*remote sensing*

MDPI

*Article*

# Reliability-Aware Feedback-Controlled Fusion for Spatially Disjoint HSI–LiDAR Classification

**Davaajargal Myagmarsuren, Haibin Wu *, Aili Wang and Minhui Wang**

Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China; 2212200001@stu.hrbust.edu.cn (D.M.); aili925@hrbust.edu.cn (A.W.)
* Correspondence: woo@hrbust.edu.cn

## Highlights

### What are the main findings?

- RAFC-Fusion couples a library of HSI–LiDAR fusion operators with a validation-gated closed-loop controller and feedback-guided refinement.
- Under sample-center spatially disjoint evaluation, with seven labels per class, the framework achieves mean OA values of 70.00%, 73.20%, and 66.34% on Houston2013, Muufl, and Augsburg, respectively; its advantage over the baseline is preserved under a patch-disjoint protocol.

### What are the implications of the main findings?

- No single fusion operator is best on all three datasets; a single adaptive configuration matches the dataset-specific best fixed operator.
- Validation-gated closed-loop training is an effective alternative to training a fixed fusion model once, and adaptive operator switching becomes viable as the label budget increases.

## Abstract

This study presents a reliability-aware feedback-controlled fusion framework for hyperspectral image (HSI) and LiDAR classification under spatially disjoint conditions. Instead of committing to a fixed fusion configuration, the framework couples a library of fusion operators (geometry-aware fusion, adaptive gating, cross-modal transformer fusion and physics-informed fusion), integrated within a unified GhostConv-based architecture, with a closed-loop controller that evaluates candidate operators on validation data, accepts them only when they improve validation accuracy, and applies feedback-guided residual refinement. Experiments are conducted on Houston2013, Muufl, and Augsburg under a sample-center spatially disjoint protocol, in which training and validation samples are drawn from development blocks that are spatially disjoint from the test blocks, with seven labeled samples per class and ten random partitions. The framework achieves mean overall accuracies of 70.00%, 73.20%, and 66.34%, improving the Baseline-Ghost model on Houston2013, Muufl, and Augsburg; the improvement achieved by closed-loop training is preserved under a patch-disjoint protocol. Budget-matched analyses show that, with seven labels per class, the gain stems mainly from validation-gated closed-loop training and feedback-guided refinement, whereas the benefit of adaptive operator switching emerges as the label budget increases.

Check for updates

Academic Editor: Jining Yan

Received: 24 September 2026
Revised: 22 October 2026
Accepted: 29 October 2026
Published: 30 October 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

**Keywords:** hyperspectral image; LiDAR data; multimodal fusion; spatially disjoint; feedback control; remote sensing

## 1. Introduction

Hyperspectral image (HSI) and light detection and ranging (LiDAR) fusion have attracted increasing attention in remote sensing because the two modalities provide highly complementary information for land-cover classification. HSI captures rich spectral signatures for material discrimination, whereas LiDAR provides elevation and structural information that improves spatial interpretation and geometric understanding. Their combination has therefore demonstrated superior classification performance compared with single-modal approaches in complex urban and natural scenes [1–4]. Recent multimodal fusion methods have introduced various deep learning architectures, including convolutional fusion networks, graph-based models, transformer-based fusion, and adaptive attention mechanisms [5–7].

CNN-based methods effectively capture local spectral–spatial representations, while transformer architectures improve long-range cross-modal interaction modeling [8–10]. In addition, several architectures have recently been proposed to reduce computational redundancy while preserving representation capability in multimodal remote sensing systems [11–15]. Despite these advances, most existing approaches rely on fixed fusion architectures that remain unchanged throughout training and across datasets. To ensure equitable comparison, baselines were selected based on methodological relevance, reproducibility, and compatibility with strict multimodal HSI–LiDAR evaluation.

However, multimodal remote sensing datasets often exhibit substantially different spectral distributions, spatial complexity, modality agreement, and scene heterogeneity. Consequently, a fusion strategy that performs well on one dataset may not generalize effectively to another. This limitation becomes more critical under low-sample and spatially disjoint evaluation settings, where the optimal fusion behavior may evolve during training as the network transitions from coarse feature extraction to fine-grained semantic discrimination. Current fusion methods in general lack any mechanism to dynamically adjust fusion operators based on dataset characteristics and feedback from the learning process. Recently, adaptive optimization and feedback-control strategies have been explored in deep learning and multimodal systems [16–21]. These methods demonstrate that closed-loop feedback can improve optimization stability and adaptive learning behavior. Nevertheless, existing approaches mainly focus on parameter optimization or network tuning within fixed architectures rather than dynamic multimodal fusion operator selection. Moreover, most HSI–LiDAR fusion frameworks still employ static operator configurations without explicitly considering reliability-aware adaptive fusion transitions during training.

Another important challenge in HSI–LiDAR classification is evaluation reliability. Many existing studies adopt random pixel-wise sampling, which may introduce spatial overlap between development and testing samples, thereby overestimating spatial generalization performance [10,22,23]. Recent studies increasingly advocate spatially disjoint evaluation to better assess robustness under realistic deployment conditions, particularly for disjoint hyperspectral classification [24,25].

To address these limitations, this article proposes a Reliability-Aware Feedback-Controlled Fusion (RAFC-Fusion), a reliability-aware closed-loop fusion framework for HSI–LiDAR classification. Instead of relying on a single fixed fusion architecture, the proposed framework dynamically selects and updates fusion operators during training using validation-driven feedback control. The framework integrates multiple complementary fu-

sion operators, including geometry-aware fusion, adaptive gating, physics-informed fusion, cross-modal transformer fusion, and ghost-convolution-based fusion modules within a unified adaptive framework. The main contributions of this work are summarized as follows:

- A reliability-aware feedback-controlled fusion framework is proposed for HSI–LiDAR classification, in which fusion-operator prioritization is formulated as a prior-initialized, validation-driven greedy search. The procedure is closely related to validation-based model selection—its distinguishing elements are a training-derived feedforward prior, a history-based reliability score, an acceptance margin, and warm-started candidate training, while a PID-inspired signal re-weights operators according to the validation error and regulates the strength of feedback-guided refinement. A budget-matched analysis shows that, with seven labeled samples per class, the improvement stems from validation-gated closed-loop training and feedback-guided refinement, and that adaptive operator switching becomes viable as the label budget increases;
- A reliability-aware fusion framework integrating geometry-aware, adaptive gating, cross-modal transformer, and physics-informed operators is developed within a unified GhostConv backbone;
- A hybrid reliability controller combining training-derived feedforward characterization and validation-driven feedback is introduced for operator prioritization under spatially disjoint supervision, and it is compared explicitly with direct validation-based operator selection and adaptive operator weighting under matched training budgets;
- A unified sample-center spatially disjoint evaluation protocol is applied across three benchmark datasets for reproducible comparison, and its optimism relative to a patch-disjoint protocol is quantified.

## 2. Related Works

### *2.1. Multimodal Fusion Strategies*

Multimodal fusion in remote sensing encompasses a spectrum of approaches ranging from early fusion (concatenation-based) to intermediate fusion (feature-level) and late fusion (decision-level), and hybrid fusion [26]. Especially, recent surveys have developed comprehensive taxonomies that categorize fusion methods based on their structural approach and the stage at which they combine HSI and LiDAR modalities [27]. The way hyperspectral imaging and LiDAR data are used together to characterize land cover has changed significantly in the last decade. HSI offers a wealth of spectral information across hundreds of bands, while LiDAR provides detailed elevation and geometric features [28]. Traditional fusion methods have relied on simple concatenation or manually designed features [1], but deep learning has revolutionized this domain.

Convolutional neural networks (CNNs) have been proven to be effective in hyperspectral imaging (HSI) and light detection and ranging (LiDAR) classification, and many architectures have been proposed to extract spectral, spatial, and elevation information [5,6]. Wang et al. [29] introduced a two-branch CNN framework for HSI-LiDAR classification, which brings improvements over single-source methods. Hierarchical fusion strategies include triple-branch CNN architectures [23] that extract features from each model separately before fusion, and transformer-based methods have emerged as powerful alternatives that use self-attention mechanisms to model long-range spectral dependencies and spatial relationships [9,30].

Attention-based fusion methods have emerged as the main paradigm because of their ability to detect long-range correlations [10,31], and adaptive gating mechanisms learn whether to trust each mode based on the input attributes [32]. Cross-modal transformers, on the other hand, use both self-attention (intra-modal) and cross-attention (inter-modal) modules to model complex relationships [33,34]. Graph-based methods have also shown

promise, and graph convolutional networks (GCNs) [35] and hypergraph networks [36] model the high-order relationships between pixels to represent the complex spatial–spectral elevation interactions. Adversarial learning approaches [37] have been explored to extract similar high-level semantic information and complementary discriminative features from the two modalities. Wang et al. [38] showed that integrating semantic thinking with HSI-LiDAR fusion can improve the accuracy of feature discrimination and classification through a multimodal semantic collaborative classification network, which uses visual language methodologies to produce a domain-specific, knowledge-rich land cover classifier. Recent methods further combine global and local feature extraction [39] or exploit directional sparsity and cross-modal contrast-consensus learning [40] to adapt the fusion of hyperspectral and SAR/LiDAR data; these approaches adapt the fusion within a single architecture, whereas the present work adapts the choice among several fusion operators through validation feedback.

Recent research has concentrated on progressively advanced fusion architecture. Pande and Banerjee [41] presented self-looping convolutional networks with weight sharing to improve interaction between different types of data. However, many existing modular fusion methods mainly use static fusion architectures that process all inputs equally, regardless of the data properties or the complexity of the dataset [10,42]. This limitation has motivated our work on dynamic fusion and has motivated us to explore more efficient strategies.

### *2.2. Efficient Neural Networks and Ghost Convolutions*

As hyperspectral imaging moves to mobile and embedded platforms, computational efficiency is becoming increasingly important [43]. Han et al. [11] introduced GhostNet, a lightweight architecture that generates redundant feature maps with low cost and reduces computational cost while maintaining accuracy. This technique splits feature generation into intrinsic features generated by normal convolution and ghosts generated by low-cost linear transformation. Paoletti et al. [12] successfully adapted ghost convolution for single-source hyperspectral image classification and showed that the ghost module can achieve competitive accuracy at a fraction of the computational cost compared to standard CNNs.

The remote sensing community has explored a variety of efficiency-oriented techniques. MobileNet architectures [44–46] are based on deep separable convolutions, and the recent MobileNetV4 [13] has demonstrated high performance with state-of-the-art results on mobile platforms. Meng et al. [47] suggested a global-local multi-grain transformer for HSI, while Li et al. [14] introduced MambaHSI employing linear complex state-space models for spatial spectrum processing. These are examples of efficient transformers that could work. Recent extensive research [15] has chronicled the progression of the discipline towards lightweight architectures tailored for real-time implementation.

Hybrid CNN–transformer architectures represent a promising direction. iFormer [48] shows that it is possible to obtain MobileNet-level performance while preserving the transformation expression by integrating convolutions and transforming components. LowFormer [49] presents hardware-efficient models for edge devices and ensures real-time performance on resource-constrained platforms.

However, to our knowledge, there has been little research on lightweight methods such as Ghost convolution in multi-dimensional HSI-LiDAR fusion. This motivates the integration of GhostConv with multiple candidate fusion operators and reliability-aware feedback-controlled operator prioritization.

### *2.3. Control Theory in Deep Learning*

The intersection of control theory and machine learning is gaining increasing attention as researchers seek a principled approach to algorithm design and optimization. Neural

architecture search (NAS) [21,50] pioneered the automated discovery of optimal network architectures, but it requires significant computational resources (thousands of GPU-days) and lacks theoretical convergence guarantees. Control theory has been less applied to deep learning, and An et al. [17] pioneered the use of PID controllers in the stochastic optimization of deep networks and demonstrated improved convergence properties at CVPR 2018. This important work has stimulated further research into control-based optimization strategies.

Recent breakthrough studies have validated control theory approaches in high-impact areas. Chen et al. [18] proposed a detailed framework for faster optimization using proportional-integral-derivative controllers. They performed theoretical convergence analysis and showed that gradient-based optimization can be considered a beneficial continuous-time dynamical system for feedback control. Zhang et al. [51] introduced a gradient-decreasing probability-dependent PID controller in adaptive model tuning, demonstrating that control principles can reduce overconfidence in neural networks predictions.

Dai et al. [19] introduced an adaptive PID optimizer that integrates proportional, integral, and conventional control theory coefficients to mitigate overconfidence in Adam optimization. This significantly speeds up and improves the convergence accuracy. Shi et al. [52] and Hao et al. [53] showed that the DDPG and TD3 algorithms can effectively optimize PID parameters in real time for complex nonlinear systems and allow the use of feedback control for adaptive parameter tuning.

The specific case of multimodal fusion has limited previous work. Wang et al. [20] applied PID control to infrared-visible image fusion (FusionPID) and built a closed-loop system to optimize the fusion quality performance. However, this work focuses on parameter tuning into a single fusion method for image quality optimization, rather than selecting a strategy among multiple operators for the classification task.

Our work applies control-inspired feedback to the selection among multiple fusion operators. Unlike existing methods that mainly use PID-type control for hyperparameter tuning [19,51] or optimization acceleration [18,20], the controller considered here prioritizes candidate fusion operators according to validation feedback. Procedurally, this is closely related to validation-based model selection and greedy architecture search [21,50]; candidate operators are trained and retained only when they improve validation accuracy. The proposed controller differs from plain validation-based selection mainly in its training-derived feedforward prior, history-based reliability score, acceptance margin and warm-started candidate training. Section 4.6.4 therefore compares it with direct validation-based selection and with adaptive operator weighting under matched training budgets.

## 3. Methodology

### *3.1. Overview of the Proposed Framework*

Hyperspectral–LiDAR fusion methods typically employ a fixed fusion architecture throughout training and inference, assuming that a single fusion strategy is universally optimal across datasets. However, under sample-center spatially disjoint evaluation, the statistical characteristics of development and held-out regions often differ substantially because of variations in terrain morphology, spectral diversity, spatial fragmentation, and cross-modal consistency. Consequently, a fixed fusion operator may not provide robust performance across datasets with heterogeneous spatial distributions.

To address this limitation, we propose a Reliability-Aware Feedback-Controlled Fusion (RAFC-Fusion) framework that formulates multimodal fusion as a closed-loop adaptive optimization process. Rather than relying on predefined fusion architecture, the proposed framework dynamically prioritizes fusion operators according to both dataset characteristics and online validation feedback.

As illustrated in Figure 1, the framework consists of five sequential stages:

(1) Multimodal feature extraction;
(2) Feedforward dataset characterization to estimate scene complexity from five complementary indicators;
(3) Reliability-aware feedback control that continuously evaluates validation performance and updates controller states;
(4) Adaptive operator prioritization over the proposed multimodal fusion operator library;
(5) Closed-loop optimization that iteratively refines operator preference until the stopping criterion is reached.

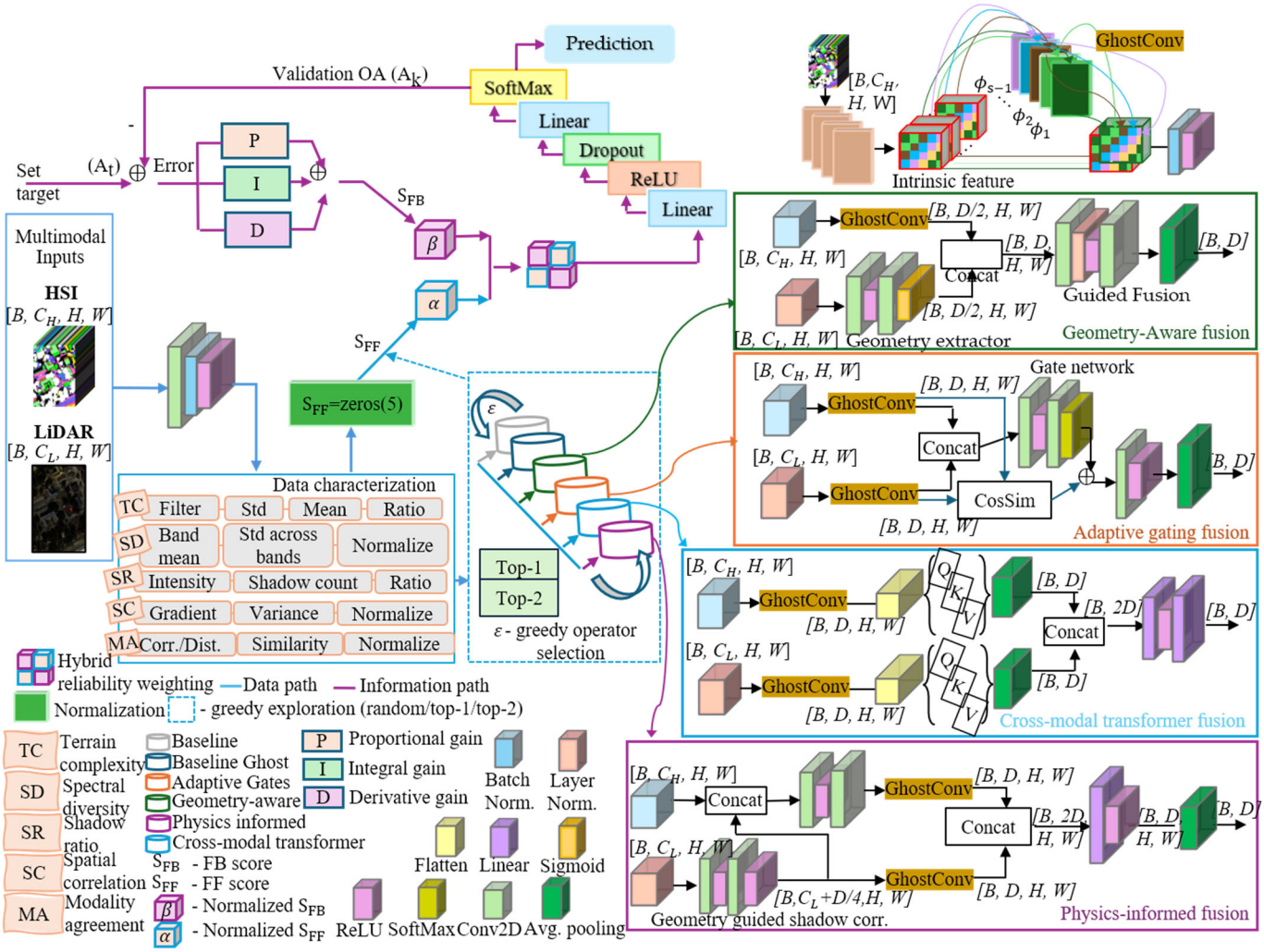



**Figure 1.** Overview of the proposed reliability-aware feedback-controlled multimodal fusion framework. The HSI and LiDAR inputs (**left**) are characterized by five indicators (**bottom center**) that yield the feedforward operator prior (Section 3.2). The closed-loop controller (upper **left** and **center**) compares the validation OA with the target accuracy, computes the PID-inspired signal and combines feedforward and feedback scores for ε-greedy operator selection (Section 3.4). The candidate operators of the GhostConv-based library (right; Section 3.3) are trained and evaluated through the classification head (**top**).

Unlike conventional dynamic fusion methods, the proposed framework jointly exploits prior dataset characterization and online reliability feedback to continuously adjust fusion behavior during optimization. This closed-loop design enables adaptive fusion decisions under extremely limited supervision while maintaining stable generalization across spatially disjoint regions.

