as shown in Equation (15).

$$L_{\text{Seg}} = L_{CE}\left(q^{M}_{\text{final}}, q^{M}_{GT}\right) \tag{15}$$

Auxiliary loss L_{Aux} constrains support pseudo-masks from APCM as:

$$L_{\text{Aux}} = L_{CE}\left(\widetilde{M}_s, s^{M}_{GT}\right) \tag{16}$$

Total loss weights two tasks with fixed hyperparameters $\lambda_1=1.0$, $\lambda_2=0.5$ as shown in Equation (17).

$$L_{\text{Total}} = \lambda_1\cdot L_{\text{Seg}}+\lambda_2\cdot L_{\text{Aux}} \tag{17}$$

TABLE 1 | Performance comparison on PASCAL-5^i in terms of mIoU.

Method	1-Shot					5-Shot				
	Fold-0	Fold-1	Fold-2	Fold-3	mIoU	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
CANet (Zhang et al. 2019)	52.50	65.90	51.30	51.90	55.40	55.50	67.80	51.90	53.20	57.10
CRNet (Liu, Zhang, Lin, and Liu 2020)	—	—	—	—	55.70	—	—	—	—	58.80
PPNet (Liu, Zhang, Zhang, and He 2020)	48.58	60.58	55.71	46.47	52.84	58.85	68.28	66.77	57.98	62.97
HSNet (Min et al. 2021)	64.30	70.70	60.30	60.50	63.95	70.30	73.20	67.40	67.10	69.45
PFENet (Tian et al. 2020)	61.70	69.50	55.40	56.30	60.73	63.10	70.70	55.80	57.90	61.88
ASNet (Kang and Cho 2022)	68.90	71.70	61.10	62.70	66.10	72.60	74.30	65.30	67.10	69.83
DCP (Lang et al. 2022)	63.81	70.54	61.16	55.69	62.80	67.19	73.15	66.39	64.48	67.80
BAM (Lang et al. 2023)	69.15	74.66	67.79	61.72	68.33	71.84	75.67	71.95	67.54	71.75
Ours	**73.20**	**78.50**	**72.10**	**66.80**	**71.55**	**76.50**	**79.30**	**76.20**	**72.40**	**75.04**

Note: The bold values represent the performance of our proposed method on the corresponding datasets. They are used to highlight the results of our method for clearer presentation.

TABLE 2 | Performance comparison on COCO-20^i in terms of mIoU and FB-IoU.

Method	1-Shot						5-Shot					
	Fold-0	Fold-1	Fold-2	Fold-3	MIoU	FB-IoU	Fold-0	Fold-1	Fold-2	Fold-3	MIoU	FB-IoU
RPMMs (Yang et al. 2020)	29.53	36.82	28.94	27.02	30.58	—	33.82	41.96	32.99	33.33	35.52	—
HFA (Liu et al. 2021)	28.65	36.02	30.16	33.28	32.03	—	32.69	42.12	30.35	36.19	35.34	—
RePRI (Boudiaf et al. 2021)	32.00	38.70	32.70	33.10	34.13	—	39.30	45.40	39.70	41.80	41.55	—
ASGNet (Li et al. 2021)	—	—	—	—	34.56	60.39	—	—	—	—	42.48	66.96
MMNet (Wu et al. 2021)	34.90	41.00	37.20	37.00	37.53	—	37.00	40.30	39.30	36.00	38.15	—
HSNet (Min et al. 2021)	36.30	43.10	38.70	38.70	39.20	68.20	43.30	51.30	48.20	45.00	46.95	70.70
NTRENet (Liu et al. 2022)	36.80	42.60	39.90	37.90	39.30	68.50	38.20	44.10	40.40	38.40	40.28	69.20
DCP (Lang et al. 2022)	40.89	43.77	42.60	38.29	41.39	—	45.82	49.66	43.69	46.62	46.48	—
BAM (Lang et al. 2023)	43.86	51.38	47.86	44.47	46.89	72.32	49.76	55.43	52.31	50.19	51.92	74.65
APENet (Chang et al. 2026)	44.60	51.40	49.00	47.70	48.2	—	52.90	56.20	53.50	51.10	53.40	—
Ours	**48.50**	**55.70**	**52.30**	**49.20**	**49.87**	**75.64**	**54.30**	**60.10**	**57.80**	**55.60**	**55.87**	**78.32**

Note: The bold values represent the performance of our proposed method on the corresponding datasets. They are used to highlight the results of our method for clearer presentation.

5 | Experiments and Results

Our model is implemented in PyTorch. We adopt ImageNet-pretrained ResNet-50 as the backbone and resize all input images to 473×473 for unified feature dimensions. SGD serves as the optimizer with an initial learning rate of 0.0025, polynomial decay, weight decay 0.0001, and momentum 0.9. The model is trained for 10,000 epochs on RTX4090 GPUs. Loss weights $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ are validated on the development set. The pretrained CLIP model acts as the text encoder to supply high-quality cross-modal embeddings for CFEU and QF-DAM.

TABLE 3 | Performance comparison on DFVS dataset in terms of mIoU.

Method	1-Shot					5-Shot				
	Fold-0	Fold-1	Fold-2	Fold-3	mIoU	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
CANet (Zhang et al. 2019)	58.32	62.47	56.89	59.13	59.20	61.54	65.38	60.27	62.81	62.50
CRNet (Liu, Zhang, Lin, and Liu 2020)	59.15	63.28	57.64	59.87	60.00	62.36	66.19	61.05	63.62	63.30
PPNet (Liu, Zhang, Zhang, and He 2020)	57.28	61.35	55.76	58.02	58.10	60.43	64.27	59.16	61.74	61.40
HSNet (Min et al. 2021)	65.43	69.28	63.75	66.01	66.12	68.65	72.43	67.98	70.24	70.08
PFENet (Tian et al. 2020)	63.17	67.02	61.54	63.86	63.89	66.39	70.15	65.72	67.98	67.56
ASNet (Kang and Cho 2022)	67.85	71.63	65.98	68.24	68.43	71.07	74.82	70.16	72.41	72.11
DCP (Lang et al. 2022)	64.92	68.76	63.21	65.53	65.61	68.14	71.98	67.43	69.76	69.33
BAM (Lang et al. 2023)	70.36	74.12	68.59	71.03	71.03	73.58	77.34	72.81	75.15	74.72
APENet (Chang et al. 2026)	72.24	75.36	70.31	74.34	74.45	74.48	80.32	75.43	78.84	76.42
Ours	**75.69**	**79.45**	**73.82**	**76.27**	**76.31**	**79.93**	**83.68**	**78.95**	**81.31**	**80.97**

Note: The bold values represent the performance of our proposed method on the corresponding datasets. They are used to highlight the results of our method for clearer presentation.

5.1 | Comparison With SOTA Methods

We compare the proposed framework against current mainstream FSS baselines, including CANet (Zhang et al. 2019), CRNet (Liu, Zhang, Lin, and Liu 2020), PPNet (Liu, Zhang, Zhang, and He 2020), HSNet (Min et al. 2021), PFENet (Tian et al. 2020), ASNet (Kang and Cho 2022), DCP (Lang et al. 2022), BAM (Lang et al. 2023), RPMMs (Yang et al. 2020), HFA (Liu et al. 2021), RePRI (Boudiaf et al. 2021), ASGNet (Li et al. 2021), MMNet (Wu et al. 2021), NTRENet (Liu et al. 2022), and APENet (Chang et al. 2026). As tabulated in Table 1, the proposed method achieves an average mIoU of 71.55% under the 1-shot setting and 75.04% under the 5-shot setting, outperforming the state-of-the-art BAM method by 3.22% and 3.29%, respectively. This superiority stems from the enhanced feature discrimination provided by the CFEU alongside the highly representative prototypes generated by the APCM. The consistent performance gains across all PASCAL-5^i folds demonstrate the favorable generalization capability of the proposed framework under different few-shot settings.