Let the hyperspectral image and LiDAR data observations be denoted as $X^H \in R^{B\times C_H\times H\times W}$, $X^L \in R^{B\times C_L\times H\times W}$, where B denotes batch size, $C_H$ and $C_L$ represent spectral and LiDAR channels, and H and W indicate spatial dimensions. The objective is to learn a multimodal mapping F: $(X^H, X^L) \rightarrow Y$, that predicts land-cover labels Y. As shown in Figure 1, the proposed framework jointly integrates feature extraction, feedforward dataset characterization, reliability-aware feedback control, and adaptive operator prioritization.

### *3.2. Feedforward Dataset Characterization*

To guide operator selection through feedforward prediction, we introduce five quantitative metrics that capture key physical and statistical properties of multimodal HSI–LiDAR data. These metrics estimate terrain variability, spectral variation, illumination artifacts, spatial structure, and modality consistency. Altogether, they provide an interpretable, heuristic initial ordering of the fusion operators before training, which forms the feedforward (FF) stage of the adaptive fusion system; the ordering is subsequently corrected by validation feedback, and its predictive value is examined in Section 4.6.4 (Appendix A Table A1).

1. Terrain Complexity (TC)

Given LiDAR elevation map $X_L$, horizontal and vertical gradients are computed as

$$\begin{aligned}\nabla_x X_L(i,j) &= X_L(i+1,j) - X_L(i,j),\\ \nabla_y X_L(i,j), &= X_L(i,j+1) - X_L(i,j).\end{aligned} \tag{1}$$

where $\nabla_x$ and $\nabla_y$ are gradients. The local terrain slope magnitude is

$$G(i,j) = \sqrt{\left(\nabla_x X_L(i,j)\right)^2 + \left(\nabla_y X_L(i,j)\right)^2} \tag{2}$$

Terrain complexity (normalized) is defined as

$$TC = \frac{\mathrm{std}(G)}{\mathrm{std}(X_L) + \varepsilon} \tag{3}$$

where $\mathrm{std}(\cdot)$ is the elevation standard deviation and $\varepsilon = 10^{-8}$ avoids numerical instability. Larger TC indicates stronger terrain variation and favors geometry-aware fusion. Since the raw values are dataset-dependent, TC is subsequently normalized into a reliability score in (0,1) before operator prioritization.

2. Spectral Diversity (SD)

For each hyperspectral band *c*,

$$CV_c = \frac{\sigma_c}{\mu_c + \varepsilon} \tag{4}$$

where $\sigma_c$ and $\mu_c$ are the standard deviation and the mean of the c-th spectral band c, respectively. SD is computed as follows

$$SD = \frac{1}{C_H}\sum_{c=1}^{C_H} CV_c \tag{5}$$

where $C_H$ is the number of hyperspectral bands. A higher SD indicates greater spectral heterogeneity and favors transformer-based fusion. Such cases favor fusion mechanisms capable of modeling long-range spectral dependencies. SD is subsequently normalized using the training data statistics described in Section 3.2.

3. Shadow Ratio (SR)

Shadow areas are detected using a combination of luminance and slope criteria, which follows the physical principle that shadows generally have low spectral reflectance and occur on steep terrain [54]. Average image brightness is computed by

$$B(i,\ j) = \frac{1}{C_H}\sum_{c=1}^{C_H} X_H(c,\ i,\ j) \tag{6}$$

Pixels satisfying

$$B(i,\ j) < \overline{B} \wedge G(i,\ j) > \overline{G} \tag{7}$$

are regarded as shadow pixels, where $\overline{B}$ and $\overline{G}$ denote the mean brightness and mean slope, respectively. The shadow ratio is

$$SR = \frac{1}{H \cdot W}\sum_{i,j} 1_{Shadow}(i,j) \tag{8}$$

where $1(\cdot)$ is the indicator function. High SR indicates significant illumination variation and favors the physics-informed fusion strategies. SR is normalized before reliability estimation.

4. Spatial Correlation (SC)

Spatial correlation measures the similarity of neighboring HSI patches through lag-1 autocorrelation [55]. For each spectral band, a lag-1 shifted image is constructed,

$$X_H^{\leftarrow}[c,i,j] = X_H[c,i,j+1] \tag{9}$$

where $X_H^{\leftarrow}$ is the lag-shifted version of HSI for autocorrelation. Spatial correlation is then

$$SC = clip\left(\frac{\sum_{b,c,p} \tilde{x}_{b,c}(p)\tilde{x}_{b,c}(p^+)}{\sum_{b,c,p} \tilde{x}_{b,c}(p)^2},\ 0,\ 1\right) \tag{10}$$

where $\tilde{x}_{b,c}(p)$ is the value of band c at position p of the b-th training patch after subtraction of the patch-wise band mean, and $p^+$ is the next position in row-major order. Lower SC reflects fragmented spatial structures and generally benefits cross-modal transformer fusion. SC is converted into a bounded reliability score before feedback initialization.

5. Modality Agreement (MA)

Modality agreement quantifies how well HSI and LiDAR features correspond spatially [56]. First,

$$F_H = \frac{1}{C_H}\sum_{c=1}^{C_H} X_H^{(c)}, F_L = \frac{1}{C_L}\sum_{c=1}^{C_L} X_L^{(c)}. \tag{11}$$

After spatial alignment (bilinear interpolation if necessary), both features are flattened and mean-centered $\overline{F}_H$ and $\overline{F}_L$. Their normalized correlation is

$$r = \frac{\tilde{F}_H^T \tilde{F}_L}{|\tilde{F}_H|_2 \cdot |\tilde{F}_L|_2} \tag{12}$$

which is mapped into

$$MA = \frac{r+1}{2},\ MA \epsilon [0,1]. \tag{13}$$

A low MA value indicates weak consistency between HSI and LiDAR representations and therefore increases the relevance of adaptive gating. The centered cosine correlation is mapped to the interval [0,1], so no additional scaling is required.

6. Reliability Normalization

The five characterization metrics—terrain complexity (TC), spectral diversity (SD), shadow ratio (SR), spatial correlation (SC), and modality agreement (MA)—capture complementary geometric, spectral, illumination-related, spatial, and cross-modal properties of the HSI–LiDAR data. Because these metrics differ in their numerical ranges and physical interpretations, they are standardized and transformed into bounded scores before being used to initialize the feedforward fusion-operator priorities.

Although these metrics provide dataset-level characterization when aggregated, their normalization statistics are estimated from the distribution of characterization values computed across the training mini-batches, rather than from a single dataset-level scalar. For each characterization metric, the training-derived mean is calculated as

$$\mu_x = \frac{x_1 + x_2 + \cdots x_B}{B}, \tag{14}$$

where $x_1$,$x_2$, ..., $x_B$ denote the scalar values of a particular characterization metric x, computed separately from the first through the B-th training mini-batches, B is the total number of training mini-batches, and $\mu_x$ represents the corresponding training-derived mean.

The standard deviation of each characterization metric across the training mini batches is calculated as

$$\sigma_x = \sqrt{\frac{(x_1 - \mu_x)^2 + (x_2 - \mu_x)^2 + \cdots (x_B - \mu_x)^2}{B}}, \tag{15}$$

where $\sigma_x$ denotes the training-derived standard deviation of metric $x$, quantifying the variability of its scalar values across the training mini-batches. Each mini batch contributes one observation to the reference distribution of the corresponding metric.

Using these training-derived statistics, each characterization value is standardized as

$$z_x = \frac{x - \mu_x}{\sigma_x + \varepsilon}, \tag{16}$$

where $z_x$ denotes the standardized characterization value, x is the original metric value to be normalized, $\mu_x$ and $\sigma_x$ are its training-derived mean and standard deviation, respectively, and $\varepsilon = 10^{-8}$ is a small positive constant introduced to ensure numerical stability when the standard deviation approaches zero.

The standardized characterization value is then transformed into a bounded, direction-adjusted score using the sigmoid function

$$r(x) = \frac{1}{1 + \exp(-s_x z_x)}. \tag{17}$$

where $r(x)$ denotes the normalized characterization score associated with metric x, $z_x$ is its standardized value, $s_x$ is a direction coefficient, and $\exp(\cdot)$ denotes the exponential function. Specifically, $s_x = +1$ for TC, SD, and SR, whereas $s_x = -1$ for SC and MA. The resulting score is bounded between zero and one.

The direction coefficients reflect the intended relationship between the observed scene characteristics and the suitability of different fusion mechanisms. Higher TC indicates greater terrain complexity, higher SD reflects stronger spectral diversity, and higher SR indicates a greater proportion of shadow-affected regions. These three metrics therefore retain their positive standardized directions. Conversely, lower SC indicates weaker spatial continuity and greater structural fragmentation, while lower MA indicates reduced agreement between HSI and LiDAR representations. Accordingly, SC and MA are inversely standardized so that their normalized scores increase as spatial fragmentation or cross-modal disagreement becomes more pronounced.

The normalized indicator scores are mapped to the feedforward priors of the six operators by the fixed mapping

$$\begin{gathered} S_{FF}(\text{base}) = 0.10,\ S_{FF}(\text{ghost}) = 1 - 1/4(r(TC) + r(SD) + r(SR) + r(MA)), \\ S_{FF}(\text{geo}) = 1/3(r(TC) + r(SR) + r(MA)),\ S_{FF}(\text{gate}) = r(MA), \\ S_{FF}(\text{trans}) = 1/2(r(SD) + r(SC)),\ S_{FF}(\text{phys}) = r(SR) \end{gathered} \tag{18}$$

where each score is finally clipped to $[0.05, 0.95]$. Each operator receives the indicators that describe the condition its fusion mechanism is designed to address (Table 1): geometry-aware fusion is favored by strong relief (TC), shadow-affected regions (SR) and weak cross-modal agreement (MA); adaptive gating by weak modality agreement (MA); the cross-modal transformer by high spectral diversity (SD) and fragmented spatial structures (low SC); and physics-informed fusion by a high shadow ratio (SR). Baseline-Ghost receives the complement of the mean complexity score and is therefore preferred when terrain, spectral, illumination and cross-modal complexity are all low, whereas the plain concatenation baseline receives a constant low prior (0.10) and serves only as a fallback. The mapping encodes heuristic design priors; it is not learned and is not claimed to be optimal. The priors only initialize the operator ranking, which is subsequently corrected by validation feedback (Section 3.4).

**Table 1.** Feedforward characterization indicators, their normalization direction $s_x$ (Equation (17)) and the operators whose prior they increase (Equation (18)).

| Signal | Physical Meaning | Data Source | Direction $s_x$ | Operators Whose Prior It Increases (Rationale) |
|---|---|---|---|---|
| TC | Terrain complexity | LiDAR | +1 | Geometry-aware (elevation-guided gating is most useful under strong relief); lowers Baseline-Ghost |
| SD | Spectral diversity | HSI | +1 | Cross-modal transformer (long-range spectral–spatial interaction); lowers Baseline-Ghost |
| SR | Shadow ratio | HSI + LiDAR | +1 | Physics-informed (illumination compensation) and geometry-aware; lowers Baseline-Ghost |
| SC | Spatial correlation (lag-1) | HSI | −1 | Cross-modal transformers (fragmented structures require non-local interaction) |
| MA | Modal agreement | HSI + LiDAR | −1 | Adaptive gating (per-location modality re-weighting) and geometry-aware; lowers Baseline-Ghost |

Importantly, the normalized characterization scores are not direct measures of intrinsic data reliability or probabilities of correct classification. Instead, they serve as direction-adjusted indicators that encode heuristic, data-dependent priors regarding fusion-operator suitability. These scores are used to initialize the feedforward operator priorities, which are subsequently refined through validation-based feedback. Thus, a higher normalized score indicates a stronger characterization signal relevant to operator prioritization, rather than necessarily implying higher data quality.

All normalization statistics are estimated exclusively from the training mini-batches and remain fixed during subsequent operator prioritization. Neither validation nor spatially held-out test samples are used to estimate these statistics, thereby preserving the separation between model development and held-out evaluation. All five indicators are computed on mini batches of $7 \times 7$ training patches, and the feedforward priors are evaluated once, on a training batch, before the closed loop. Because the standardization of Equations (14)–(16) is performed with respect to the distribution of training mini-batches, the resulting scores describe batch-relative characteristics of the training data rather than absolute differences between datasets.

The characterization itself adds negligible cost, because it is computed once on the training mini batches before the closed loop, and the controller updates only six scalar operator scores per iteration. The main computational cost of the search arises from the candidate training (at most two 200-epoch training sessions per iteration and five iterations).

### *3.3. Multimodal Fusion Operator Library*

All fusion operators are implemented using GhostConv-based feature extraction modules. Given modality-specific inputs, embeddings are extracted as

$$F_H = E_H(X^H), F_L = E_L(X^L), \tag{19}$$

where $E_H(\cdot)$ and $E_L(\cdot)$ denote modality encoders.

1. Baseline Fusion

The baseline operator serves as the simplest feature-level fusion strategy and provides a reference for evaluating the effectiveness of more sophisticated adaptive operators.

$$F_{base} = \phi([F^H,F^L]), \tag{20}$$

where $[\cdot]$ denotes channel concatenation and $\phi(\cdot)$ represents convolutional fusion.

2. Baseline-Ghost Fusion

GhostConv replaces redundant feature generation with inexpensive linear transformations, thereby reducing computational redundancy while preserving representative feature responses. The Baseline-Ghost operator employs GhostConv-based feature extractors to generate modality-specific representations with reduced feature redundancy.

$$F_H^g = E_H^g\left(X^H\right), F_L^g = E_L^g\left(X^L\right), \tag{21}$$

where $E_H^g$ and $E_L^g$ denote GhostConv-based feature extractors. The extracted representations are subsequently fused through concatenation,

$$F_{ghost} = \phi\left(\left[F_H^g, F_L^g\right]\right) \tag{22}$$

The resulting features are subsequently concatenated for multimodal representation learning.

3. Geometry-Aware Fusion

Geometry-aware fusion explicitly exploits LiDAR-derived terrain information to guide feature interaction, which is particularly beneficial for scenes exhibiting strong elevation variation.

$$G_{geo} = \sigma(g(X^L)), \tag{23}$$

where $g(\cdot)$ denotes geometry extraction and $\sigma$ is sigmoid activation.

$$F_{geo} = F_{base} + \lambda c_g R(F_{base} \odot G_{geo}), \tag{24}$$

where $c_g$ is a geometry confidence, $\lambda$ controls residual contribution, and $R(\cdot)$ represents refinement.

4. Adaptive Gating Fusion

Adaptive gating combines learnable gating and cosine-based modality consistency to estimate modality reliability. Cosine similarity provides a scale-invariant estimate of cross-modal consistency and is therefore employed as the gating prior.

$$G_{learn} = \psi([F_H, F_L]), \tag{25}$$

where $\psi$ denotes gate prediction.

$$G_{cos} = \frac{CosSim(F_H, F_L) + 1}{2} \tag{26}$$

The final adaptive gate becomes

$$G = \gamma G_{cos} + (1 - \gamma) G_{learn}, \tag{27}$$

where $\gamma$ balances the cosine-similarity prior and learnable gate. Fusion is then performed as

$$F_{gate} = G \odot F_H + (1 - G) \odot F_L. \tag{28}$$

5. Cross-Modal Transformer Fusion

The GhostConv features $F_H^g, F_L^g \in R^{B\times d\times 7\times 7}(d = 256)$ are flattened into token sequences $T_H$,$T_L \in R^{B\times N\times d}$, with N = 49 tokens, i.e., one per spatial position of the patch. They are given by

$$T_H = \text{flatten}\left(F_H^g\right), \quad T_L = \text{flatten}\left(F_L^g\right) \tag{29}$$

and two cross-attention blocks are applied,

$$Z_H = \text{Attn}\left(T_H W_1^Q, T_L W_1^K, T_L W_1^V\right), \quad Z_L = \text{Attn}\left(T_L W_2^Q, T_H W_2^K, T_H W_2^V\right) \tag{30}$$

where Attn$(\cdot)$ denotes multi-head scaled dot-product attention with eight heads $(d_h = 32)$ and attention dropout 0.1,

$$\text{Attn}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}}\right)V. \tag{31}$$

and $W^Q$, $W^K$ and $W^V$ are separate learnable projections for each direction. Thus, in the HIS → LiDAR block, the hyperspectral tokens form the queries, and the LiDAR tokens form the keys and values, and vice versa in the LiDAR → HSI block. No positional encoding is used. The outputs are averaged over the tokens, concatenated and projected by a two-layer MLP $(512 \to 256 \to 256)$:

$$F_{trans} = \text{MLP}([\text{mean}(Z_H), \text{mean}(Z_L)]) \tag{32}$$

6. Physics-Informed Fusion

Physics-informed fusion compensates illumination-induced spectral degradation using LiDAR-derived geometric cues, improving feature consistency in shadowed regions. Under a simplified Lambertian illumination model, the at-sensor radiance of pixel p at wavelength $\lambda$ can be written as

$$L_\lambda(p) \approx \rho_\lambda(p)\left[E_\lambda^{dir} V(p) \cos\theta_i(p) + E_\lambda^{dif} \Psi(p)\right] \tag{33}$$

where $\rho_\lambda$ is the surface reflectance, $E_\lambda^{dir}$ and $E_\lambda^{dif}$ are the direct and diffuse irradiance, $V(p) \in \{0, 1\}$ denotes cast-shadow visibility, $\theta_i(p)$ is the local solar incidence angle and $\Psi(p)$ is the sky-view factor. In shadowed pixels, V = 0 and only the diffuse term remains, which reduces brightness and alters the spectral shape. Because V, $\theta_i$ and $\Psi$ are determined by the surface geometry, the LiDAR elevation model contains the information required to predict where and how strongly the hyperspectral signal is attenuated. The physics-informed operator therefore learns geometry-conditioned spectral correction

$$G = g(X^L), \quad \Delta X^H = C([X^H, G]) \tag{34}$$

$$X_{corr}^H = X^H + \Delta X^H \tag{35}$$

$$F_{phys} = GAP\left(\phi\left(\left[E_H^g\left(X_{corr}^H\right), E_L^g\left(X^L\right)\right]\right)\right) \quad (36)$$

where $g(\cdot)$ consists of two $3 \times 3$ convolutions with ReLU (64 output channels), $C(\cdot)$ consists of two $3 \times 3$ convolutions that output a $C_H$-channel correction, $\phi(\cdot)$ is a $1 \times 1$ convolution with ReLU, and GAP denotes global average pooling. Equation (33) motivates the conditioning variable (LiDAR geometry) and the residual form of the correction, but it is not imposed as a hard constraint; neither solar geometry nor an explicit shadow mask is provided, and $\Delta X^H$ is learned only through classification loss. The operator is therefore physics-motivated rather than physics-constrained, and the term physics-informed is used in this sense. Moreover, the additive residual in Equation (35) is only a local approximation of the multiplicative shading effect in Equation (33).

### *3.4. Hybrid Reliability-Aware Feedback Control*

In this work, reliability refers to the expected suitability of a fusion operator for the current training data, estimated from development region information only. The controller is executed for at most K iterations after a Baseline-Ghost model has been trained; the validation accuracy of this model is denoted by $A_0$. The values of all constants introduced below, including the action priors, are listed in Appendix A Table A3.