Table 2 reports the performance on the COCO-20^i dataset. Under the 1-shot setting, the average mIoU and FB-IoU reach 49.87% and 75.64% (surpassing BAM by 2.98% and 3.32%); under the 5-shot setting, they reach 55.87% and 78.32% (surpassing BAM by 3.95% and 3.67%). The performance advantage on the COCO-20^i benchmark indicates that the proposed feature-refinement and query-alignment mechanisms help suppress the influence of complex background information.

As shown in Table 3, the performance advantages of our framework are most pronounced on the self-built domain-specific DFVS dataset.

The performance improvement on DFVS is particularly meaningful because this dataset contains debris flow scenes with irregular target morphology, heterogeneous backgrounds, and substantial appearance variations. The comparison between 1-shot and 5-shot settings further demonstrates how additional support examples improve the representation of the target class.

Qualitative results illustrated in Figure 3 visually confirm that the proposed method delivers superior, high-fidelity pixel-level mapping with precise edge discrimination across diverse debris flow scenarios.

5.2 | Ablation Study

To verify the individual contributions and mathematical validity of each proposed module, rigorous ablation experiments are executed on the DFVS dataset under the 1-shot protocol as shown in Table 4.

5.2.1 | Baseline Limitations

The vanilla prototypical network yields a baseline mIoU of only 56.32%. This deficit underscores that standard metric-learning pipelines lack the structural capacity to handle intense target-background spectral overlaps and morphological variations inherent to chaotic wilderness debris flow imagery.

5.2.2 | Support Purification via CFEU

Integrating the CFEU induces a substantial accuracy surge (+8.43% mIoU). This confirms that anchoring support visual streams with high-level cross-modal linguistic constraints ("debris flow") effectively gates out highly cluttered geomorphic background noise (e.g., weathered scoria, scree, and alpine canopy).

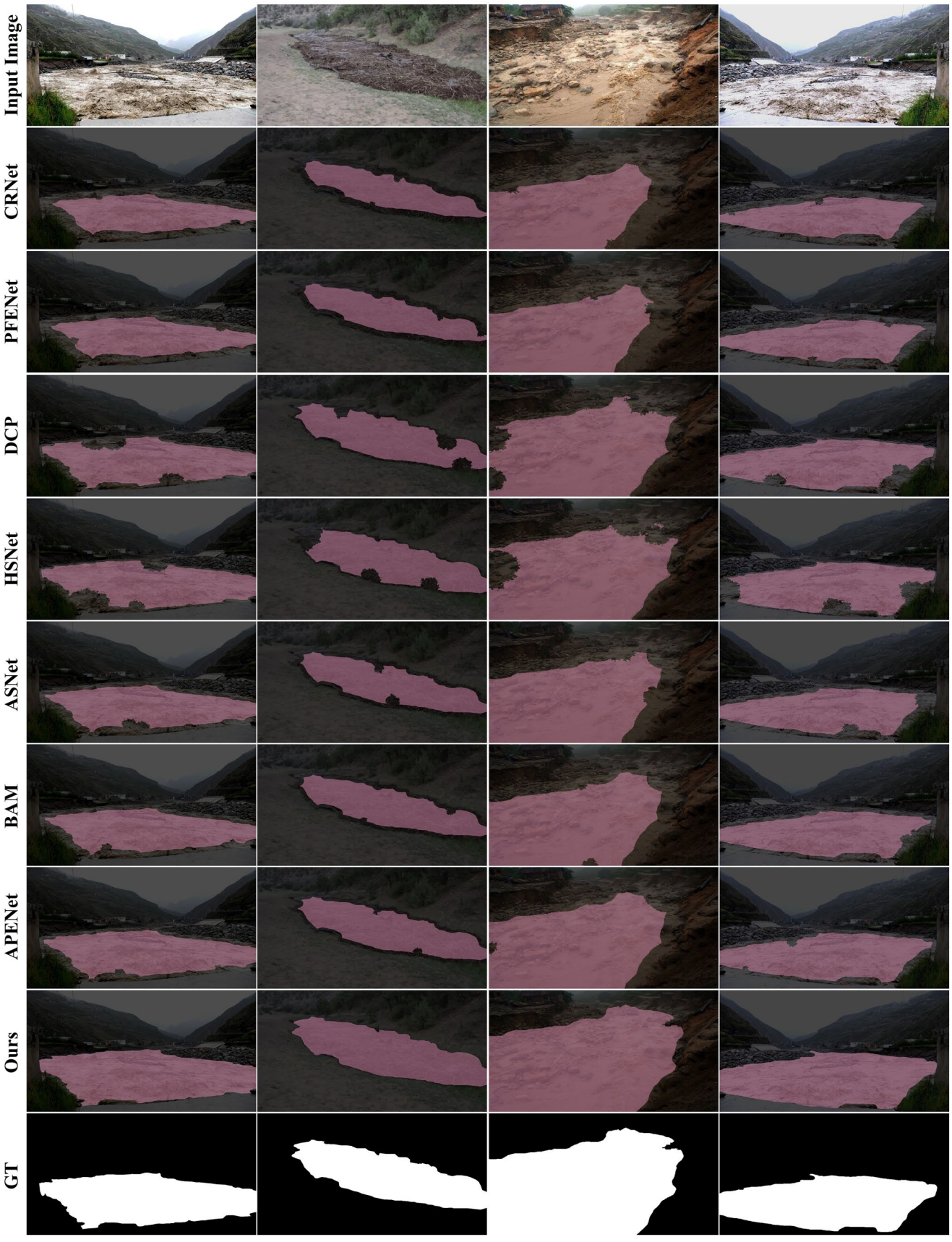

FIGURE 3 | Segmentation visualization of the proposed method on the DFVS dataset.

5.2.3 | Heterogeneity Modeling via APCM

Incorporating the APCM boosts performance by 6.96% mIoU. Unlike rigid global pooling that flattens target features, the multilevel prototypes driven by an inverse-segmentation auxiliary task successfully capture the multiphase texture dispersion and complex boundary uncertainties of active landslide/debris surges.

5.2.4 | Query Refinement and Alignment via QF-DAM

Deploying query feature disentanglement without alignment lifts the baseline to 67.53% mIoU (+11.21%), proving that pseudo-mask cues and spatial detail recovery can cleanly isolate target primitives from unconstrained environments. Activating the BiAttn-driven CBFA mechanism yields an additional 4.36% mIoU gain (Config 5), demonstrating that deep interbranch mutual learning effectively neutralizes "prototype bias" induced by disparate catchment topographies or illumination variations.

TABLE 4 | Ablation study results of each module (mIoU %).

Model configuration	mIoU	Performance improvement
Baseline	56.32	—
Baseline + CFEU	64.75	+8.43
Baseline + APCM	63.28	+6.96
Baseline + QF-DAM (w/o CBFA)	67.53	+11.21
Baseline + QF-DAM (w/CBFA)	71.89	+15.57
Baseline + CFEU + APCM + QF-DAM (w/CBFA)	76.31	+19.99

5.2.5 | Systemic Synergistic Cohesion

Engaging all components simultaneously achieves the peak performance of 76.31% mIoU, outperforming the baseline by 19.99%.

The ablation results indicate that the three proposed modules provide complementary contributions to segmentation performance. CFEU improves support feature discriminability, APCM enhances prototype representation by capturing heterogeneous target characteristics, and QF-DAM further reduces query-side background interference and support-query discrepancy. The complete framework achieves the best overall performance, demonstrating the effectiveness of their collaborative design.