1. Adaptive target.

   Before the loop, a reachable target accuracy is derived from $A_0$ and then kept fixed,

$$A_t = \max\left(\tilde{A}_t - \Delta_c(1[TC > \theta_{TC}] + 1[MA < \theta_{MA}]), A_0 + \Delta_{min}\right), \quad (37)$$

   where $\tilde{A}_t = \min(A_0 + \Delta_1, {A_1}^{max})$ if $A_0 < a_1$, $\tilde{A}_t = \min(A_0 + \Delta_2, {A_2}^{max})$ if $a_1 \le A_0 \le a_2$, and $\tilde{A}_t = {A_3}^{max}$ otherwise, if $1[\cdot]$ is the indicator function, and TC and MA are the terrain complexity and modality agreement values of the training batch (Section 3.2). The target thus requests a larger gain when the baseline is weak and is lowered for terrain-complex or weakly agreeing scenes, in which large validation gains are less likely.

2. Error signal.

   At iteration k, the accepted validation accuracy $A_k$ is smoothed and compared with the target

$$\overline{A}_k = \mu_A \overline{A}_{k-1} + (1 - \mu_A)A_k, \ \ \overline{A}_0 = A_0 \quad (38)$$

   During training, operator performance is evaluated using validation accuracy. Let

$$e_k = A_t - \overline{A}_k \quad (39)$$

   Rather than performing the classical control of a physical system, the PID-inspired controller adaptively adjusts operator priorities according to validation performance.

3. PID-inspired signal and control action. With an anti-windup bound $I_{max}$ on the integral term,

$$I_k = \text{clip}(I_{k-1} + e_k, -I_{max}, I_{max}), \quad (40)$$

$$u_k = K_p e_k + K_i I_k + K_d(e_k - e_{k-1}). \quad (41)$$

   The signal is mapped to a discrete control action

$$a_k = \begin{cases} \text{enable-complex}, & u_k > \theta_1 \\ \text{enable-simple}, & \theta_2 < u_k \le \theta_1 \\ \text{maintain}, & |u_k| \le \theta_2 \\ \text{disable}, & u_k < -\theta_2 \end{cases} \quad (42)$$

and each action is associated with a fixed operator-specific prior vector $\pi_a(m)$. The magnitude of $u_k$ also sets the gain of the feedback-guided residual refinement used when the candidates of iteration k are trained,

$$\rho_k = \mathrm{clip}(|u_k|, \rho_{min}, \rho_{max}). \quad (43)$$

Thus, a large error (validation accuracy far below the target) results in stronger refinement during candidate training, whereas a small error results in weaker refinement.

4. Operator reward and reliability. When operator m is evaluated at iteration k, its reward is

$$r_m^{(k)} = (1 - \lambda_c)A_m^{(k)} + \lambda_c C_m^{(k)} - \eta\sigma_m^{(k)}, \quad (44)$$

where $A_m^{(k)}$ is the best validation accuracy of the candidate, $C_m^{(k)}$ is its mean maximum SoftMax probability on the validation samples, and $\sigma_m^{(k)}$ is the standard deviation of the previous rewards of operator m (zero when fewer than two are available). The reward is smoothed over successive evaluations of the same operator,

$$\tilde{r}_m^{(k)} = \mu_r \tilde{r}_m^{(k-1)} + (1 - \mu_r) r_m^{(k)}, \quad (45)$$

and the reliability of operator m is obtained from its smoothed reward history $H_m = \{\tilde{r}_m^{(1)}, \ldots, \tilde{r}_m^{(n)}\}$, as

$$R_m = \mathrm{sigmoid}\Big(k\Big[\Big(\tilde{r}_m^{(n)} - \mathrm{mean}\Big(\tilde{r}_m^{(1:n-1)}\Big)\Big) - \eta\,\mathrm{std}(H_m)\Big]\Big), n \geq 2, \quad (46)$$

with $R_m = 0.5$ for operators evaluated fewer than twice. Operators whose reward is increasing and stable therefore receive higher reliability; with the small number of evaluations per operator and the smoothing of Equation (45), $R_m$ remains close to 0.5 in practice (Section 4.7), so that $S_{FB}$ is dominated by the action priors. The feedback score combines reliability with the action prior and is normalized over the operator library $\Omega$,

$$S_{FB}(m) = \frac{R_m \pi_{a_k}(m)}{\sum_{j\in\Omega} R_j \pi_{a_k}(j)}. \quad (47)$$

5. Hybrid score and operator selection. The feedforward scores of Equation (18) are normalized in the same way, $\bar{S}_{FF}(m) = S_{FF}(m)/\sum_{j\in\Omega} S_{FF}(j)$, and combined with the feedback scores using fixed convex weights,

$$S(m) = \alpha\bar{S}_{FF}(m) + \beta S_{FB}(m), \alpha + \beta = 1. \quad (48)$$

Because both terms sum to one over the six operators, $\alpha$ and $\beta$ directly set their relative influence; their sensitivity, including equal weighting. At each iteration, the highest-ranked operator is selected together with the second-ranked operator with probability $1 - \varepsilon_g$, or a randomly chosen lower-ranked operator otherwise; equal scores are ranked in the fixed order of the operator library. An operator whose validation accuracy falls more than $\Delta_s$ below the current accepted accuracy in $n_s$ consecutive evaluations is suppressed for the next $T_s$ iterations; if fewer than two candidates remain, the list is completed from a fixed fallback order.

6. Candidate training, acceptance, and termination.

Each candidate is initialized from the currently accepted model (so that operators not yet accepted start from previously untrained branches) and trained for $E_c$ epochs with refinement gain $\rho_k$ and a learning rate of $\eta_0 \gamma_{lr}^{k-1}$, so that later iterations perform progressively finer adjustments. Validation accuracy is evaluated every 10 epochs, and the

best-validation checkpoint of each candidate is retained. Within an iteration, candidates are compared using $\tilde{r}_m^{(k)} - w_v \sigma^{val}{}_m$, where $\sigma^{val}{}_m$ is the standard deviation of the candidate's last validation accuracies. The best candidate is accepted only if

$$A_{new} > A_{best} + \delta. \tag{49}$$

Otherwise, the previously accepted model is restored. The loop terminates when $A_{best} \geq A_t$ or after $K$ iterations. No gradient-based convergence criterion is used; each candidate is trained for a fixed number of epochs with best-validation checkpointing.

7. Reliability-weighted inference.

After the search, the operators whose best validation accuracy lies within $\Delta_e$ of $A_{best}$ are retained (the plain concatenation baseline is excluded, and the finally accepted operator is always included), each with the score $s_m = \omega \max(A_m) - (1-\omega)\mathrm{std}(A_m)$ computed from its validation history. At most, the M highest-scoring operators form the set $\mathcal{M}$, with weights

$$w_m = \frac{\exp(s_m/\tau)}{\sum_{j\in\mathcal{M}} \exp\left(s_j/\tau\right)}, \tag{50}$$

and the class probabilities at inference are $p(y|x) = \sum_{m\in\mathcal{M}} w_m \mathrm{SoftMax}(f_m(x))$, where $f_m$ denotes the classifier output obtained with operator m. If only one operator satisfies the criterion, the single accepted operator is used. It is disabled in the component-wise ablation so that the contribution of each controller component can be isolated.

8. Constant and practical stability.

All constants of the controller (Appendix A Table A3) were fixed before the experiments and are identical for the three datasets; none was tuned on held-out test data. The most influential ones were examined in the sensitivity analysis, whereas the remaining smoothing factors, reward weights, reliability scale, ensemble temperature, and suppression rule are heuristic design choices that were not optimized. The controller operates on bounded scores rather than on physical system dynamics; therefore, no formal Lyapunov stability guarantee is claimed. Bounded behavior is promoted by the integral clipping in Equation (40), the bounded gain in Equation (43), the bounded and normalized scores in Equations (46)–(48), the acceptance margin in Equation (49), the decreasing candidate learning rate, and the iteration limit K. The closed loop operates between training runs: the controller quantities are computed from validation accuracies after each candidate has been trained and are not differentiated, so gradients are never propagated across controller iterations, and the refinement is a single residual correction with a bounded gain rather than an iterated feature update. By Equation (49), the accepted validation accuracy is non-decreasing over the iterations.

The complete closed-loop training procedure is summarized in Algorithm 1, where feedforward characterization initializes operator priors and validation-driven feedback iteratively refines operator priorities.

**Algorithm 1.** Closed-Loop Reliability-Aware Fusion Optimization

Input: Training dataset $D_{train}$, validation set $D_{val}$, operator library $\Omega$, Baseline-Ghost model with validation accuracy $A_0$
Output: final model $M^*$ with reliability-weighted inference
Parameters: $K_p$, $K_i$, $K_d$, $\alpha$, $\beta$, $\varepsilon_g$, $\Delta$, K
Compute TC,SD,SR,SC,MA on the training mini-batches and estimate $\mu_x$, $\sigma_x$
Compute $r_x$ and $S_{FF}(m)$ on the training batches; normalize to $\overline{S}_{FF}(m)$
Set the target $A_t$; $A_{best} \leftarrow A_0$; $M^* \leftarrow$ Baseline-Ghost model
for k = 1, ..., K do

**Algorithm 1.** *Cont.*

Update $\overline{A}_k$ and $e_k$; compute $u_k$
Determine the action $a_k$ and the refinement gain $\rho_k$
Compute $r_m$, $S_{FB}(m)$ and $S(m)$
Select the top-ranked and one further operator ($\varepsilon_g$-greedy); remove suppressed operators
for each candidate m: initialize from M*, train for 200 epochs with learning rate $10^{-4} \cdot 0.5^{k-1}$, and compute its reward
if $A_{new} > A_{best} + \Delta$ then accept: M* ← candidate, $A_{best} \leftarrow A_{new}$; else restore M*
if $A_{best} \geq A_t$ then break
end for
Build the reliability-weighted ensemble and return M*

## 4. Experimental Results and Analysis

### *4.1. Datasets Description*

Three publicly available HSI–LiDAR benchmark datasets were used for evaluation: Houston2013, Muufl, and Augsburg. These datasets exhibit substantially different spectral resolutions, spatial resolutions, terrain characteristics, and scene complexities, providing a comprehensive benchmark for evaluating cross-dataset robustness. Their main characteristics are summarized in Table 2. The Augsburg dataset is used at the 30 m ground sampling distance at which it is distributed; together with Houston2013 (2.5 m) and Muufl (1.0 m), it allows the framework to be evaluated at both fine and coarse spatial resolutions, the latter with more mixed pixels and less structural detail in the elevation data.

**Table 2.** Summary of benchmark datasets used for experimental evaluation.

| Dataset | Size | Bands | Classes | Resolution | Labeled Pixels |
|---|---|---|---|---|---|
| Houston2013 | 349 × 1905 | 144 | 15 | 2.5 m | 15,029 |
| Muufl | 325 × 220 | 64 | 11 | 1.0 m | 53,687 |
| Augsburg | 332 × 485 | 180 | 7 | 30 m | 78,294 |

### *4.2. Spatially Disjoint Sample-Center Evaluation Protocol*

To ensure a consistent and reproducible evaluation, all experiments were conducted using a unified spatially disjoint sample-center protocol. In this study, spatial disjointness refers specifically to the separation of development and held-out test sample centers across non-overlapping spatial blocks, rather than complete disjointness of input patch footprints. Unlike conventional random pixel sampling, the protocol uses a fixed 60/40 spatial block partition, with training and validation sample centers restricted to development blocks and test sample centers restricted to held-out blocks, as summarized in Table 3.

**Table 3.** Statistics of the spatially disjoint benchmark protocol used in all experiments.

| Dataset | Total Labeled | Dev Ratio | Test Ratio | Train Samples | Validation Samples | Held-Out Test Samples |
|---|---|---|---|---|---|---|
| Houston2013 | 15,029 | 0.6 | 0.4 | 105 | 105 | 6139 |
| Muufl | 53,687 | 0.6 | 0.4 | 77 | 77 | 21,588 |
| Augsburg | 78,294 | 0.6 | 0.4 | 49 | 49 | 31,064 |

Only the development region is used during model development and label-based decision-making. Specifically, seven labeled samples per class are randomly selected for training and another seven samples per class are randomly selected for validation from the

remaining development pixels. The held-out test labels are excluded from model training, hyperparameter tuning, controller initialization, operator prioritization, and model selection.

For reproducibility, all experiments employ matched spatial partitions across competing fusion operators, with random sampling repeated using ten predefined seeds (42–51). Consequently, the compared methods are evaluated under consistent training, validation, and testing conditions for each seed.

Both the feedforward characterization signals and the feedback controller are derived from the development region information. However, because 7 × 7 HSI-LiDAR patches are extracted from the full scenes without boundary masking, patches centered near development–test block boundaries may include contextual image pixels from the adjacent partition. Accordingly, the protocol ensures spatial disjointness at the sample-center level but does not guarantee the complete separation of patch footprints. The extent of this overlap and its relationship to classification performance are examined in Section 4.6.5.

Figure 2 illustrates the partition. For each seed (42–51), the scene is divided into a regular grid of blocks (6 × 6 for Houston2013 and Augsburg; 8 × 8 for Muufl), and 60% of the blocks are randomly assigned to the development region and 40% to the held-out region. The same partition and sample centers are used for all compared methods with the same seed. Seven training and seven validation samples per class are drawn at random from the development blocks; consequently, validation samples are spatially disjoint from the test samples but not from the training samples, and validation accuracy is an optimistic estimate of held-out accuracy. Because the partition is seed-specific, the number of held-out test samples varies between seeds (Houston2013, 5266–6901; Muufl, 20,112–22,533; Augsburg, 29,016–34,338); Table 3 reports the averages over the ten seeds.

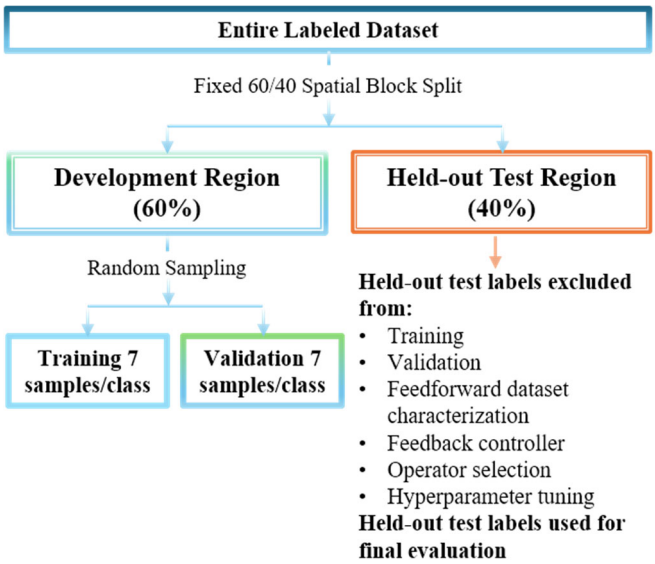



**Figure 2.** Spatially disjoint sample-center evaluation protocol. Development and held-out test sample centers are assigned to separate spatial blocks, while input patch footprints may cross block boundaries.

### *4.3. Implementation Setting Details*

All experiments were implemented in PyTorch 2.8.0 and executed on a workstation equipped with an Intel Core i5-14600KF CPU, an NVIDIA RTX 5070 GPU (12 GB, NVIDIA Corp., Santa Clara, CA, USA), and 32 GB RAM. Models were trained for 200 epochs using the AdamW optimizer with an initial learning rate of $10^{-4}$, weight decay of $5 \times 10^{-4}$, a cosine annealing scheduler, a batch size of 32, and label smoothing of 0.05. Dropout rates of 0.5 (classifier), 0.2 (refinement module), and 0.1 (attention) are used. Validation accuracy is evaluated every 10 epochs, and the best-validation checkpoint is retained. Each controller

candidate is trained for 200 epochs with a learning rate that is halved at every controller iteration (Section 3.4).

All methods used the same training, validation, and test samples for each seed. Validation accuracy was evaluated after every epoch for the competing methods and every 10 epochs for RAFC-Fusion, and the checkpoint with the highest validation OA was used for testing. Budget-matched comparisons are reported in Sections 4.6.4 and 4.7. To ensure reproducibility, experiments were repeated using ten official spatially disjoint seeds (42–51).

### *4.4. Quantitative Comparison*

Tables 4–6 compare the proposed RAFC-Fusion framework with six established HSI–LiDAR classification methods on the Houston2013, Muufl, and Augsburg datasets, respectively, under the spatially disjoint sample-center evaluation protocol. The reported metrics include Overall Accuracy (OA), Average Accuracy (AA), Cohen's Kappa (×100), and macro-averaged F1-score (Macro-F1). All results are presented as mean ± standard deviation over ten experimental runs (seeds 42–51). Macro-F1 is included to complement OA by evaluating the balance between precision and recall across land-cover classes, with equal weight assigned to each class. In addition to HybridSN, CCR-Net, MFT and CALC, they include two recent adaptive fusion methods, NCGLF2 and MERGE-Net, which were adapted to the same splits from their released code (Section 4.3). The competing methods were selected for their methodological relevance, reproducibility and applicability to multimodal HSI–LiDAR classification under the same evaluation protocol.

**Table 4.** Quantitative comparison of HSI–LiDAR classification performance on Houston2013 under the spatially disjoint sample-center evaluation protocol. Results are reported as mean ± standard deviation over ten runs. The highest mean value for each metric is shown in bold.

| **Methods** | **Houston2013** | | | |
|---|---|---|---|---|
| | **OA (%)** | **AA (%)** | **Kappa × 100** | **Macro-F1 (%)** |
| HybridSN [22] | 41.60 ± 6.36 | 45.00 ± 6.68 | 37.77 ± 6.88 | 40.09 ± 7.35 |
| CCR-Net [57] | 59.55 ± 5.12 | 62.68 ± 3.63 | 56.30 ± 5.51 | 59.03 ± 5.27 |
| MFT [58] | 68.31 ± 7.53 | 71.20 ± 5.86 | 65.59 ± 8.35 | 68.11 ± 5.29 |
| CALC [37] | 68.07 ± 6.44 | 71.65 ± 4.73 | 65.43 ± 6.99 | 68.31 ± 4.94 |
| MERGE-Net [39] | 62.66 ± 5.81 | 65.51 ± 5.36 | 59.54 ± 6.40 | 62.80 ± 5.82 |
| NCGLF2 [40] | 65.67 ± 6.01 | 69.66 ± 5.40 | 62.79 ± 6.60 | 66.54 ± 5.40 |
| Proposed | **70.00 ± 6.64** | **74.07 ± 5.10** | **67.59 ± 7.15** | **70.17 ± 6.50** |

**Table 5.** Quantitative comparison of HSI–LiDAR classification performance on Muufl under the spatially disjoint sample-center evaluation protocol. Results are reported as mean ± standard deviation over ten runs. The highest mean value for each metric is shown in bold.

| **Methods** | **Muufl** | | | |
|---|---|---|---|---|
| | **OA (%)** | **AA (%)** | **Kappa × 100** | **Macro-F1 (%)** |
| HybridSN [22] | 58.67 ± 4.47 | 54.00 ± 4.41 | 48.02 ± 3.59 | 39.42 ± 3.31 |
| CCR-Net [57] | 49.08 ± 5.07 | 52.41 ± 7.04 | 38.69 ± 5.40 | 36.27 ± 4.85 |
| MFT [58] | 68.49 ± 4.23 | 61.26 ± 3.81 | 59.35 ± 5.03 | 46.11 ± 4.64 |
| CALC [37] | 70.82 ± 2.60 | 65.60 ± 3.98 | 62.36 ± 3.27 | 49.22 ± 3.86 |
| MERGE-Net [39] | 58.24 ± 6.73 | 50.03 ± 7.02 | 47.29 ± 5.78 | 39.84 ± 4.75 |
| NCGLF2 [40] | 66.10 ± 7.19 | 59.00 ± 7.18 | 56.96 ± 6.82 | 46.54 ± 7.58 |
| Proposed | **73.20 ± 5.45** | **68.34 ± 3.47** | **65.36 ± 5.77** | **52.14 ± 4.59** |

As shown in Table 4, the proposed framework achieves the highest mean performance in Houston2013, with an OA of 70.00 ± 6.64%, AA of 74.07 ± 5.10%, Kappa of 67.59 ± 7.15,

and Macro-F1 of 70.17 ± 6.50%. On this dataset, MFT is the strongest competing method in terms of mean OA (68.31%) and Kappa (65.59), whereas CALC provides the highest competing AA (71.65%) and Macro-F1 (68.31%). Relative to the strongest competing value of each metric, the proposed framework improves OA by 1.69, AA by 2.42, Kappa by 2.00 and Macro-F1 by 1.86 percentage points; these differences are small compared with the run-to-run standard deviations (5–8 percentage points) and are obtained with a larger training budget (see the end of this section). The relatively large variability observed for HybridSN indicates substantial run-to-run instability in this evaluation.

**Table 6.** Quantitative comparison of HSI–LiDAR classification performance on Augsburg under the spatially disjoint sample-center evaluation protocol. Results are reported as mean ± standard deviation over ten runs. The highest mean value for each metric is shown in bold.

| Methods | Augsburg | | | |
|---|---|---|---|---|
| | **OA (%)** | **AA (%)** | **Kappa × 100** | **Macro-F1 (%)** |
| HybridSN [22] | 51.46 ± 7.02 | 52.47 ± 3.64 | 39.28 ± 7.12 | 37.41 ± 3.85 |
| CCR-Net [57] | 40.49 ± 7.74 | 44.69 ± 4.07 | 28.41 ± 7.67 | 30.58 ± 3.90 |
| MFT [58] | 58.03 ± 10.97 | 55.93 ± 4.49 | 46.23 ± 12.00 | 42.46 ± 6.20 |
| CALC [37] | 59.70 ± 9.46 | 58.28 ± 4.58 | 48.90 ± 9.12 | 43.70 ± 6.42 |
| MERGE-Net [39] | 56.25 ± 13.83 | 53.64 ± 5.78 | 45.73 ± 13.98 | 40.34 ± 9.27 |
| NCGLF2 [40] | 64.72 ± 10.88 | 60.60 ± 4.63 | 55.31 ± 11.68 | 47.51 ± 6.25 |
| Proposed | **66.34 ± 6.73** | **62.09 ± 5.46** | **56.33 ± 7.99** | **48.15 ± 4.36** |

For the Muufl dataset (Table 5), the proposed framework achieves an OA of 73.20 ± 5.45%, AA of 68.34 ± 3.47%, Kappa of 65.36 ± 5.77, and Macro-F1 of 52.14 ± 4.59%. Compared with CALC, which provides the strongest competing mean OA and Macro-F1, the proposed method improves these metrics by 2.38 and 2.92 percentage points, respectively. The improvement in Macro-F1 indicates that the overall classification gain is accompanied by better average precision–recall performance across the 11 land-cover classes.

On Augsburg (Table 6), the proposed framework achieves an OA of 66.34 ± 6.73%, an AA of 62.09 ± 5.46%, a Kappa of 56.33 ± 7.99, and a Macro-F1 of 48.15 ± 4.36%. Among the competing methods, the recent NCGLF2 achieves the highest values of all four metrics (64.72%, 60.60%, 55.31 and 47.51%); the proposed framework exceeds them by 1.62, 1.49, 1.02 and 0.64 percentage points, respectively, which is small compared with the run-to-run standard deviations, and it exceeds CALC by 6.64 percentage points in OA. On Augsburg, the proposed framework therefore performs comparably to the strongest recent method.

Overall, the proposed RAFC-Fusion framework achieves the highest mean OA, AA, Kappa, and Macro-F1 across all three datasets. The consistent improvement in Macro-F1 provides complementary evidence that the observed gains are not limited to overall classification accuracy but also extend to class-balanced precision–recall performance. Nevertheless, the reported standard deviations indicate variability across experimental runs, and differences in mean performance alone should not be interpreted as evidence of statistical significance.

Figures 3–5 further demonstrate the qualitative advantages of the proposed framework on Houston2013, Muufl, and Augsburg, respectively. Across all three datasets, the proposed method produces more spatially coherent classification regions, preserves object boundaries more consistently, and exhibits fewer isolated misclassified pixels within the held-out test blocks (red outlines) than the competing methods.

On Houston2013, the proposed framework better maintains the continuity of elongated road and railway structures. On Muufl, it is better at suppressing salt-and-pepper artifacts in fragmented vegetation and urban areas while preserving fine object boundaries. The proposed method leads to more homogeneous land-cover regions with less fragmented

predictions, especially in forest, residential and low-vegetation areas for Augsburg. These qualitative observations are consistent with the quantitative improvements reported in Tables 4–6, indicating improved cross-region generalization under the spatially disjoint sample-center evaluation protocol.

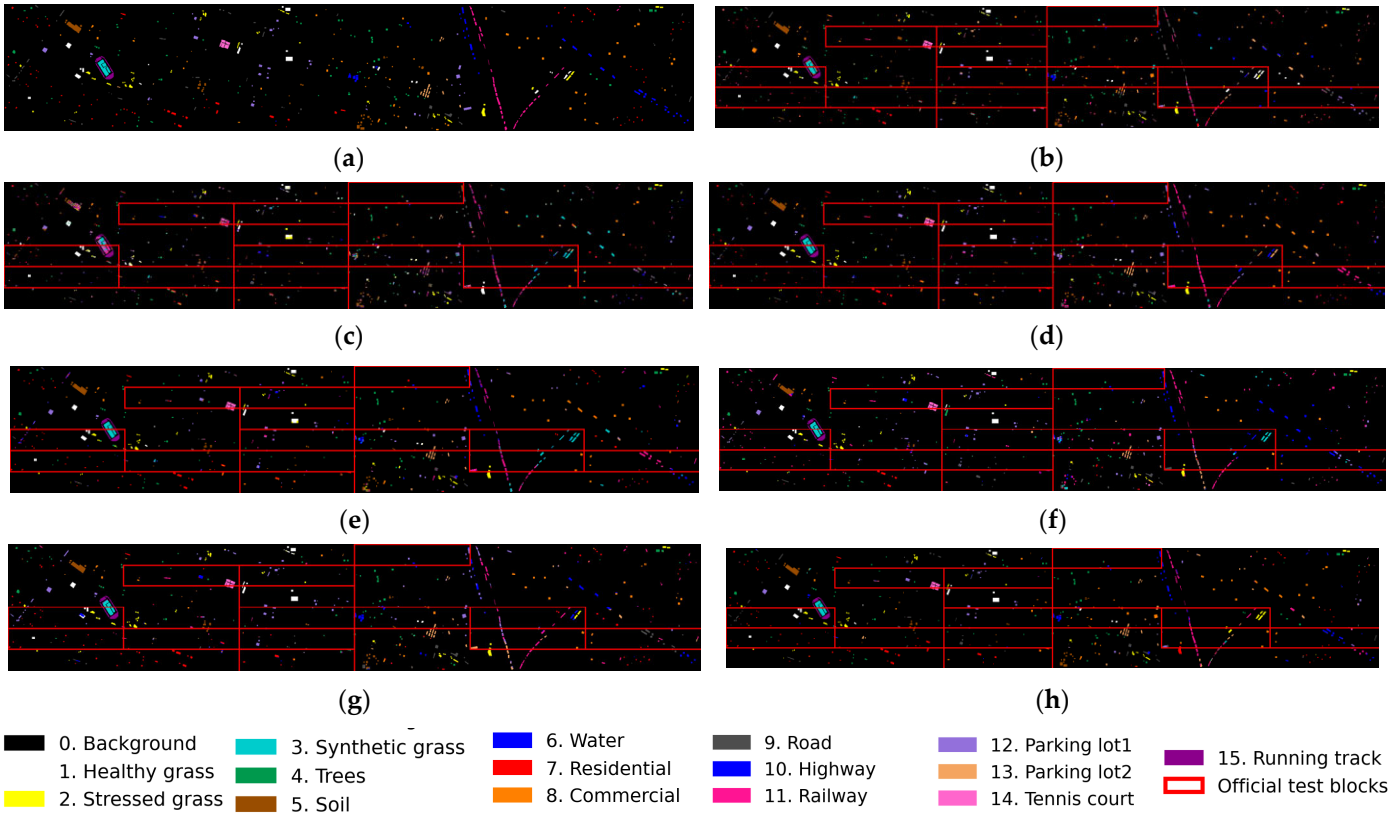



**Figure 3.** Classification maps of Houston2013. (**a**) Ground-truth; (**b**) HybridSN; (**c**) CCR-Net; (**d**) MFT; (**e**) CALC; (**f**) MERGE-Net; (**g**) NCGLF2; (**h**) Proposed. Held-out test blocks are outlined in red.

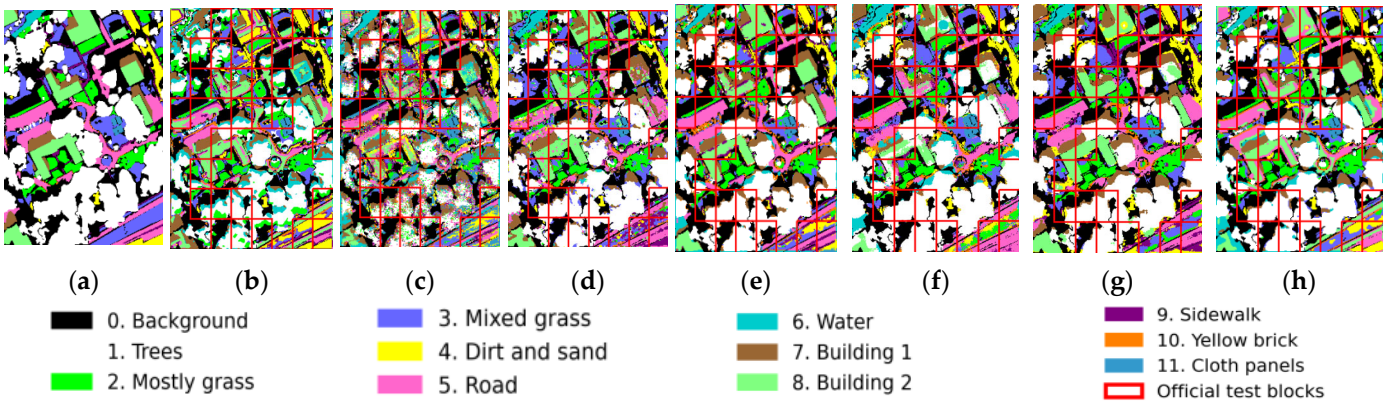



**Figure 4.** Classification maps of Muufl. (**a**) Ground-truth; (**b**) HybridSN; (**c**) CCR-Net; (**d**) MFT; (**e**) CALC; (**f**) MERGE-Net; (**g**) NCGLF2; (**h**) Proposed. Held-out test blocks are outlined in red.

Finally, although relatively large standard deviations are observed for all methods, this behavior is expected under extremely limited supervision (seven training samples per class) and the sample-center strict spatially disjoint protocol. The robustness of the proposed framework is further analyzed through multi-seed experiments and statistical analysis in Table 7.

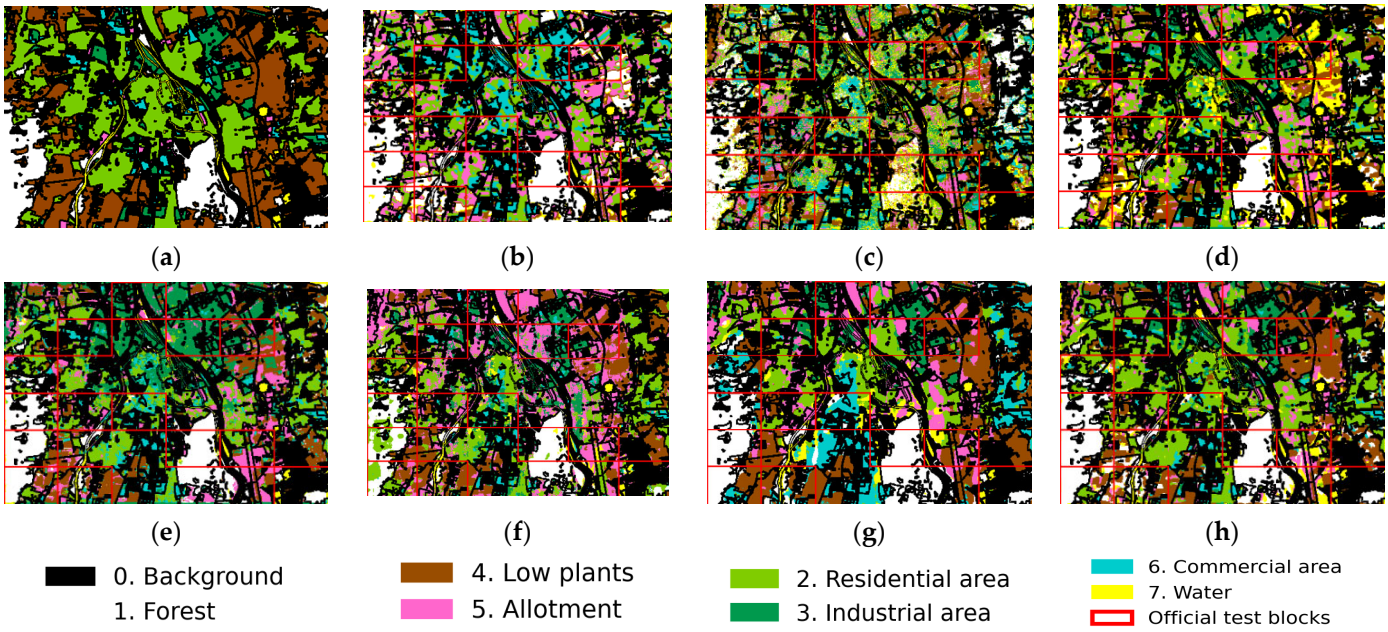



**Figure 5.** Classification maps of Augsburg. (**a**) Ground-truth; (**b**) HybridSN; (**c**) CCR-Net; (**d**) MFT; (**e**) CALC; (**f**) MERGE-Net; (**g**) NCGLF2; (**h**) Proposed. Held-out test blocks are outlined in red.

**Table 7.** Multi-seed robustness evaluation.

| Dataset | OA (%) | Val-Test Gap (%) | Positive-Gain Runs |
|---|---|---|---|
| Houston2013 | 70.00 ± 6.64 | 16.57 ± 7.15 | 10/10 |
| Muufl | 73.20 ± 5.45 | 2.13 ± 4.80 | 9/10 |
| Augsburg | 66.34 ± 6.73 | 3.72 ± 5.11 | 8/10 |

### *4.5. Robustness Evaluation*

Table 7 summarizes the robustness of the proposed framework over ten predefined spatially disjoint runs. Three complementary indicators are reported: (1) the average overall accuracy (OA), (2) the validation–test performance gap, and (3) the positive gain ratio, which represents the number of runs in which the feedback-controlled optimization achieved a higher final test OA than the corresponding baseline initialization.

Although only seven validation samples per class are used, the proposed framework demonstrates stable behavior across all datasets. On Houston2013, Muufl, and Augsburg, the controller achieves positive improvements in 10/10, 9/10, and 8/10 runs, respectively. These results indicate that the closed-loop procedure improves on its baseline initialization in most runs, rather than relying on favorable random validation subsets.

The smaller gaps on Muufl and Augsburg suggest that validation feedback remains reasonably representative on these datasets, whereas the substantially larger Houston2013 gap highlights the limitation of validation-guided adaptation under stronger spatial distribution shift.

Overall, the consistent positive gain ratios indicate that the closed-loop procedure does not depend on a particular validation split, whereas the large validation–test gap on Houston2013 shows that its validation signal can be optimistic under strong spatial distribution shift (Section 5.3).

These multi-seed results indicate that the closed-loop procedure improves on its Baseline-Ghost initialization consistently across different validation subsets rather than only for favorable validation sampling; the sources of this improvement are analyzed in Section 4.7.

### *4.6. Ablation Results*

#### 4.6.1. Controller Parameter Sensitivity

Table 8 investigates the sensitivity of the proposed framework to representative controller parameter configurations across the three benchmark datasets. The configuration used in the main experiments ($K_p = 0.4$, $K_i = 0.05$, $K_d = 0.3$, $\alpha = 0.8$, $\beta = 0.2$, $\varepsilon_g = 0.05$) is identified as the reference configuration and highlighted in bold. The same reference configuration is used across all three datasets rather than applying dataset-specific controller settings. The weights $\alpha$ and $\beta$ are applied to the normalized feedforward and feedback scores of Equation (48), and the first row of Table 8 ($\alpha = \beta = 0.5$) corresponds to equal weighting.

**Table 8.** Controller parameter sensitivity across three benchmark datasets. Values denote held-out test OA (%). The bold row indicates the reference configuration used in the main experiments; candidate configurations are ranked separately using validation-based stability scores.

| $K_p$ | $K_i$ | $K_d$ | $\alpha$ | $\beta$ | $\varepsilon$ | Houston2013 | Muufl | Augsburg |
|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.1 | 0.2 | 0.5 | 0.5 | 0.1 | 64.32 | 69.47 | 70.03 |
| 0.4 | 0.05 | 0.3 | 0.7 | 0.3 | 0.05 | 65.09 | 70.14 | 71.40 |
| **0.4** | **0.05** | **0.3** | **0.8** | **0.2** | **0.05** | **64.54** | **70.35** | **71.68** |
| 0.3 | 0.03 | 0.4 | 0.8 | 0.2 | 0.0 | 64.43 | 69.73 | 72.66 |
| 0.3 | 0.03 | 0.4 | 0.9 | 0.1 | 0.0 | 65.93 | 70.39 | 71.38 |

In the controller parameter sensitivity experiment, candidate configurations are ranked using a validation-based stability criterion, defined as the development-region validation accuracy minus 0.5 times its standard deviation across repeated runs. This criterion considers both predictive performance and variability without using held-out test OA for the ranking. Importantly, the highest-ranked configuration under this criterion is not identical across datasets. Therefore, the reference configuration should not be interpreted as the individually validation-optimal setting for every benchmark.

The held-out test OA values in Table 8 are reported separately to characterize the sensitivity of classification performance to controller parameter variations. The results show that several configurations yield comparable performance, with different settings achieving the highest test OA on different datasets. The reference configuration achieves test OA values of 64.54%, 70.35%, and 71.68% on Houston2013, Muufl, and Augsburg, respectively. These findings indicate that the controller can operate under a common parameter configuration, while its sensitivity and the relative ranking of alternative configurations remain dataset-dependent. Values obtained under this setting are not directly comparable with Tables 4–6.

#### 4.6.2. Patch Size Analysis

Table 9 studies the sensitivity of the proposed framework to spatial neighborhood size across the benchmark datasets. Smaller patches preserve local spatial information, whereas larger neighborhoods provide broader contextual information but may introduce additional spatial heterogeneity.

The 7 × 7 patch size, used as the reference configuration in the main experiments, achieves competitive performance across all three datasets, with OA values of 69.49%, 74.42%, and 70.41% on Houston2013, Muufl, and Augsburg, respectively. Although the 9 × 9 patch achieves a slightly higher OA on Houston2013 (70.61%), larger patches do not consistently improve classification performance across the benchmarks. The results in Table 7 are reported as a sensitivity assessment of spatial neighborhood size under the spatially disjoint sample-center evaluation protocol, rather than as a test-OA-based hyperparameter selection criterion. These values were obtained in separate sensitivity runs and are not directly comparable with Tables 4–6.