5.3 | On-Site Monitoring Deployment in Dabaini Gully

To evaluate the operational robustness of the proposed framework, an intelligence-driven vision sensor integrated with our model was deployed in Dabaini Gully (Dongchuan District,

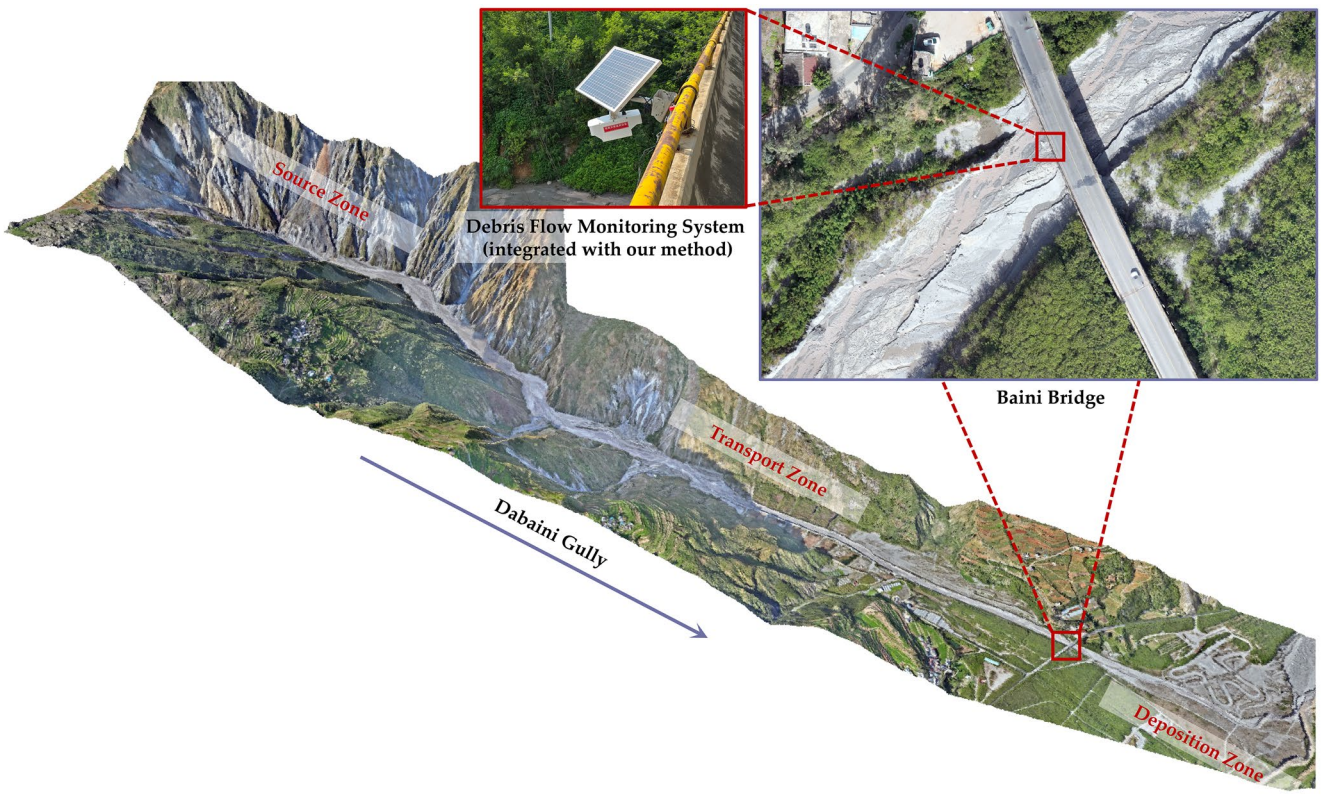

FIGURE 4 | Field monitoring system and geomorphic zonation of Dabaini Gully: Complete topographic structure of the gully, deployment of our integrated monitoring device in the transport zone, and division of source, transport, and deposition zones.

Yunnan Province, China)—this highly heterogeneous, fragmented geomorphic structure, coupled with prominent vertical precipitation and dense mountain fog during the monsoon season (May to October), creates an exceptionally chaotic outdoor environment that imposes extreme visual noise and severe illumination fluctuations on optical monitoring streams.