**Table 9.** Sensitivity of classification performance to spatial neighborhood size across three benchmark datasets. Values denote held-out test OA (%); the 7 × 7 patch is the reference configuration used in the main experiments. The reference configuration in the main experiments for each metric is shown in bold.

| Patch Size | Houston2013 | Muufl | Augsburg |
|---|---|---|---|
| **7 × 7** | **69.49** | **74.42** | **70.41** |
| 9 × 9 | 70.61 | 69.26 | 64.94 |
| 11 × 11 | 64.89 | 71.92 | 68.16 |
| 13 × 13 | 61.68 | 70.79 | 66.66 |
| 15 × 15 | 61.35 | 65.87 | 66.81 |

#### 4.6.3. Fusion Operator Contribution

Table 10 compares the proposed framework with individual fusion operators. Across ten seeds, the best individual operator differs between datasets (physics-informed on Houston2013 and Muufl; Baseline-Ghost on Augsburg), and the ranking of the remaining operators changes substantially. RAFC-Fusion attains accuracy comparable to that of the best fixed operator on each dataset (differences of −0.04, +1.71 and −0.45 percentage points; $p \geq 0.33$) without prior knowledge of which operator is best.

**Table 10.** Fusion operator contribution analysis (test OA, %). Ten seeds (42–51), mean ± standard deviation: best fixed operator of each dataset, trained individually with the same refinement module, and RAFC-Fusion re-run in the same session (ensemble disabled).

| Test OA (%) | Houston2013 | Muufl | Augsburg |
|---|---|---|---|
| Best fixed operator (per dataset) | 69.57 ± 6.78 (Physics) | 72.79 ± 4.67 (Physics) | 65.73 ± 7.41 (Ghost) |
| RAFC-Fusion (Full, re-run) | 69.53 ± 5.29 | 74.50 ± 4.82 | 65.28 ± 7.36 |
| Difference (paired *t*-test) | −0.04 ($p = 0.978$) | +1.71 ($p = 0.333$) | −0.45 ($p = 0.863$) |

Random operator selection yields the worst performance on all three datasets in the single-run comparison (Table 11), which shows that the observed improvements are not merely caused by increased model diversity. In this comparison, the proposed framework achieves the highest overall accuracy; over ten seeds it matches the dataset-specific best fixed operator (Table 10), and the budget-matched analysis of Section 4.7 shows that its improvement over a fixed operator stems mainly from validation-gated closed-loop training and feedback-guided refinement.

**Table 11.** Single run of all operators and of random operator selection. The highest mean value for each metric is shown in bold.

| Fusion Operators | Houston2013 | Muufl | Augsburg |
|---|---|---|---|
| Baseline concatenation | 56.39 | 66.61 | 54.04 |
| Baseline Ghost | 60.70 | 68.87 | 60.29 |
| Geometry-Aware | 53.91 | 50.75 | 44.83 |
| Adaptive Gates | 56.75 | 56.18 | 61.21 |
| Cross-Modal Transformer | 61.79 | 64.37 | 51.31 |
| Physics-Informed | 66.96 | 70.89 | 61.85 |
| Random selection | 47.21 | 27.38 | 44.83 |
| **Ours (Full)** | **69.04** | **74.31** | **70.47** |

Notably, different operators become optimal on different datasets, further motivating adaptive selection instead of fixed fusion architecture. The ablation results indicate that the proposed framework does not rely on a universally superior fusion operator; instead,

validation-gated adaptation reaches the accuracy of the best operator of each dataset under sample-center spatially disjoint evaluation.

#### 4.6.4. Component-Wise Ablation and Controller Decision Analysis

To examine whether the controller provides benefits beyond conventional validation-based model selection, three simpler strategies were evaluated on the same splits, seeds, encoders, operator library, classifier head, and training schedule, with the residual refinement module (fixed gain 0.10) enabled in all of them: (a) direct validation-based selection, in which each of the six operators is trained independently for 200 epochs and the operator with the highest validation OA is retained; (b) budget-matched direct selection, in which operators are drawn in a seeded random order until the number of 200-epoch trainings equals that of RAFC-Fusion on the same seed; and (c) adaptive weighting, in which the pooled features of all six operators are combined with learnable SoftMax v4.6 weights trained end-to-end on the training loss.

All strategies, including RAFC-Fusion, were run in a separate paired experiment that shares the splits, seeds, and Baseline-Ghost initializations; the RAFC-Fusion values therefore differ slightly from Tables 4–6 (by −0.47, +1.30, and −1.06 percentage points), whereas the Baseline-Ghost values are identical.

Training—number of 200-epoch training courses, including the Baseline-Ghost initialization; time—total training wall-clock time on an RTX 5070 (NVIDIA, Santa Clara, CA, USA). The lower block gives the paired differences.

In Table 12 and Figure 6, under a matched training budget, RAFC-Fusion yields mean OA differences of +1.65, +3.11 and +1.06 percentage points relative to direct selection on Houston2013, Muufl and Augsburg, respectively, none of which is statistically significant ($p = 0.28$, 0.16 and 0.59). Adaptive weighting, whose weights are learned from the training loss, achieves lower OA on all datasets (66.41%, 63.97% and 59.28%). These results indicate that the benefit of RAFC-Fusion over a fixed fusion operator derives mainly from validation-gated closed-loop training (Section 4.7), and that the specific controller design yields comparable, but not significantly higher, accuracy than direct validation-based selection under the evaluated low-sample conditions.

**Table 12.** Comparison with simpler operator-selection strategies (test OA, mean ± standard deviation over ten seeds; independent paired run).

| Strategy (Test OA, %) | Houston2013 | Muufl | Augsburg | Training (H/M/A) | Time, s (H/M/A) |
|---|---|---|---|---|---|
| Baseline-Ghost (fixed) | 61.23 ± 5.76 | 67.07 ± 6.31 | 61.66 ± 7.43 | 1/1/1 | 3.8/3.2/2.1 |
| Direct selection (exhaustive) | 69.70 ± 6.45 | 71.45 ± 6.22 | 63.61 ± 9.31 | 6/6/6 | 34.2/26.0/17.3 |
| Direct selection (budget-matched) | 67.87 ± 7.32 | 71.39 ± 5.12 | 64.22 ± 8.75 | 7.8/7.2/8.0 | 45.8/33.3/23.5 |
| Adaptive weighting (SoftMax) | 66.41 ± 6.07 | 63.97 ± 4.56 | 59.28 ± 10.90 | 1/1/1 | 16.9/12.3/8.5 |
| RAFC-Fusion (Full, re-run) | 69.53 ± 5.29 | 74.50 ± 4.82 | 65.28 ± 7.36 | 7.8/7.2/8.0 | 43.2/35.3/23.6 |

Adaptive weighting learns its weights from the training loss. With seven samples per class, this loss is minimized most easily by high-capacity operators, so the weights concentrate on the concatenation baseline and the cross-modal transformer (0.33–0.44 each on Houston2013 and Muufl), whereas operators that generalize better receive small weights (e.g., Baseline-Ghost 0.02–0.05 on all datasets, although it is the best fixed operator on Augsburg). This supports the use of a separate validation signal for operator prioritization.

To examine the individual contributions of the proposed framework, we conducted a sequential component-wise ablation across Houston2013, Muufl, and Augsburg using ten random seeds (Table 13). Starting from feedforward-only operator prioritization, PID-inspired feedback, reliability-aware operator selection, and residual refinement were in-

crementally introduced. The complete framework improved mean test overall accuracy relative to the feedforward-only configuration by 0.71, 2.32, and 1.90 percentage points on Houston2013, Muufl, and Augsburg, respectively. However, the corresponding paired comparisons did not reach statistical significance at the 0.05 level.

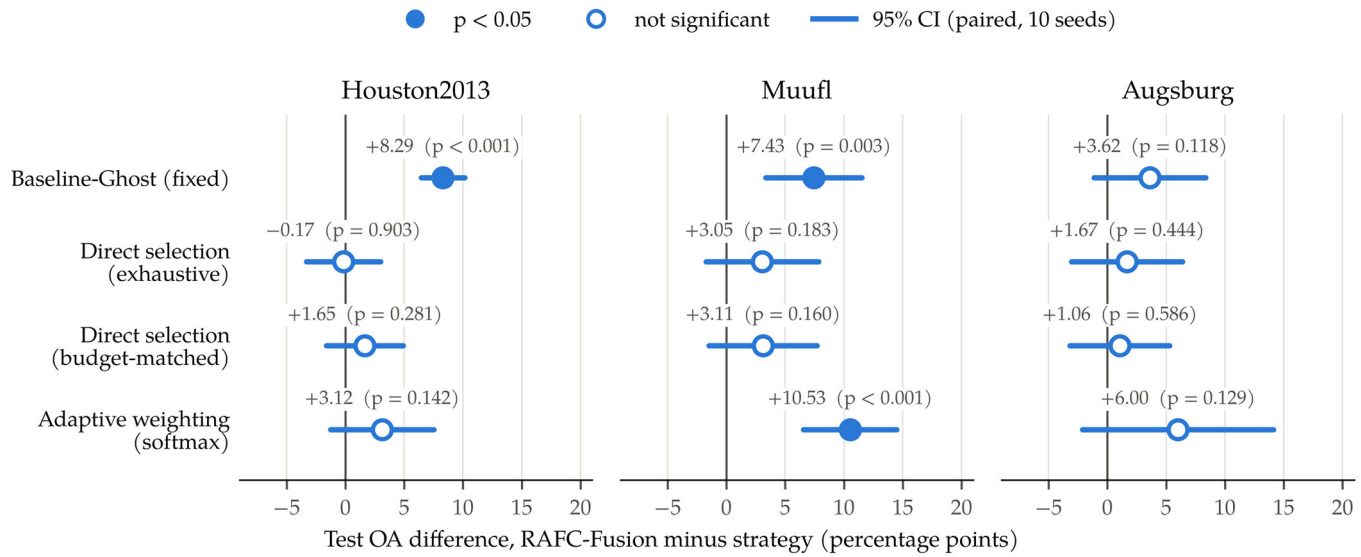



**Figure 6.** Paired differences in test OA between RAFC-Fusion and simpler operator-selection strategies (mean and 95% confidence interval over ten seeds). Filled markers indicate $p < 0.05$ (paired $t$-test); hollow markers indicate non-significant differences.

**Table 13.** Component-wise ablation (test OA, %, mean ± standard deviation over ten seeds; five controller iterations, ensemble disabled). Feedback-only: the feedforward prior is removed. Full without PID: the action is fixed to "maintain" and the refinement gain to 0.10. The last rows give paired differences (percentage points; paired $t$-test).

| Fusion Operators | Houston2013 | Muufl | Augsburg |
|---|---|---|---|
| Baseline-Ghost | 61.23 | 67.07 | 61.66 |
| FF-only | 69.17 | 72.37 | 64.08 |
| +PID-inspired feedback | 69.46 | 72.52 | 62.88 |
| +Reliability-aware selection | 68.29 | 72.33 | 66.25 |
| +Residual refinement (Full) | 69.89 | 74.69 | 65.98 |
| FB-only (Full without FF prior) | 70.65 ± 6.30 | 76.40 ± 2.73 | 71.93 ± 6.35 |
| Full without PID | 70.68 ± 5.34 | 72.88 ± 7.23 | 66.90 ± 10.56 |
| Full − FB -only | −0.77 ($p = 0.257$) | −1.71 ($p = 0.272$) | −5.95 ($p = 0.054$) |
| Full − Full without PID | −0.80 ($p = 0.012$) | +1.82 ($p = 0.352$) | −0.92 ($p = 0.673$) |
| FB -only − Baseline-Ghost | +9.42 ($p < 0.001$) | +9.34 ($p < 0.001$) | +10.27 ($p = 0.002$) |

The leave-one-out configurations show that, under the evaluated conditions, neither the feedforward prior nor the PID-inspired term contributes measurably to accuracy. Removing the feedforward prior increases mean OA by 0.77, 1.71 and 5.95 percentage points ($p = 0.26$, 0.27 and 0.054), and removing the PID signal changes it by +0.80 ($p = 0.012$), −1.82 ($p = 0.35$) and +0.92 ($p = 0.67$) percentage points on Houston2013, Muufl and Augsburg, respectively. The improvement over the Baseline-Ghost model is driven by validation-gated closed-loop training with feedback-guided refinement (feedback-only: +9.42, +9.34 and +10.27 percentage points, $p \leq 0.002$; Table 13). The feedforward prior does not yet contribute because its batch-relative standardization removes the differences between scenes and its heuristic mapping does not reflect the label efficiency of the operators (Section 4.6.4), and the PID term operates in a narrow action range under the present target definition (Section 5.4); both are discussed as directions for improvement in Section 5.5.

Results are reported as mean ± standard deviation of test overall accuracy (OA, %) over ten random seeds (42–51). Each successive configuration incrementally adds one component to the preceding configuration. The PID-only configuration disables historical reliability weighting and residual refinement. The reliability-aware configuration enables historical operator scoring without residual refinement. All controller configurations use a fixed five-iteration search budget, and the final reliability-weighted ensemble is disabled. Bold values indicate the highest mean OA for each dataset.

Although the PID-inspired component does not produce statistically significant standalone accuracy improvements, its contribution to adaptive operator selection can be examined through controller decision traces. As shown in Table 14, the PID-enabled configuration generated 39 non-maintain actions across 150 matched decision points, corresponding to 26.0% of the recorded decisions. Compared with feedforward-only prioritization, the addition of PID-inspired feedback resulted in different ordered operator-selection lists at 15 decision points (10.0%), including 11 cases with different operator-set membership (7.3%).

**Table 14.** Decision trace audit of PID-inspired feedback and reliability-aware operator selection.

| Datasets | PID Non-Maintain Actions | PID Selection Disagreement | History Selection Disagreement |
|---|---|---|---|
| Houston2013 | 11/50 (22.0%) | 5/50 (10.0%) | 16/50 (32.0%) |
| Muufl | 16/50 (32.0%) | 6/50 (12.0%) | 15/50 (30.0%) |
| Augsburg | 12/50 (24.0%) | 4/50 (8.0%) | 17/50 (34.0%) |
| Overall | 39/150 (26.0%) | 15/150 (10.0%) | 48/150 (32.0%) |

Each dataset contributes 50 matched decision points (10 seeds × 5 control iterations). PID non-maintain actions are measured in the FF + PID configuration. PID-related selection disagreement compares the ordered operator-selection lists of FF and FF + PID. History-related disagreement compares the filtered operator-selection lists of FF + PID and FF + PID + History. Among the 15 PID-related list disagreements, 11 correspond to changes in operator-set membership and four to ordering differences only. These comparisons characterize observed decision-trace differences between configurations rather than isolated per-decision causal effects.

Introducing reliability-aware selection further changed the filtered operator-selection lists at 48 decision points (32.0%). This indicates that historical operator performance and associated selection mechanisms influence the set of candidates considered during controlled fusion optimization.

These observations demonstrate that the proposed feedback-related components are operational and affect the operator-selection process. However, the observed decision changes do not consistently translate into improved classification accuracy. The decision-trace audit therefore complements the ablation results by separating the controller's operational contribution from its predictive effectiveness.

To validate the feedforward indicators, the full framework was re-run on the ten seeds with (1) a uniform prior, (2) leave-one-indicator-out priors, in which the score of one indicator is fixed at 0.5, and (3) a random permutation of the six prior values across operators. In addition, the Spearman rank correlation between the six feedforward priors and the standalone validation and test accuracy of the six operators was computed for each seed and tested by permutation (10,000 permutations).

Replacing the feedforward prior by a uniform, leave-one-indicator-out or randomly permuted prior does not change the test OA significantly after correction for multiple comparisons (Table A1; all Holm-adjusted $p \geq 0.17$), and the feedforward scores are not

positively rank-correlated with the standalone accuracy of the operators (pooled $\rho_s = -0.16$ for validation OA, $p = 0.053$). The uniform prior is numerically higher on all three datasets (+2.01 pp on average) and requires less candidate training. Under the present protocol, the characterization indicators therefore do not identify the most suitable operator; they favor higher-capacity operators in complex scenes, whereas with seven labeled samples per class these operators are among the least accurate when trained standalone. The final accuracy is nevertheless insensitive to the priority, including a randomly permuted one, because every candidate is accepted or rejected based on validation feedback (Equation (49)). This robustness to a miss-specified prior is a property of the reliability-aware feedback loop, not of the indicators. A step-by-step diagnosis shows that the batch-relative normalization of Equations (14)–(16) removes the differences between scenes and makes the scores sensitive to the sampled batch, and that, even on a fixed reference scale, the mapping of Equation (18) favors operators that are difficult to train from seven labeled samples per class.

A step-by-step diagnosis without training identified two causes. First, the batch-relative standardization of Equations (14)–(16), estimated from only two to four training mini-batches, removes the differences between the scenes (the scene explains 30–94% of the variance of the raw indicators but only 0–13% after standardization) and makes the scores sensitive to the sampled batch. Second, even on a fixed reference scale, the mapping of Equation (18) favors the cross-modal transformer and geometry-aware operators, which are among the least accurate with seven labeled samples per class (pooled Spearman $\rho_s = +0.01$ for validation OA, versus +0.63 for the mean ranking observed on other seeds of the same scene).

#### 4.6.5. Patch-Boundary Sensitivity Analysis

To further examine the spatial separation properties of the evaluation protocol, we conducted a patch-footprint audit using the original spatial splits and 7 × 7 HSI–LiDAR input patches. Although training, validation, and test sample centers belong to spatially separated development and held-out blocks, input patches are extracted from the full scenes without boundary masking. Therefore, some development patches may include contextual pixels from held-out test blocks, and some training and test patches may share spatial footprints.

Figure 7 visualizes the actual spatial partition and sample-center assignment for Houston2013 under seed 42, including an example of cross-boundary contextual exposure in a 7 × 7 training patch.

Table 15 quantifies three distinct forms of spatial-context overlap across ten predefined seeds (42–51): training patches entering held-out test blocks, validation patches entering held-out test blocks, and training patches intersecting at least one test-patch footprint.

These measurements demonstrate that the evaluation protocol provides spatially separated sample centers but does not enforce complete patch footprint disjointness. The reported overlap rates refer to input patches rather than unique pixels or held-out test labels. In particular, the overlap arises from spatial image context and should not be interpreted as the direct use of test labels for model optimization.

In Table 16, to assess whether the reported classification performance was concentrated near development–test boundaries, we additionally performed checkpoint-based boundary sensitivity analysis on Houston2013. Five ablation configurations were evaluated using their saved checkpoints across ten seeds, without retraining. Test samples were divided into boundary and interior subsets according to whether their 7 × 7 input patch intersected a development block. The reproduced full-test accuracies were compared with the original checkpoint evaluation records.

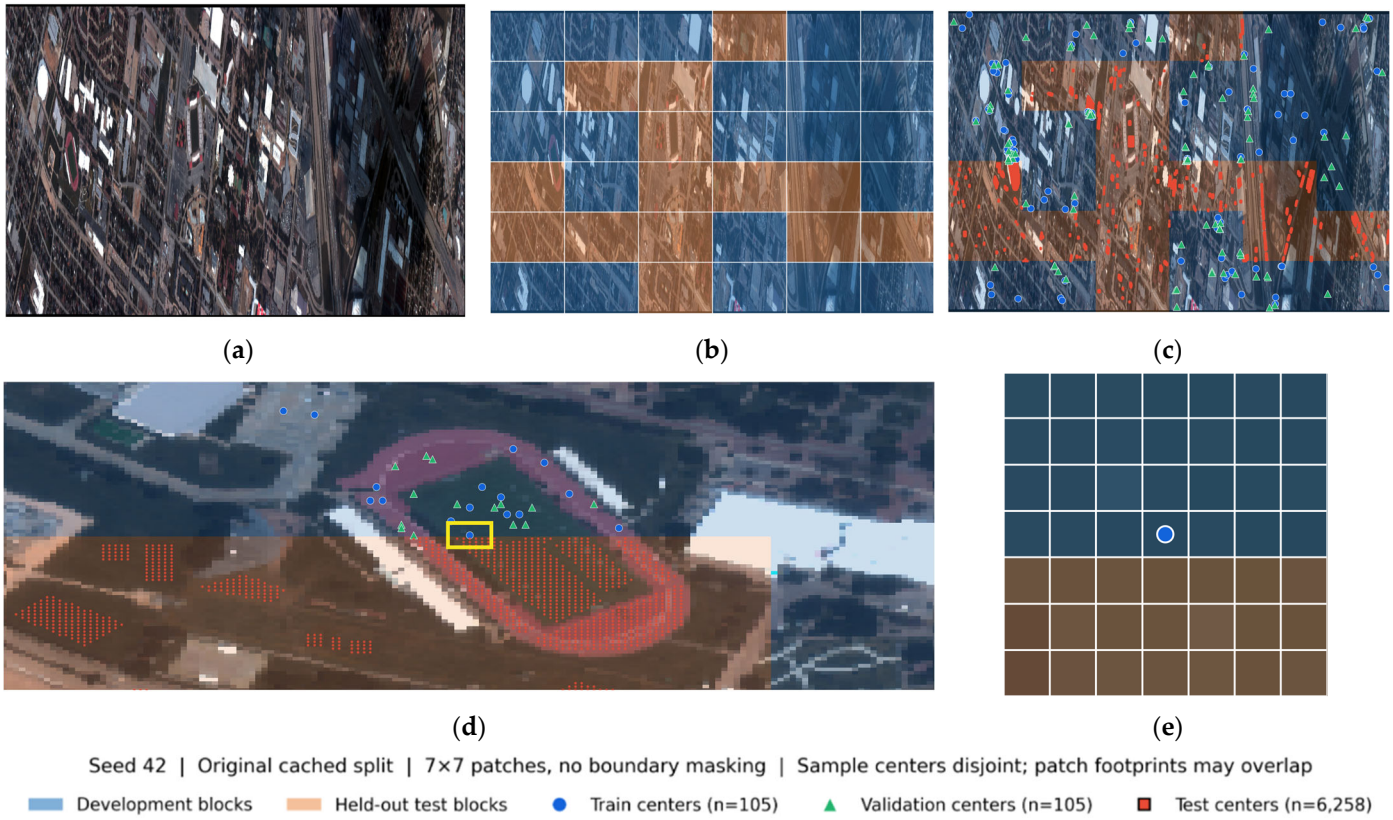



**Figure 7.** Actual spatially disjoint sample-center partition on Houston2013 (seed 42). (**a**) HSI false-color composite; (**b**) development and held-out test blocks; (**c**) training, validation, and test sample centers; (**d**) a boundary neighborhood; and (**e**) an actual 7 × 7 training patch containing pixels from a held-out test block. The protocol separates sample centers but does not enforce complete patch footprint disjointness.

**Table 15.** Patch footprint overlap audit across ten spatial split seeds using 7 × 7 input patches.

| Dataset | Training Patches Entering Test Blocks | Validation Patches Entering Test Blocks | Training Patches Overlapping Test-Patch Footprints |
|---|---|---|---|
| Houston2013 | 99/1050 (9.43%) | 101/1050 (9.62%) | 135/1050 (12.86%) |
| Muufl | 105/770 (13.64%) | 111/770 (14.42%) | 207/770 (26.88%) |
| Augsburg | 39/490 (7.96%) | 39/490 (7.96%) | 65/490 (13.27%) |

**Table 16.** Houston2013 boundary sensitivity analysis: overall accuracy (OA, %, mean ± standard deviation across ten seeds). The highest mean value for each metric is shown in bold.

| Dataset | Full-Test OA | Boundary OA | Interior OA |
|---|---|---|---|
| Baseline-Ghost | 61.23 ± 5.76 | 72.32 ± 7.12 | 60.27 ± 6.12 |
| FF-only | 69.17 ± 4.96 | 76.11 ± 5.92 | 68.60 ± 5.28 |
| +PID-inspired feedback | 69.46 ± 5.23 | 77.21 ± 6.82 | 68.80 ± 5.41 |
| +Reliability-aware selection | 68.29 ± 5.71 | 76.36 ± 6.67 | 67.60 ± 5.92 |
| **+Residual refinement (Full)** | **69.89 ± 5.43** | **77.42 ± 6.82** | **69.25 ± 5.61** |

Across the ten Houston2013 seeds, 5176 of 61,398 test sample evaluations (8.43%) were assigned to the boundary subset, while 56,222 (91.57%) were classified as interior samples. On the interior subset, the complete configuration achieved 69.25 ± 5.61% OA compared with 60.27 ± 6.12% for the baseline, corresponding to a mean paired improvement of 8.98 percentage points. This improvement indicates that the overall advantage over the baseline is not confined to test samples whose input patches intersect development blocks.

However, the difference between Full and FF on the interior subset was comparatively small (0.64 percentage points) and was not statistically significant in the paired analysis ($p = 0.383$). Moreover, the interior-only evaluation does not eliminate cross-boundary contextual exposure that may have occurred during training. Consequently, these results provide a post-hoc boundary sensitivity assessment under the original sample-center-disjoint protocol, rather than proof of leakage-free or strictly patch-disjoint model training.

To test whether the overlap affects the conclusions, the comparison was repeated under two stricter protocols that keep the same test samples for every seed: (P) a patch-disjoint protocol, in which every training and validation center lies more than 2r = 6 pixels (Chebyshev distance) from any pixel of a held-out block, so that no training or validation patch contains a held-out pixel or overlaps a test patch; and (PV) P combined with blocked validation, in which one third of the development blocks is reserved for validation and the training centers also lie more than 6 pixels from the validation blocks. Training and validation samples that satisfy these constraints were kept, and only the violating ones were replaced (Appendix A Table A2).

Under the patch-disjoint protocol, test OA decreased by 1.8–2.8 pp for Baseline-Ghost, by 0.1–4.8 pp for closed-loop training with refinement, and by 4.2 pp on Houston2013 and Muufl for RAFC-Fusion, whose accuracy increased by 2.0 pp on Augsburg; RAFC-Fusion remained 4.5–9.5 pp above Baseline-Ghost ($p \leq 0.02$); with validation drawn from separate blocks, the decreases were larger (up to 10.9 pp on Houston2013), partly because the buffer reduced the number of admissible labeled samples for some classes. The improvement over Baseline-Ghost was preserved under all protocols (+7.8 to +15.7 pp, $p \leq 0.01$). The sample-center protocol is therefore moderately optimistic in absolute accuracy, but it does not change the relative comparison.

#### 4.6.6. Computational Complexity Analysis

Table 17 compares the computational characteristics of the evaluated methods on Houston2013 under the same experimental setting. The reported measurements include model parameters, training time, inference latency, and peak GPU memory consumption.

**Table 17.** Computational complexity analysis on Houston2013.

| Method | Parameters, Total/Active (M) | Training Time incl. Search (s) | Inference Time (ms/Sample) | Peak GPU Memory (GB) |
|---|---|---|---|---|
| HybridSN | 0.74/0.74 | 37.18 | 0.17 | 0.70 |
| CCR-Net | 0.07/0.07 | 5.34 | 0.07 | 1.95 |
| MFT | 0.32/0.32 | 5.83 | 0.06 | 0.05 |
| CALC | 0.28/0.28 | 20.31 | 0.02 | 0.10 |
| MERGE-Net | 80.93/80.93 | 16.52 | 0.19 | 2.86 |
| NCGLF2 | 0.64/0.64 | 73.52 | 1.03 | 0.66 |
| Ours (Full) | 4.01/0.77–1.52 | 40.41 | 0.03 | 0.09 |

The proposed RAFC-Fusion framework contains 4.01 M parameters, of which 0.77–1.52 M are active at inference because only the selected operator branch is executed, and requires 40.41 s of training time on Houston2013, including the Baseline-Ghost initialization and the complete operator search (about eight trainings of 200 epochs). This reflects the additional model capacity and computational cost associated with reliability-aware fusion and feedback-controlled operator search. Nevertheless, its measured inference latency of 0.03 ms per sample, with a peak GPU memory usage of 0.09 GB, is lower than that of HybridSN (0.17 ms/sample with 0.70 GB of memory), CCR-Net (0.07 ms/sample with 1.95 GB), and MFT (0.06 ms/sample with 0.05 GB).

An important distinction is that feedback-controlled operator search is performed during model development rather than repeated in full for every inference sample. In the standard inference path, the model executes the selected fusion operator, followed by the classification head and optional feedback-guided feature refinement. Consequently, the inference cost is determined by the active execution path rather than by the simultaneous execution of all available fusion operators. When reliability-weighted multi-operator ensembling is enabled, however, additional operators are executed, and the inference cost may increase.

Computational profiling implementation measures training time using synchronized wall-clock timing and evaluates inference in model evaluation mode without gradient computation. Inference measurements exclude warm-up passes and are averaged over three times repetitions, with CUDA synchronization applied to obtain consistent GPU timing. The reported inference latency therefore characterizes the measured model-forward execution rather than the complete large-area remote sensing processing pipeline.

Although these results demonstrate relatively low inference latency and GPU memory consumption under the evaluated conditions, they do not establish real-time performance for large-area imagery. The total computational workload increases with the number of spatial locations to be classified, while data loading, patch extraction, and output reconstruction introduce additional processing costs. Therefore, the proposed framework should be regarded as an accuracy–complexity trade-off rather than a verified real-time solution. Future work will investigate tiled inference, memory-efficient batching, and further optimization of the fusion and feedback components to assess scalability in operational remote sensing applications.

#### 4.6.7. Robustness to Degraded Inputs, Shadows and Class Boundaries

To examine robustness without retraining, the saved models of Table 11 (Baseline-Ghost and RAFC-Fusion, ten seeds) were evaluated on the held-out test set after the normalized HSI cube had been degraded by spatial Gaussian blur ($\sigma = 0.5$, 1 and 2 pixels), reduced illumination (gain 0.7, 0.5 and 0.3), reduced contrast (factor 0.7, 0.5 and 0.3 around the band mean) or additive Gaussian noise (standard deviation 0.01, 0.02 and 0.05). In addition, the clean test set was divided into shadowed pixels, whose mean HSI brightness is below half of the scene mean, and class-boundary pixels, whose $3 \times 3$ ground-truth neighborhood contains another class, and which serve as a proxy for mixed pixels (Appendix A Table A5). Spatial blur and weak noise had little effect: with $\sigma = 2$ pixels or a noise standard deviation of 0.02, the OA of RAFC-Fusion decreased by at most 3.0 percentage points, and its advantage over Baseline-Ghost was preserved (+7.8 and +7.0 points on Houston2013 and Muufl at $\sigma = 2$, $p \leq 0.004$; +3.3 points on Augsburg).

Radiometric changes were more harmful. A mild contrast reduction (factor 0.7) reduced the OA by at most 5.4 points, and RAFC-Fusion remained the more accurate model on all datasets, whereas a global reduction in illumination degraded both models strongly (by 9–54 points), because the intensity normalization is fixed by the scene on which the models were developed. Under reduced illumination, RAFC-Fusion retained its advantage on Muufl (+8.2 to +9.8 points, $p \leq 0.008$), but on Houston2013 its accuracy decreased more than that of Baseline-Ghost (for example, by 33.9 versus 17.7 points at contrast factor 0.3 and by 12.5 versus 4.8 points at a noise standard deviation of 0.05), which indicates that part of its additional accuracy on clean data relies on spectral detail that is sensitive to radiometric shifts. AA and Kappa changed in the same direction as OA.

On the clean test set, shadowed pixels were classified much less accurately than the test set by both models, and RAFC-Fusion was more accurate on them on Houston2013 (38.97% versus 34.48%, $p = 0.007$) and Muufl (56.19% versus 49.44%, $p = 0.011$). Class-

boundary pixels could be identified only on Muufl, whose ground truth is dense; RAFC-Fusion classified them with 45.10% compared with 39.59% for Baseline-Ghost ($p = 0.004$), far below the accuracy on the whole test set (74.69% and 67.06%). The framework is therefore robust to spatial blur and moderate noise, whereas strong illumination or contrast differences between the development and test regions call for radiometric augmentation or normalization (Section 5.4).

### *4.7. Analysis of Operator Selection Under Label Scarcity*

Because the closed-loop procedure trains several candidates, part of its improvement may stem from the additional training. Its sources were therefore separated with controls that start from the same Baseline-Ghost checkpoint and use the same optimizer, schedule, and validation-based checkpoint selection as a controller candidate, but involve no operator switching (Table 18). To examine operator switching under fair conditions, a variant with symmetric warm starts was also evaluated, in which every operator first receives the same 200-epoch training and each candidate continues its own trajectory, and it was compared with a budget-matched single-operator control at 7, 15, and 30 labeled samples per class (Figure 8).

**Table 18.** Sources of the improvement over Baseline-Ghost under a matched training budget (test OA, %, mean ± standard deviation over seeds 42–51; all configurations start from the same Baseline-Ghost checkpoint; ensemble disabled).

| Configuration (Test OA, %; Seeds 42–51) | Houston2013 | Muufl | Augsburg |
|---|---|---|---|
| Baseline-Ghost (200 epochs) | 61.23 ± 5.76 | 67.07 ± 6.31 | 61.66 ± 7.43 |
| +one continuation of 200 epochs | 67.60 ± 5.45 | 74.86 ± 2.51 | 70.58 ± 5.73 |
| +one continuation with feedback-guided refinement | 70.70 ± 5.77 | 76.54 ± 1.90 | 70.23 ± 6.98 |
| +five validation-gated warm restarts with refinement | 72.04 ± 6.11 | 76.60 ± 2.77 | 71.35 ± 7.58 |
| RAFC-Fusion, controller restricted to Baseline-Ghost | 70.75 ± 5.79 | 76.59 ± 3.45 | 70.74 ± 6.29 |
| RAFC-Fusion, feedback-only (uniform prior) | 70.65 ± 6.30 | 76.40 ± 2.73 | 71.93 ± 6.35 |
| RAFC-Fusion (Full) | 69.89 ± 5.43 | 74.69 ± 4.77 | 65.98 ± 7.33 |

RAFC-Fusion improves the Baseline-Ghost model on all datasets (+8.65, +7.62, and +4.32 pp; Table 18), and closed-loop training with refinement improves it by 9–11 pp—an improvement that is preserved under a patch-disjoint protocol (Section 4.6.5). Budget-matched controls (Table 18) show that, with seven labeled samples per class, this improvement is attributable to validation-gated closed-loop training and to the feedback-guided refinement, whereas switching between operators does not add accuracy beyond that of a controller restricted to the initial operator. When all operators are trained on an equal footing, the controller identifies the operator that performs best in isolation; its cost relative to a budget-matched single-operator control decreases from 2.1 pp at seven labels per class to 1.4 pp at 15 labels, and vanishes at 30 labels per class (Figure 8). Adaptive operator selection therefore becomes viable as the label budget, and with it the reliability of the validation signal, increases.

Figure 8 shows how the effect of adaptive operator switching depends on the label budget. With seven labeled samples per class, the controller is 2.1 pp less accurate than a control that spends the same training budget on a single operator ($p = 0.049$), because the validation signal is coarse and the budget is distributed over operators that cannot yet be trained to their potential. This gap decreases to 1.4 pp at 15 labels per class ($p = 0.106$) and vanishes at 30 labels per class (+0.1 pp, $p = 0.931$), consistently on all three datasets. Operator switching thus becomes competitive as the label budget, and with it the reliability of the validation signal, increases, which identifies the label budget as the main condition for exploiting the operator library.

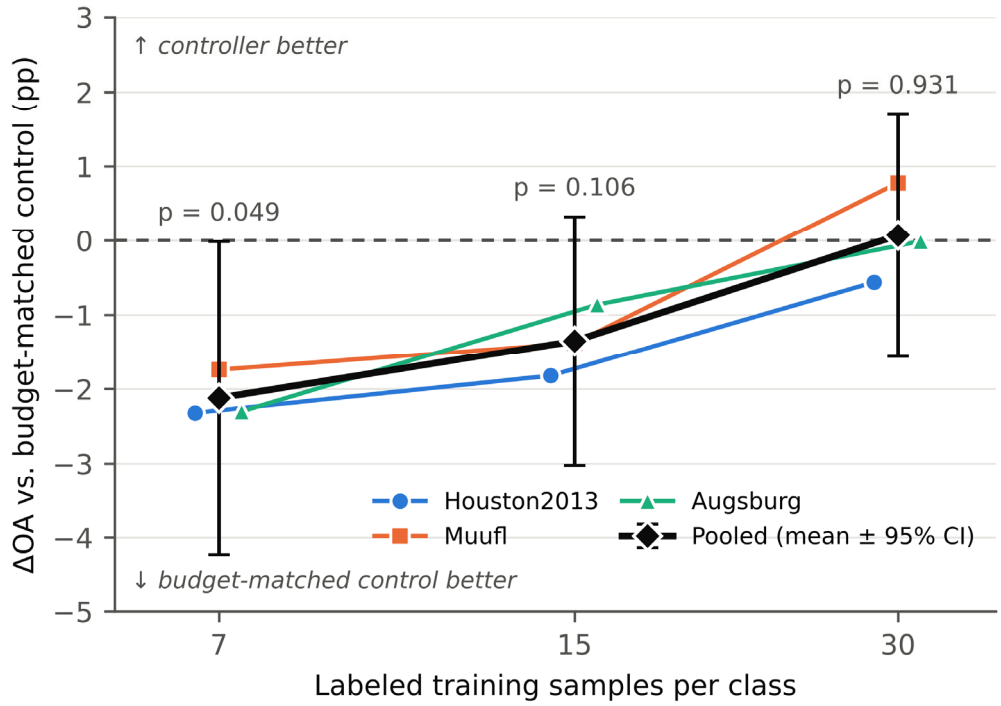



**Figure 8.** Test OA of the controller (symmetric warm starts) is relative to a budget-matched single-operator control, as a function of the label budget; mean ± 95% CI pooled over datasets and seeds.

## 5. Discussion

The experimental results demonstrate that the main benefit of RAFC-Fusion improves on a fixed Baseline-Ghost model on all three datasets and attains the accuracy of the dataset-specific best fixed operator without knowing it in advance. Under the sample-center spatially disjoint protocol, the proposed framework achieved mean OA values of 70.00%, 73.20%, and 66.34% on Houston2013, Muufl, and Augsburg, respectively. These results are particularly relevant because only seven training and seven validation samples per class were available, and the held-out spatial regions were excluded from label-based model development. Therefore, the observed improvements reflect cross-region generalization under limited supervision rather than performance under conventional random pixel-wise sampling.