For real-time disaster perception, our solar-powered monitoring device was strategically installed at the Baini Bridge within the transport zone as shown in Figure 4 to capture continuous, unconstrained visual signals during peak energy transmission. During this field deployment, the integrated system successfully recorded and processed a complete paroxysmal debris flow event. Figure 5 systematically illustrates the pixel-level segmentation performance generated by our model during three critical evolutionary stages of this live event: the initial front arrival, structural development, and peak propagation phases.

The proposed cross-modal FSS framework achieves exceptional boundary delineation accuracy and structural alignment throughout the entire disaster life cycle. Our method cleanly isolates the active foreground hazard body while suppressing background environmental clutter.

6 | Discussion

Their complementary modeling accurately describes the nonrigid, chaotic spatial evolution of debris flows. This RGB-based framework still has inherent limitations in extreme field environments. Segmentation accuracy drops sharply for tiny debris flow targets, heavy shadow occlusion, and muddy flows with spectra similar to turbid rivers; dense vegetation fully covering flow channels leads to complete detection failure due to missing visible cues, as the model only relies on surface RGB signals without geophysical reasoning. Despite the demonstrated effectiveness, several challenging conditions may still affect the robustness of the proposed framework. Severe illumination variations can increase the visual similarity between debris flow materials and surrounding terrain, while dense vegetation may partially obscure the target region. Extremely small debris flow targets provide limited spatial information for reliable feature matching, and adverse weather conditions such as fog and rainfall may further degrade visual observations. Future research will investigate multimodal sensing and vision-language foundation models to provide complementary physical and semantic information. Thermal infrared, radar, or other complementary sensing modalities may improve target perception when visible information is degraded, while vision-language foundation models may provide additional semantic and contextual priors for visually ambiguous or partially occluded debris flow scenes.

In an operational monitoring pipeline, the proposed segmentation model can serve as a spatial perception component between image acquisition and subsequent hazard assessment. The generated segmentation masks provide spatial information about the observed debris flow, while their temporal evolution can be further analyzed for automatic event detection and process-state assessment. These visual indicators can subsequently be integrated with rainfall, vibration, water-level, or other monitoring information to support multisource hazard assessment and warning decision-making. Therefore, the proposed framework provides a visual perception basis for operational monitoring and decision support.



FIGURE 5 | Segmentation results of the 06-30-2025 debris flow event at different stages: Original visible-light images, binary segmentation masks, and overlay visualization of our proposed method.

From the perspective of integrated natural hazard monitoring, FSS can transform continuous image observations into spatially explicit hazard information. The segmented target regions can support continuous hazard perception, while their temporal evolution can provide information for dynamic process monitoring and subsequent early-warning decision-making. In mountainous watersheds, debris flows and flash floods may be driven by related rainfall processes and may interact within the same watershed system. Therefore, visual perception of dynamic hazardous processes may provide a complementary information source for coordinated monitoring of multiple hazards. Although the present study focuses on debris flow segmentation, the proposed framework provides a potential perception component for future integrated disaster monitoring and intelligent risk-management systems.

7 | Conclusion

Extensive evaluations on DFVS, PASCAL-5^i, and COCO-20^i benchmarks demonstrate that our method consistently outperforms state-of-the-art baselines, while ablation experiments confirm the strong synergistic effects among all proposed modules. Future research will focus on structural optimization for extremely complex disaster scenarios and expanding the framework's adaptability to broader catastrophic geohazards. Beyond debris flow monitoring, the proposed framework provides a potential paradigm for data-scarce geohazard perception under complex environmental conditions. Its combination of cross-modal feature enhancement, adaptive prototype construction, and query feature decoupling and alignment may be further extended to other dynamic geohazards, such as landslides and rockfalls, where annotated observations are limited and target morphology is highly variable. Future integration with multimodal sensing, semantic foundation models, and multisource monitoring information may further promote the development of intelligent disaster perception, monitoring, and early-warning systems.

Funding

This work is supported by the Special Project for Natural Disaster Prevention Equipment of the Ministry of Industry and Information Technology of China, the National Key Research and Development Program of China (2025YFE0212800), and the Geological Survey Program of China (DD20251300201).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Boudiaf, M., H. Kervadec, Z. I. Masud, P. Piantanida, I. B. Ayed, and J. Dolz. 2021. "Few-Shot Segmentation Without Meta-Learning: A Good Transductive Inference Is All You Need?" In *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13974–13983. IEEE.