### *5.1. Interpretation of the Adaptive Fusion Behavior*

A central observation from the fusion-operator analysis is that no individual operator consistently provides the strongest performance across all three datasets. As shown in Table 10, the physics-informed operator is the best fixed operator on Houston2013 and Muufl, and Baseline-Ghost on Augsburg (ten seeds), and the ranking of the remaining operators changes between datasets. RAFC-Fusion attains accuracy comparable to that of the best fixed operator on each dataset without prior knowledge of which operator is best. This result supports the working hypothesis that heterogeneous HSI–LiDAR scenes cannot always be represented effectively by a single predefined fusion strategy.

This behavior is consistent with the diversity of multimodal fusion mechanisms reported in previous HSI–LiDAR studies. CNN-based methods emphasize local spectral–spatial representation, transformer-based methods improve long-range multimodal interactions, and attention or gating mechanisms provide mechanisms for emphasizing complementary modality information. These approaches have demonstrated that different forms of feature interaction can be advantageous depending on scene characteristics. RAFC-Fusion extends this perspective by treating the choice among complementary fusion mechanisms as an adaptive decision rather than fixing the same interaction mechanism for all datasets.

The poor performance of random operator selection provides additional evidence for this interpretation. Random selection produces OA values of 47.21%, 27.38%, and 44.83% on Houston2013, Muufl, and Augsburg, respectively, substantially below the complete adaptive framework (Table 11). Consequently, simply incorporating multiple fusion operators is insufficient; the operators must be selected and trained under validation control. Direct validation-based selection over the same operator library achieves comparable accuracy under a matched training budget (Table 12), and the budget-matched analysis of Section 4.7 attributes the improvement over a fixed operator mainly to validation-gated closed-loop training and feedback-guided refinement. Adaptive operator switching does not yet add accuracy with seven labels per class, but its cost relative to single-operator training vanishes as the label budget increases (Figure 8). The benefit over a fixed operator should therefore be attributed mainly to validation-guided operator search, residual refinement and warm-started training, rather than to a selection principle that is fundamentally different from validation-based model selection.

### *5.2. Feedback Control and Reliability-Aware Adaptation*

The results also clarify what feedback-control principles contribute to multimodal fusion. Previous control-inspired deep-learning studies have primarily used PID-related mechanisms for optimization, parameter adaptation, or the improvement of convergence. In contrast, the proposed framework applies to a PID-inspired feedback mechanism at the architectural decision level, where validation performance influences the priority assigned to different fusion operators. The controller therefore does not replace gradient-based learning; instead, it provides a higher-level mechanism for deciding which fusion behavior should receive greater emphasis during optimization.

The controller sensitivity results further indicate that the adaptive behavior is not restricted to a single narrowly tuned parameter configuration. Neighboring controller settings maintain broadly comparable performance across the three datasets, suggesting that validation-driven acceptance, rather than any single controller coefficient, determines the outcome. This observation is relevant for practical application because extensive dataset-specific controller tuning would otherwise reduce the usefulness of the proposed adaptive strategy.

The feedforward and feedback components also play complementary roles. Dataset characterization provides a heuristic initial ordering based on terrain complexity, spectral diversity, shadow ratio, spatial correlation, and modality agreement, while validation feedback subsequently adjusts these priorities according to observed classification behavior. The indicator ablation and correlation analysis (Appendix A Table A1) shows that this ordering does not predict the most accurate operator under seven labeled samples per class, and that the final accuracy is insensitive to it because candidate acceptance is decided by validation feedback. This combination distinguishes the proposed framework from purely static fusion and from adaptation based only on validation performance. In particular, the feedback component allows the initial preference to be corrected when actual validation behavior does not agree with that prior.

### *5.3. Generalization Under Spatially Disjoint Evaluation*

The spatial block-based protocol separates development and test sample centers, but does not completely isolate their input patch footprints. As quantified in Section 4.6.5, a subset of training and validation patches includes contextual pixels from held-out test blocks. The additional Houston2013 boundary sensitivity analysis indicates that the complete framework retains an overall accuracy advantage over the baseline on interior test samples. Retraining under a patch-disjoint protocol (Appendix A Table A2) reduced abso-

lute accuracies moderately, but preserved the improvement over the baseline. Therefore, the main results should be interpreted as cross-region classification under sample-center-disjoint evaluation, whose absolute accuracies are moderately optimistic compared with patch-disjoint training.

The multi-seed experiments provide further evidence regarding this generalization behavior. Positive gains are observed in 10/10 Houston2013 runs, 9/10 Muufl runs, and 8/10 Augsburg runs. Thus, the improvement obtained through feedback-controlled adaptation is not limited to a single favorable training–validation sample. This is particularly important because the development set contains only seven training and seven validation samples per class, making the optimization process inherently sensitive to sample selection.

Nevertheless, the validation–test results also reveal an important limitation. Houston2013 exhibits a validation–test gap of $16.57 \pm 7.15$ percentage points, considerably larger than the corresponding gaps of $2.13 \pm 4.80$ for Muufl and $3.72 \pm 5.11$ for Augsburg. This indicates that successful adaptation to a small validation subset does not necessarily guarantee equivalent performance in a spatially distinct test region. Rather than contradicting the usefulness of validation-guided fusion, this result highlights a fundamental limitation of development-set feedback under strong spatial distribution shift. When the validation samples do not adequately represent the spatial and spectral conditions of the held-out region, the controller can optimize operator priorities only with respect to the information available during development.

This finding also suggests that future adaptive fusion systems should move beyond dataset-level reliability estimation. Region-level or sample-level reliability modeling could allow fusion behavior to respond to local variations in terrain, illumination, spectral distribution, and modality consistency. Such an extension may be particularly useful for large heterogeneous scenes in which different spatial regions favor different multimodal interactions.

### *5.4. Accuracy–Complexity Trade-Off*

The computational results provide an additional perspective on the practical implications of the proposed framework. RAFC-Fusion contains 4.01 M parameters, more than all compared methods except MERGE-Net (80.93 M), and requires 40.41 s of training on Houston2013, including the complete operator search, which is more than HybridSN, CCR-Net, MFT, CALC and MERGE-Net but less than NCGLF2 (73.52 s; Table 17). Therefore, the proposed method should not be interpreted as a minimum-complexity or lightweight architecture. The larger training cost primarily reflects the availability and evaluation of multiple candidate fusion mechanisms during adaptive optimization.

However, the computational behavior during inference is considerably different from the training stage cost. The final configuration requires approximately 0.03 ms per sample and 0.09 GB peak GPU memory in the reported experimental environment. These values indicate that the additional adaptive search cost is concentrated mainly during model development rather than necessarily translating into proportionally high deployment cost. Consequently, the contribution of RAFC-Fusion is more appropriately characterized as an accuracy–adaptability–complexity trade-off than as a reduction in computational complexity.

This distinction is also important when considering the use of GhostConv. Ghost convolution was introduced to reduce redundant feature generation [11] and has previously been adapted to hyperspectral classification [12]. In the proposed framework, GhostConv reduces redundancy within individual feature-processing components, but this does not make the complete adaptive architecture lightweight because multiple candidate operators increase the total model capacity and training procedure. The computational results there-

fore provide a more balanced interpretation: efficient internal feature extraction can coexist with increased system-level complexity when adaptive architectural selection is introduced.

### *5.5. Limitations and Future Directions*

Several limitations should be considered when interpreting the present findings. First, although three HSI–LiDAR benchmarks with substantially different characteristics were evaluated, additional datasets would be necessary to establish broader generalization across sensors, spatial resolutions, geographic regions, and acquisition conditions. Second, the current reliability characterization operates primarily at the dataset level. This design provides stable priors under extremely limited supervision but cannot fully represent within-scene spatial heterogeneity. Moreover, the feedforward prior did not predict the most accurate operator in the present experiments: a uniform or randomly permuted prior gave statistically indistinguishable accuracy (Appendix A Table A1). Two causes were identified (Section 4.6.4): the batch-relative normalization, which is estimated from only two to four training mini-batches under seven labeled samples per class and removes the differences between scenes, and the heuristic mapping of Equation (18), which favors operators that are difficult to train from very few labels. Learning this mapping from a larger collection of scenes and characterizing the operators by their label efficiency in addition to the condition they are designed to address are directions for future work.

Third, the feedback controller depends on a small validation subset. The larger validation–test gap observed on Houston2013 demonstrates that validation feedback may become less representative when strong spatial distribution shifts occur between development and deployment regions. Future work could therefore investigate uncertainty-aware validation signals, spatially stratified reliability estimation, or region-specific controller states to reduce dependence on a small global validation subset.

Fourth, adaptive operator prioritization increases training complexity because several candidate fusion mechanisms must be maintained and evaluated. Future research should investigate parameter sharing among operators, lightweight shared representations, operator pruning, and early termination of the controller search. These strategies could retain the benefits of adaptive fusion while reducing the additional optimization cost.

Fifth, the benefit of adaptive operator switching depends on the label budget. With seven labeled samples per class, the validation signal is coarse (one validation sample corresponds to 1–2 percentage points), and higher-capacity operators cannot be trained to their potential, so that switching does not yet add accuracy beyond budget-matched single-operator training (Section 4.7, Figure 8). As the label budget increases, this gap closes, which indicates that operator adaptation becomes viable when the validation signal and the operators themselves become more reliable. More reliable validation estimates, for example repeated or cross-validated estimates within the development region, are a natural next step.

Sixth, under the present target definition, the PID-inspired term operates within a narrow action range (enable-simple and maintain), because the adaptive target bounds the initial error; normalizing the error activates the full range but did not change accuracy in a pilot analysis on the same seeds. Formulating feedback on a continuous training signal, such as the generalization gap, would allow it to act as a genuine controller. Similarly, the reliability score of Equation (46) is only weakly discriminative when each operator is evaluated a few times and its rewards are smoothed. In an exploratory analysis (five seeds), a score calibrated to the resolution of the validation set (accuracy differences expressed in validation samples, without smoothing) improved the validation accuracy reached by the controller by 2.1 percentage points at seven labels per class and matched budget-matched

single-operator training at 30 labels per class but did not exceed it; more reliable validation estimates therefore remain the main prerequisite.

Seventh, operator selection is performed once on development data and does not adapt to degradations of the HSI modality that occur only at inference. The models are robust to spatial blur and weak noise, but strong illumination or contrast changes reduce the accuracy of both models considerably, and remove the advantage of RAFC-Fusion on Houston2013 (Section 4.6.7, Appendix A Table A5). Radiometric augmentation during training, per-image normalization and sample-level, degradation-aware weighting of the two modalities are directions for future work.

Finally, the present framework focuses on dataset-level operator prioritization under closed-set classification. A natural extension is to develop sample- or region-level dynamic fusion in which operator weights respond directly to local multimodal reliability. Combining such mechanisms with uncertainty estimation and open-set recognition could further improve robustness when the deployed scene contains spatial distributions or semantic categories that are insufficiently represented during development.

An additional limitation concerns the spatial extent of input patches. Although the evaluation separates development and test sample centers, the original patch extraction procedure permits contextual overlap across block boundaries. Retraining under a patch-disjoint protocol showed that this makes absolute accuracies moderately optimistic (by up to 5 percentage points; Appendix A Table A2) but does not change the relative comparison; future benchmarks should nevertheless adopt patch-disjoint sampling with spatial buffer zones by default.

In addition, a fixed 7 × 7 neighborhood is used for all classes and scenes, although the best neighborhood size depends on the dataset (Table 9), and class-boundary pixels are classified less accurately than the test set as a whole (Section 4.6.7). Multi-scale and adaptive neighborhoods, which adapt the size and shape of the spatial support to the local structure [59], and boundary-aware semantic segmentation, which explicitly models object edges in high-resolution imagery [60], could reduce the loss of features characteristic of small or elongated objects. Incorporating such adaptive patch reshaping into the operator library is a promising extension.

Overall, the results support the central hypothesis of this study: the appropriate HSI–LiDAR fusion behavior is dataset-dependent, and validation-guided closed-loop training can exploit this variability, matching the dataset-specific best fixed operator without prior knowledge of it, while adaptive operator switching becomes more effective as the label budget increases. At the same time, the Houston2013 validation–test gap and the increased training complexity demonstrate that adaptive fusion should be viewed as a mechanism for improving the accuracy–generalization trade-off rather than as a universal solution to spatial distribution shift.

## 6. Conclusions

This work presented RAFC-Fusion, a reliability-aware feedback-controlled framework for HSI-LiDAR classification under sample-center spatially disjoint and extremely limited supervision. Rather than assuming that a single fusion mechanism is optimal across heterogeneous scenes, the proposed framework formulates multimodal fusion as a closed-loop operator-prioritization process. Training-derived feedforward characterization provides dataset-level reliability priors, while validation-driven feedback iteratively refines operator priorities through a PID-inspired controller, acceptance-based updates, and restoration of the best reliable model. This design enables complementary fusion operators to be exploited without using information from the held-out test regions during model development.

Under the unified ten-run spatially disjoint protocol, RAFC-Fusion achieved mean OA values of 70.00%, 73.20%, and 66.34% on Houston2013, Muufl, and Augsburg, respectively, providing the highest mean OA, AA, Kappa and Macro-F1 among the evaluated methods on all three datasets with about eight times the training budget of a single model. The positive-gain behavior observed in 10/10, 9/10, and 8/10 runs further indicates that the controller provides consistent benefits across different development samples rather than depending on a single favorable split. The ten-seed operator analysis additionally showed that no individual fusion operator was best on all three datasets, and that RAFC-Fusion matches the dataset-specific best fixed operator without prior knowledge of it. Together with the budget-matched analyses, these findings indicate that, under heterogeneous spatially disjoint conditions with very few labels, validation-gated closed-loop training with feedback-guided refinement is an effective alternative to training a single fixed fusion model once; RAFC-Fusion attains the accuracy of the dataset-specific best operator without prior knowledge of it, and adaptive operator switching becomes viable as the label budget, and with it the reliability of the validation signal, increases.

The computational analysis further clarifies the scope of this contribution. RAFC-Fusion is not designed to minimize parameter count or training cost; its 4.01 M parameters (0.77–1.52 M active at inference) and 40.41 s training time, which includes the complete operator search, reflect the additional capacity of the multi-operator adaptive framework. Nevertheless, it retains an inference latency of 0.03 ms per sample and a peak GPU memory requirement of 0.09 GB on Houston2013. Thus, the proposed method provides an accuracy–complexity trade-off in which increased model capacity is accompanied by improved classification performance while maintaining low inference latency and modest GPU memory usage.

A current limitation is that reliability characterization and operator prioritization are performed primarily at the dataset level, and the PID-inspired controller provides bounded empirical adaptation rather than a formal control-theoretic stability guarantee. Future work will therefore investigate sample- and region-level reliability estimation, more efficient operator sharing to reduce training complexity, and theoretically grounded controller formulations for adaptive multimodal fusion under broader cross-region and cross-dataset conditions.

**Author Contributions:** Conceptualization, D.M., H.W., A.W. and M.W.; methodology, D.M., H.W., A.W. and M.W.; software, D.M.; validation, D.M.; writing—original draft preparation, D.M.; writing—review and editing, H.W., A.W. and M.W.; visualization, D.M.; supervision, A.W. and H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by Technology Project of Heilongjiang Provincial Department of Transportation (HJK2024B002), Shandong Provincial Natural Science Foundation Youth Project (Category C) under Grant Number (ZR2026QC0815). and the National Key Support Project for Foreign Experts of Northeast Special Project (D20250098).

**Data Availability Statement:** Houston2013 and Augsburg: https://drive.google.com/file/d/1UaeUWqTHhXzpwGHZElcF8AHwmVo0IwaV/view (accessed on 21 May 2021); Muufl: https://github.com/GatorSense/MUUFLGulfport (accessed on 17 April 2017). The code is freely available at https://github.com/Davaajargal-M/RAFC-Fusion-for-spatially-disjoint (accessed on 24 May 2026).

**Conflicts of Interest:** The authors declare no conflicts of interest.

# Appendix A

**Table A1.** Indicator ablation (test OA, %, mean ± standard deviation over ten seeds; in parentheses, the paired difference from the full prior in percentage points and the paired *t*-test *p*-value), Spearman correlation between the feedforward priors and the standalone accuracy of the operators, and number of candidate trainings. Mean column: mean over the three datasets, with the paired test over the 30 seed–dataset pairs. Holm-adjusted $p \geq 0.56$ within each dataset and $\geq 0.17$ pooled.

| Prior USED by RAFC-Fusion (Test OA, %) | Houston2013 | Muufl | Augsburg | Mean |
|---|---|---|---|---|
| Full prior (TC, SD, SR, SC, MA) | 69.53 ± 5.29 | 74.50 ± 4.82 | 65.28 ± 7.36 | 69.77 |
| Uniform prior (feedback-only) | 70.68 ± 6.37 (+1.16; 0.236) | 76.13 ± 2.82 (+1.63; 0.287) | 68.51 ± 6.95 (+3.23; 0.126) | 71.77 (+2.01; 0.024) |
| w/o TC | 69.48 ± 5.31 (−0.05; 0.343) | 72.93 ± 6.96 (−1.57; 0.343) | 65.71 ± 8.34 (+0.43; 0.717) | 69.37 (−0.40; 0.545) |
| w/o SD | 69.59 ± 5.37 (+0.06; 0.384) | 72.67 ± 6.59 (−1.83; 0.261) | 67.84 ± 6.42 (+2.56; 0.161) | 70.03 (+0.27; 0.742) |
| w/o SR | 68.73 ± 5.45 (−0.80; 0.242) | 71.12 ± 7.00 (−3.38; 0.080) | 64.75 ± 8.75 (−0.53; 0.656) | 68.20 (−1.57; 0.041) |
| w/o SC | 70.18 ± 6.38 (+0.66; 0.343) | 74.50 ± 4.92 (−0.00; 0.995) | 65.28 ± 7.36 (identical in 10/10 seeds) | 69.99 (+0.22; 0.362) |
| w/o MA | 68.74 ± 6.34 (−0.79; 0.245) | 76.10 ± 2.68 (+1.60; 0.239) | 64.87 ± 8.99 (−0.41; 0.847) | 69.90 (+0.14; 0.870) |
| Permuted operator assignment | 71.51 ± 4.68 (+1.98; 0.239) | 73.75 ± 6.15 (−0.75; 0.711) | 64.50 ± 12.60 (−0.78; 0.802) | 69.92 (+0.15; 0.906) |
| Spearman $\rho_s$ (SFF, standalone val. OA) | −0.02 ± 0.39 ($p = 0.916$) | −0.23 ± 0.57 ($p = 0.105$) | −0.24 ± 0.43 ($p = 0.093$) | −0.16 ± 0.47 ($p = 0.053$) |
| Spearman $\rho_s$ (SFF, standalone test OA) | −0.06 ± 0.47 ($p = 0.710$) | −0.33 ± 0.34 ($p = 0.017$) | −0.19 ± 0.39 ($p = 0.177$) | −0.19 ± 0.40 ($p = 0.017$) |
| 200-epoch training until termination (full/uniform) | 7.8/7.4 | 7.2/5.4 | 8.0/7.2 | 7.7/6.7 |

**Table A2.** Test OA (%) under the sample-center (S), patch-disjoint (P) and patch-disjoint blocked-validation (PV) protocols. S = sample-center protocol of the main experiments; P = patch-disjoint; PV = patch-disjoint with blocked validation. Ten seeds (42–51); identical test samples in all protocols; all improvements of closed-loop training over Baseline-Ghost $p \leq 0.01$ (paired *t*-test). RAFC-Fusion (Full) uses the ablation setting of Table 11 (five iterations, ensemble disabled); its S values are those of Table 12. The S values were obtained in re-runs and therefore differ slightly from Tables 4–6. With the 6-pixel buffer, fewer than seven admissible samples remained for some classes on Houston2013 (P, 101 of 105 training samples on average; PV, 95 training and 83 validation samples) and on Muufl (PV, 6 of 10 seeds).