Chang, Z., X. Gao, D. Chang, Y. Li, and Y. Lu. 2026. "APENet: Task-Aware Adaptation Prototype Evolution Network for Few-Shot Semantic Segmentation." *Expert Systems With Applications* 306: 130906.

Chang, Z., Y. Lu, X. Ran, X. Gao, and X. Wang. 2023. "Few-Shot Semantic Segmentation: A Review on Recent Approaches." *Neural Computing and Applications* 35: 18251–18275.

Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. 2018. "DeepLab: Semantic Image Segmentation With Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40: 834–848.

Chen, Y., R. Jiang, Y. Zheng, B. Sheng, Z.-X. Yang, and E. Wu. 2024. "Dual Branch Multi-Level Semantic Learning for Few-Shot Segmentation." *IEEE Transactions on Image Processing* 33: 1432–1447.

Chmiel, M., F. Walter, M. Wenner, Z. Zhang, B. W. McArdell, and C. Hibert. 2021. "Machine Learning Improves Debris Flow Warning." *Geophysical Research Letters* 48: e2020GL090874.

Cui, P., G. Zhang, Y. Wang, et al. 2025. "Evolutionary Characteristics and Frontier Scientific Issues in Flash Flood and Debris Flow Disasters Risk Under Climate Change." *BNSFC* 39: 959–967.

Ding, J., Z. Zhang, Q. Wang, and H. Wang. 2024. "SCTrans: Self-Align and Cross-Align Transformer for Few-Shot Segmentation." *Image and Vision Computing* 142: 104893.

Dissanayake, T., Y. George, D. Mahapatra, S. Sridharan, C. Fookes, and Z. Ge. 2026. "Few-Shot Learning for Medical Image Segmentation: A Review and Comparative Study." *ACM Computing Surveys* 58: 1–36.

Everingham, M., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. 2010. "The Pascal Visual Object Classes (VOC) Challenge." *International Journal of Computer Vision* 88: 303–338.

Finn, C., P. Abbeel, and S. Levine. 2017. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks." In *Proceedings of the Proceedings of the 34th International Conference on Machine Learning*, 1126–1135. PMLR.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, et al. 2020. "Generative Adversarial Networks." *Communications of the ACM* 63: 139–144.

Kang, D., and M. Cho. 2022. "Integrative Few-Shot Learning for Classification and Segmentation." In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9969–9980. IEEE.

Lang, C., G. Cheng, B. Tu, C. Li, and J. Han. 2023. "Base and Meta: A New Perspective on Few-Shot Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45: 10669–10686.

Lang, C., B. Tu, G. Cheng, and J. Han. 2022. "Beyond the Prototype: Divide-and-Conquer Proxies for Few-Shot Segmentation." In *Proceedings of the Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 1024–1030. International Joint Conferences on Artificial Intelligence Organization.

Li, G., V. Jampani, L. Sevilla-Lara, D. Sun, J. Kim, and J. Kim. 2021. "Adaptive Prototype Learning and Allocation for Few-Shot Segmentation." In *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8330–8339. IEEE.

Li, H., G. Huang, X. Yuan, et al. 2025. "Psanet: Prototype-Guided Salient Attention for Few-Shot Segmentation." *Visual Computer* 41: 2987–3001.

Lin, T.-Y., M. Maire, S. Belongie, et al. 2015. "Microsoft COCO: Common Objects in Context."

Liu, B., J. Jiao, and Q. Ye. 2021. "Harmonic Feature Activation for Few-Shot Semantic Segmentation." *IEEE Transactions on Image Processing* 30: 3142–3153.

Liu, W., C. Zhang, G. Lin, and F. Liu. 2020. "CRNet: Cross-Reference Networks for Few-Shot Segmentation." In *Proceedings of the 2020 IEEE/*

CVF Conference on Computer Vision and Pattern Recognition, 4164–4172. IEEE.

Liu, Y., N. Liu, Q. Cao, X. Yao, J. Han, and L. Shao. 2022. "Learning Non-Target Knowledge for Few-Shot Semantic Segmentation." In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11563–11572. IEEE.

Liu, Y., X. Zhang, S. Zhang, and X. He. 2020. "Part-Aware Prototype Network for Few-Shot Semantic Segmentation." In *European Conference on Computer Vision*, 142–158. Springer.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." 3431–3440.

Min, J., D. Kang, and M. Cho. 2021. "Hypercorrelation Squeeze for Few-Shot Segmentation." In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, 6941–6952. IEEE.

Nie, J., Y. Xing, G. Zhang, et al. 2024. "Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining." In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3380–3390. IEEE.

Pambala, A. K., T. Dutta, and S. Biswas. 2021. "SML: Semantic Meta-Learning for Few-Shot Semantic Segmentation☆." *Pattern Recognition Letters* 147: 93–99.

Peng, J., P. Cui, H. Wang, et al. 2025. "Conceptual Framework for the Construction of a Natural Disaster Prevention and Mitigation System." *BNSFC* 39: 864–871.

Pham, M.-V., and Y.-T. Kim. 2022. "Debris Flow Detection and Velocity Estimation Using Deep Convolutional Neural Network and Image Processing." *Landslides* 19: 2473–2488.

Snell, J., K. Swersky, and R. Zemel. 2017. "Prototypical Networks for Few-Shot Learning." In *Proceedings of the Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc.

Tian, Z., H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia. 2020. "Prior Guided Feature Enrichment Network for Few-Shot Segmentation." *IEEE Transactions on Neural Networks and Learning Systems* 2: 1050–1065.

Tian, Z., H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia. 2022. "Prior Guided Feature Enrichment Network for Few-Shot Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44: 1050–1065.

Wang, P., J. Cheng, F. Hao, L. Wang, and W. Feng. 2020. "Embedded Adaptive Cross-Modulation Neural Network for Few-Shot Learning." *Neural Computing & Applications* 32: 5505–5515.

Wu, Z., X. Shi, G. Lin, and J. Cai. 2021. "Learning Meta-Class Memory for Few-Shot Semantic Segmentation." In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, 497–506. IEEE.

Yang, B., C. Liu, B. Li, J. Jiao, and Q. Ye. 2020. "Prototype Mixture Models for Few-Shot Semantic Segmentation."

Yin, Y., and S. Gao. 2024. "Research on High-Altitude and Long-Runout Rockslides: Review and Prospects." *Chinese Journal of Geological Hazard and Control* 35: 1–18.

Yin, Y., B. Li, Y. Gao, W. Wang, S. Zhang, and S. Zhu. 2025. "Challenges and Countermeasures for High-Altitude and Long-Runout Geological Hazards in the Qinghai-Xizang Plateau." *BNSFC* 39: 986–992.

Zhang, C., G. Lin, F. Liu, R. Yao, and C. Shen. 2019. "CANet: Class-Agnostic Segmentation Networks With Iterative Refinement and Attentive Few-Shot Learning." In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5212–5221. IEEE.

Zhang, M., F. Yang, J. Ma, Z. Fu, Z. Huang, and J. Li. 2023. "Exploration of Blockchain Technology Application in the Construction of National Risk Warning System on Landslides." *Hydrogeology & Engineering Geology* 50: 168–174.

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. 2017. "Pyramid Scene Parsing Network." 2881–2890.

Zheng, Z., Z. Teng, C. Xie, et al. 2025. "Prediction of Debris Flow Hazard Based on Multi-Source Remote Sensing Data: A Case Study of Dulongjiang Township, Yunnan Province." *Remote Sensing Applications: Society and Environment* 39: 101719.

Zhu, L., Y. Jia, S. Huang, et al. 2023. "DeFlow: Self-Supervised 3D Motion Estimation of Debris Flow." In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 6508–6517. IEEE.