| Method (Test OA, %) | Protocol | Houston2013 | Muufl | Augsburg |
|---|---|---|---|---|
| Baseline-Ghost | S | 59.73 ± 6.93 | 68.73 ± 4.74 | 60.32 ± 9.71 |
| Baseline-Ghost | P | 57.08 ± 5.86 | 65.98 ± 5.87 | 58.48 ± 8.74 |
| Baseline-Ghost | PV | 51.64 ± 7.28 | 64.30 ± 4.49 | 55.59 ± 13.64 |
| Closed-loop training with refinement | S | 75.45 ± 6.81 | 76.51 ± 3.27 | 70.02 ± 8.96 |
| Closed-loop training with refinement | P | 70.70 ± 6.68 | 74.85 ± 3.46 | 69.93 ± 6.35 |
| Closed-loop training with refinement | PV | 64.53 ± 7.96 | 72.27 ± 5.23 | 70.80 ± 5.97 |
| RAFC-Fusion (Full) | S | 69.89 ± 5.43 | 74.69 ± 4.77 | 65.98 ± 7.33 |
| RAFC-Fusion (Full) | P | 65.69 ± 6.03 | 70.50 ± 7.15 | 68.02 ± 7.72 |
| Improvement of closed-loop training over Baseline-Ghost | S/P/PV | +15.72/+13.62/+12.88 | +7.78/+8.87/+7.98 | +9.70/+11.45/+15.21 |
| Improvement of RAFC-Fusion (Full) over Baseline-Ghost | P | +8.61 ($p < 0.001$) | +4.52 ($p = 0.011$) | +9.54 ($p = 0.020$) |

**Table A3.** Constants of the hybrid controller (Section 3.4), their values and how they were determined. All constants are identical for the three datasets and were fixed before the experiments; none were tuned on held-out test data.

| Symbol | Role (Equation) | Value | How Determined |
|---|---|---|---|
| $K_p$, $K_i$, $K_d$ | PID gains (41) | 0.4, 0.05, 0.3 | Sensitivity analysis (Table 8) |
| $\alpha$, $\beta$ | feedforward/feedback weights (48) | 0.8, 0.2 | Sensitivity analysis (Table 8) |
| $\varepsilon_g$ | exploration probability | 0.05 | Sensitivity analysis (Table 8) |
| $\theta_1$, $\theta_2$ | action thresholds (42) | 0.15, 0.05 | Pilot analysis of the action range (Section 5.4) |
| $\gamma_{lr}$ | candidate learning-rate decay | 0.5 | Pilot analysis (Section 5.4) |
| $\delta$ | acceptance margin (49) | 0.015 | A priori (design rationale): $\approx$ one to two validation samples (1/105, 1/77, 1/49) |
| $\Delta_1$, $\Delta_2$; $a_1$, $a_2$ | requested gain; baseline-accuracy bands (37) | 0.10, 0.05; 0.85, 0.95 | A priori (design rationale): reachable improvement, smaller when the baseline is already high |
| $A_1^{max}$, $A_2^{max}$, $A_3^{max}$ | target caps (37) | 0.98, 0.97, 0.99 | A priori (design rationale): keep the target below 1 |
| $\Delta_c$; $\theta_{TC}$, $\theta_{MA}$ | target reduction; indicator thresholds (37) | 0.02; 0.7, 0.5 | A priori heuristic, not tuned |
| $\Delta_{min}$ | minimum requested gain (37) | 0.02 | A priori (design rationale): guarantees a positive initial error |
| $I_{max}$ | integral bound (40) | 1 | A priori (design rationale): anti-windup; never reached ($\lvert e\rvert \leq 0.1$, $K = 5$) |
| $\rho_{min}$, $\rho_{max}$ | refinement-gain range (43) | 0.05, 0.20 | A priori (design rationale): keep refinement a small residual correction |
| $\mu_A$, $\mu_r$ | smoothing of accuracy/ reward (38), (45) | 0.7, 0.8 | A priori heuristic, not tuned |
| $\lambda_c$, $\eta$ | confidence weight; stability penalty (44), (46) | 0.2, 0.5 | A priori: accuracy dominates the reward because SoftMax confidence is poorly calibrated with few labels; $\eta = 0.5$ as in the validation stability criterion (Section 4.6.1) |
| $\kappa$ | reliability scale (46) | 10 | A priori heuristic, not tuned |
| $w_v$ | validation-variance penalty in candidate comparison | 0.25 | A priori heuristic, not tuned |
| $\Delta_s$, $n_s$, $T_s$ | suppression rule | 0.02, 2, 2 | A priori heuristic, not tuned |
| K, $E_c$ | iterations; epochs per candidate | 5, 200 | A priori (design rationale): $E_c$ equals the training length of a single model |
| $\Delta_e$, $\omega$, M, $\tau$ | ensemble margin, score weight, size, temperature (50) | 0.01, 0.8, 3, 0.08 | A priori heuristic, not tuned |
| $\pi_a(m)$ | action priors (47) | Table A4 | A priori heuristic, not tuned |

**Table A4.** Operator-specific action priors $\pi_a(m)$ of PID-inspired feedback (Equation (47)).

| Action $a_k$ | Baseline | Baseline-Ghost | Geometry-Aware | Adaptive Gating | Cross-Modal Transformer | Physics-Informed |
|---|---|---|---|---|---|---|
| enable-complex | 0.05 | 0.10 | 0.30 | 0.60 | 1.00 | 0.80 |
| enable-simple | 0.50 | 1.00 | 0.80 | 0.70 | 0.20 | 0.30 |
| maintain | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| disable | 0.80 | 1.00 | 0.30 | 0.20 | 0.05 | 0.10 |

**Table A5.** Robustness to degraded HSI input and accuracy on shadowed and class-boundary test pixels (test OA, %, mean over ten seeds; saved models of Table 12, no retraining). Degradations are applied to the normalized HSI cube of the test scene: spatial Gaussian blur ($\sigma$ in pixels), illumination gain, contrast factor around the band mean, and additive Gaussian noise (standard deviation). Shadow: mean HSI brightness below half of the scene mean (share of the test set: 12.2%, 9.6% and 3.1%; * on Augsburg these dark pixels are not dominated by shadow). Boundary: another class within the 3 × 3 ground-truth neighborhood, identifiable only on Muufl (11.8%), whose ground truth is dense (—: not available).

| Dataset | Configuration | Clean | Blur $\sigma$ = 0.5/1/2 | Gain 0.7/0.5/0.3 | Contrast 0.7/0.5/0.3 | Noise 0.01/0.02/0.05 | Shadow/Boundary Pixels |
|---|---|---|---|---|---|---|---|
| Houston2013 | Baseline-Ghost | 61.2 | 60.9/60.5/59.2 | 50.5/29.7/19.2 | 61.3/57.7/43.6 | 61.0/60.6/56.5 | 34.5/— |
| Houston2013 | RAFC-Fusion | 69.9 | 69.3/68.3/66.9 | 51.2/27.0/16.3 | 64.5/54.7/36.0 | 69.3/67.0/57.4 | 39.0/— |
| Muufl | Baseline-Ghost | 67.1 | 67.1/67.1/66.7 | 56.2/41.2/25.3 | 67.6/66.1/60.3 | 67.1/67.1/67.2 | 49.4/39.6 |
| Muufl | RAFC-Fusion | 74.7 | 74.7/74.7/73.8 | 65.3/49.4/35.1 | 73.8/70.6/61.4 | 74.7/74.7/74.3 | 56.2/45.1 |
| Augsburg | Baseline-Ghost | 61.7 | 61.6/61.4/61.1 | 38.0/19.8/13.7 | 59.4/54.2/41.7 | 61.7/61.6/60.4 | 90.3 */— |
| Augsburg | RAFC-Fusion | 66.0 | 65.8/65.1/64.4 | 40.2/22.0/13.8 | 62.6/54.0/38.7 | 65.8/65.5/62.2 | 86.6 */— |

## References


1. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1416–1427. [CrossRef]
2. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]
3. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4340–4354. [CrossRef]
4. Rehman, M.Z.U.; Islam, S.M.S.; Ulhaq, A.; Janjua, N.; Blake, D. Multimodal land use classification: Harnessing hsi and lidar integration. In Proceedings of the IEEE 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 27–29 November 2024; pp. 655–661.
5. Zhang, M.; Li, W.; Tao, R.; Li, H.; Du, Q. Information fusion for classification of hyperspectral and LiDAR data using IP-CNN. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5506812. [CrossRef]
6. Roy, S.K.; Deria, A.; Hong, D.; Ahmad, M.; Plaza, A.; Chanussot, J. Hyperspectral and LiDAR Data Classification Using Joint CNNs and Morphological Feature Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5530416. [CrossRef]
7. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [CrossRef]
8. Wang, A.; Lei, G.; Dai, S.; Wu, H.; Iwahori, Y. Multiscale Attention Feature Fusion Based on Improved Transformer for Hyperspectral Image and LiDAR Data Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 4124–4140. [CrossRef]
9. Ding, K.; Lu, T.; Fu, W.; Li, S.; Ma, F. Global–Local Transformer Network for HSI and LiDAR Data Joint Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5541213. [CrossRef]
10. Wang, X.; Feng, Y.; Song, R.; Mu, Z.; Song, C. Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data. *Inf. Fusion* **2022**, *82*, 1–18. [CrossRef]
11. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1577–1586.
12. Paoletti, M.E.; Haut, J.M.; Pereira, N.S.; Plaza, J.; Plaza, A. Ghostnet for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10378–10393. [CrossRef]
13. Qin, D.; Leichner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B. MobileNetV4: Universal models for the mobile ecosystem. In *Proceedings of the European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2024; pp. 78–96.
14. Wang, Y.; Liu, L.; Xiao, J.; Yu, D.; Tao, Y.; Zhang, W. MambaHSI+: Multidirectional State Propagation for Efficient Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4411414. [CrossRef]
15. Ahmad, M.; Distefano, S.; Khan, A.M.; Mazzara, M.; Li, C.; Li, H.; Aryal, J.; Ding, Y.; Vivone, G.; Hong, D. A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers and mamba models. *Neurocomputing* **2025**, *644*, 130428. [CrossRef]
16. Xu, Y.; Mao, Y.; Li, H.; Shen, J.; Xu, X.; Wang, S.; Zaman, S.; Ding, Z.; Wang, Y. A deep learning model based on RGB and hyperspectral images for efficiently detecting tea green leafhopper damage symptoms. *Smart Agric. Technol.* **2025**, *10*, 100817. [CrossRef]

17. An, W.; Wang, H.; Sun, Q.; Xu, J.; Dai, Q.; Zhang, L. A PID controller approach for stochastic optimization of deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8522–8531.
18. Chen, S.; Liu, J.; Wang, P.; Xu, C.; Cai, S.; Chu, J. Accelerated optimization in deep learning with a proportional-integral-derivative controller. *Nat*. *Commun*. **2024**, *15*, 10263. [CrossRef] [PubMed]
19. Dai, M.; Zhang, Z.; Lai, X.; Lin, X.; Wang, H. PID controller-based adaptive gradient optimizer for deep neural networks. *IET Control*. *Theory Appl*. **2023**, *17*, 2032–2037. [CrossRef]
20. Wang, X.; Guan, Z.; Qian, W.; Cao, J.; Ma, R. PID Controller-Driven Network for Image Fusion. *IEEE Trans*. *Multimed*. **2025**, *27*, 5977–5988. [CrossRef]
21. Elsken, T.; Metzen, J.H.; Hutter, F. Neural architecture search: A survey. *J*. *Mach*. *Learn*. *Res*. **2019**, *20*, 1–21.
22. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci*. *Remote Sens*. *Lett*. **2020**, *17*, 277–281. [CrossRef]
23. Zhao, X.; Tao, R.; Li, W.; Li, H.C.; Du, Q.; Liao, W.; Philips, W. Joint Classification of Hyperspectral and LiDAR Data Using Hierarchical Random Walk and Deep CNN Architecture. *IEEE Trans*. *Geosci*. *Remote Sens*. **2020**, *58*, 7355–7370. [CrossRef]
24. Jamali, A.; Roy, S.K.; Hong, D.; Atkinson, P.M.; Ghamisi, P. Attention Graph Convolutional Network for Disjoint Hyperspectral Image Classification. *IEEE Geosci*. *Remote Sens*. *Lett*. **2024**, *21*, 5503005. [CrossRef]
25. Ahmad, M.; Mazzara, M.; Distifano, S. Importance of disjoint sampling in conventional and transformer models for hyperspectral image classification. *arXiv* **2024**, arXiv:2404.14944.
26. Jiao, T.; Guo, C.; Feng, X.; Chen, Y.; Song, J. A Comprehensive Survey on Deep Learning Multi-Modal Fusion: Methods, Technologies and Applications. *Comput*. *Mater*. *Contin*. **2024**, *80*, 1–35. [CrossRef]
27. Rehman, M.Z.U.; Islam, S.M.S.; Blake, D.; Ulhaq, A.; Janjua, N. Deep learning for land use classification: A systematic review of HS-LiDAR imagery. *Artif*. *Intell*. *Rev*. **2025**, *58*, 272. [CrossRef]
28. Dalla Mura, M.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Trans*. *Geosci*. *Remote Sens*. **2010**, *48*, 3747–3762. [CrossRef]
29. Wang, J.; Zhang, J.; Guo, Q.; Li, T. Fusion of Hyperspectral and Lidar Data Based on Dual-Branch Convolutional Neural Network. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3388–3391.
30. Wang, H.; Li, Q.; Wu, L. MVSTF-Net: Multimodal remote sensing image classification based on variable scale convolution and transformer cross fusion strategies. *Int*. *J*. *Remote Sens*. **2025**, *46*, 3584–3617. [CrossRef]
31. Sun, E.; Cui, Y.; Liu, P.; Yan, J. A decade of deep learning for remote sensing spatiotemporal fusion: Advances, challenges, and opportunities. *Inf*. *Fusion* **2026**, *126*, 103675. [CrossRef]
32. Zhang, Y.; Gao, H.; Chen, Z.; Fei, S.; Zhou, J.; Ghamisi, P.; Zhang, B. Adaptive multi-stage fusion of hyperspectral and LiDAR data via selective state space models. *Inf*. *Fusion* **2026**, *125*, 103488. [CrossRef]
33. Zhang, Y.; Gao, H.; Zhou, J.; Zhang, C.; Ghamisi, P.; Xu, S.; Li, C.; Zhang, B. A cross-modal feature aggregation and enhancement network for hyperspectral and LiDAR joint classification. *Expert Syst*. *Appl*. **2024**, *258*, 125145. [CrossRef]
34. Li, L.; Liu, M.; Ma, L.; Han, L. Cross-Modal feature description for remote sensing image matching. *Int*. *J*. *Appl*. *Earth Obs*. *Geoinf*. **2022**, *112*, 102964. [CrossRef]
35. Mou, L.; Lu, X.; Li, X.; Zhu, X.X. Nonlocal graph convolutional networks for hyperspectral image classification. *IEEE Trans*. *Geosci*. *Remote Sens*. **2020**, *58*, 8246–8257. [CrossRef]
36. Wang, Q.; Liu, F.; Wang, M.; Wang, L.; Huang, J.; Shen, T. Spectral–Spatial Hypergraph Convolutional Network for Hyperspectral Image Classification. *IEEE J*. *Sel*. *Top*. *Appl*. *Earth Obs*. *Remote Sens*. **2025**, *18*, 23744–23756. [CrossRef]
37. Lu, T.; Ding, K.; Fu, W.; Li, S.; Guo, A. Coupled adversarial learning for fusion classification of hyperspectral and LiDAR data. *Inf*. *Fusion* **2023**, *93*, 118–131. [CrossRef]
38. Wang, A.; Dai, S.; Wu, H.; Iwahori, Y. Multimodal Semantic Collaborative Classification for Hyperspectral Images and LiDAR Data. *Remote Sens*. **2024**, *16*, 3082. [CrossRef]
39. Yang, J.; Du, B.; Liu, R.; Huang, J.; Chang, S.; Zhang, L. Exploring Directional Sparsity and Cross-Modal Contrast-Consensus: MERGE-Net for HSI-SAR/LiDAR Joint Classification. *IEEE Trans*. *Geosci*. *Remote Sens*. **2026**, *64*, 5532316. [CrossRef]
40. Tu, B.; Ren, Q.; Li, J.; Cao, Z.; Chen, Y.; Plaza, A. NCGLF2: Network combining global and local features for fusion of multisource remote sensing data. *Inf*. *Fusion* **2024**, *104*, 102192. [CrossRef]
41. Pande, S.; Banerjee, B. Self-supervision assisted multimodal remote sensing image classification with coupled self-looping convolution networks. *Neural Netw*. **2023**, *164*, 1–20. [CrossRef] [PubMed]
42. Wang, Q.; Zhou, B.; Zhang, J.; Xie, J.; Wang, Y. Joint classification of hyperspectral images and lidar data based on dual-branch transformer. *Sensors* **2024**, *24*, 867. [CrossRef] [PubMed]
43. Suo, J.; Zhang, W.; Gong, J.; Yuan, X.; Brady, D.J.; Dai, Q. Computational Imaging and Artificial Intelligence: The Next Revolution of Mobile Vision. *Proc*. *IEEE* **2023**, *111*, 1607–1639. [CrossRef]

44. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
45. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
46. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
47. Meng, Z.; Yan, Q.; Zhao, F.; Chen, G.; Hua, W.; Liang, M. Global-local multi-granularity transformer for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *18*, 112–131. [CrossRef]
48. Zheng, C. Iformer: Integrating ConvNet and transformer for mobile application. *arXiv* **2025**, arXiv:2501.15369.
49. Nottebaum, M.; Dunnhofer, M.; Micheloni, C. LowFormer: Hardware Efficient Design for Convolutional Transformer Backbones. In Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 26 February–6 March 2025; pp. 7008–7018.
50. Ren, P.; Xiao, Y.; Chang, X.; Huang, P.-Y.; Li, Z.; Chen, X.; Wang, X. A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–34. [CrossRef]
51. Zhang, S.; Xie, L. A PID controller approach for adaptive probability-dependent gradient decay in model calibration. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2024; pp. 114744–114764.
52. Shi, W. Research and Design of PID Control System Based on BP Neural Network. In *Proceedings of the 2024 4th International Conference on Artificial Intelligence, Automation and High Performance Computing*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 320–324.
53. Hao, X.; Xin, Z.; Huang, W.; Wan, S.; Qiu, G.; Wang, T.; Wang, Z. Deep reinforcement learning enhanced PID control for hydraulic servo systems in injection molding machines. *Sci. Rep.* **2025**, *15*, 23005. [CrossRef] [PubMed]
54. Adeline, K.R.; Chen, M.; Briottet, X.; Pang, S.; Paparoditis, N. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 21–38. [CrossRef]
55. Cressie, N. *Statistics for Spatial Data*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
56. Rasti, B.; Ghamisi, P.; Gloaguen, R. Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3997–4007. [CrossRef]
57. Wu, X.; Hong, D.; Chanussot, J. Convolutional neural networks for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5517010. [CrossRef]
58. Roy, S.; Deria, A.; Hong, D.; Rasti, B.; Plaza, A.; Chanussot, J. Multimodal Fusion Transformer for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5515620. [CrossRef]
59. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7738–7749. [CrossRef]
60. Marmanis, D.; Schindler, K.; Wegner, J.D.; Galliani, S.; Datcu, M.; Stilla, U. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 158–172. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.