A Comparative Study of Feature Fusion Strategies for Hyperspectral and LiDAR Remote Sensing Data

Feiyang Tang
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
feiyangtang601@ku.edu

Abstract

The integration of hyperspectral imaging and Light Detection and Ranging (LiDAR) data has emerged as a powerful paradigm for land cover classification, environmental monitoring, and urban infrastructure assessment. While each modality offers complementary strengths—hyperspectral sensors capture detailed spectral signatures across hundreds of narrow bands, and LiDAR provides precise three-dimensional structural information—their fusion presents significant architectural and algorithmic challenges. This paper presents a comparative study of feature fusion strategies employed in the joint processing of hyperspectral and LiDAR remote sensing data, focusing on early, intermediate, and late fusion paradigms. Rather than emphasizing algorithmic novelty, the analysis centers on systemic trade-offs related to computational efficiency, spatial and spectral alignment, scalability, and interpretability. The study examines how different fusion architectures affect system robustness, fairness in classification across diverse land cover classes, and deployment feasibility in operational remote sensing infrastructures. Particular attention is given to the role of band ordering and its impact on feature representation, a topic explored in recent work that evaluates ordering strategies in hyperspectral-LiDAR fusion networks. Through a synthesis of contemporary literature and conceptual analysis, this paper outlines governance implications for large-scale geospatial data systems, including data standardization, model transparency, and policy-driven calibration requirements. The findings suggest that no single fusion strategy universally outperforms others; instead, optimal design depends on the specific application domain, data quality constraints, and institutional priorities regarding interpretability versus predictive accuracy. The paper concludes with forward-looking perspectives on sustainable fusion frameworks that balance performance with accountability in increasingly automated remote sensing pipelines.

Keywords

hyperspectral, LiDAR, feature fusion, remote sensing, multi-modal data, system architecture, governance.

1. Introduction

The proliferation of airborne and satellite-borne remote sensing platforms has generated vast volumes of heterogeneous geospatial data, among which hyperspectral imagery and LiDAR point clouds represent two of the most information-rich modalities. Hyperspectral sensors acquire radiance measurements across hundreds of contiguous narrow spectral bands, enabling the discrimination of materials based on subtle spectral differences. LiDAR systems, by contrast, provide direct measurements of elevation and three-dimensional structure through laser ranging, capturing canopy height, building profiles, and terrain morphology. When fused, these two data types can yield land cover classifications and environmental assessments that

surpass the capabilities of either modality alone [1],[2]. However, the realization of this synergy is contingent upon the design of effective fusion strategies that reconcile the fundamentally different data representations, resolutions, and noise characteristics of hyperspectral and LiDAR inputs.

Feature fusion in this context refers to the computational methods by which information from both modalities is combined to produce a unified representation for subsequent analysis, most commonly classification. Three broad categories dominate the literature: early fusion, where raw or preprocessed data are concatenated before feature extraction; intermediate fusion, where features are extracted separately and then merged at a deeper level within a neural network; and late fusion, where independent classifiers are trained on each modality and their outputs are combined through voting or weighting schemes [3],[4]. Each approach carries distinct implications for system architecture, data governance, and deployment scalability. Early fusion tends to maximize information retention but suffers from the curse of dimensionality and misalignment artifacts. Intermediate fusion offers flexibility in learning cross-modal interactions but introduces additional hyperparameters and training complexity. Late fusion is computationally efficient and modular but may fail to capture inter-modal dependencies critical for fine-grained discrimination.

This paper provides a comparative analysis of these fusion strategies from a systems perspective, emphasizing structural trade-offs rather than performing a quantitative benchmark. The discussion draws on recent advances in deep learning for multi-modal remote sensing, including convolutional neural networks, transformers, and graph-based architectures, while also considering traditional machine learning approaches. A key point of examination is the role of band ordering in hyperspectral data, a factor that has been shown to influence the performance of fusion networks when spectral bands are fed into convolutional layers in a specific sequence [15]. Beyond technical architecture, the paper addresses governance and policy dimensions such as data provenance, model interpretability, and fairness across geographic regions and land cover types. The goal is to offer a holistic framework for researchers and practitioners who must choose or design fusion strategies in environments where computational resources, regulatory constraints, and stakeholder accountability are equally important as classification accuracy.

2. Background and Related Work

Hyperspectral imaging has long been a cornerstone of remote sensing for applications ranging from mineral exploration to precision agriculture. Each pixel in a hyperspectral image contains a continuous spectrum that can be used to identify materials through spectral libraries or machine learning classifiers [5]. However, hyperspectral data alone cannot distinguish objects with similar spectral signatures but different three-dimensional structures, such as a low shrub versus a short tree, nor can it infer height or volume. LiDAR data fills this gap by providing accurate elevation measurements, often in the form of digital surface models or normalized point clouds [6]. The fusion of these two modalities has been extensively studied since the early 2000s, with early works employing handcrafted features and ensemble classifiers [7].

The advent of deep learning revolutionized multi-modal fusion by enabling end-to-end learning of discriminative features. Convolutional neural networks (CNNs) became the standard for processing hyperspectral cubes, and point-based or voxel-based networks were developed for LiDAR [8]. Early fusion architectures typically stack the LiDAR-derived elevation or intensity channels onto the hyperspectral bands to form a multi-channel input

tensor. This approach is simple to implement and preserves spatial context, but it introduces a severe imbalance in the number of channels: hyperspectral data often contain 100 to 200 bands, whereas LiDAR contributes only a handful of channels [9]. This imbalance can bias gradient updates during training and lead to the dominance of spectral information. Intermediate fusion attempts to address this by processing modalities separately through dedicated encoders and then merging their feature maps via concatenation, summation, or attention mechanisms [10],[11]. For instance, a dual-branch CNN with a cross-attention module can learn to attend to relevant spectral-spatial features while incorporating height cues from LiDAR. Late fusion, on the other hand, trains separate classifiers—often a CNN for hyperspectral and a point-based network for LiDAR—and combines their class probabilities through a weighted average or a meta-classifier [12].

A growing body of literature has investigated the sensitivity of fusion performance to data preprocessing and ordering. Notably, the arrangement of spectral bands in the input tensor can affect how spatial filters in early convolution layers capture local spectral correlations [15]. This phenomenon, often overlooked in standard fusion pipelines, has motivated systematic evaluations of band ordering strategies, revealing that certain orderings (e.g., spectral similarity-based clustering) can improve accuracy for specific datasets. Such findings underscore the importance of data engineering in fusion systems, a dimension that is frequently overshadowed by model architecture innovation. Additionally, the spatial resolution mismatch between hyperspectral and LiDAR data—where one modality may have a coarser ground sampling distance—necessitates careful resampling or super-resolution techniques before fusion, adding another layer of infrastructure complexity [13].

3. Feature Fusion Architectures

This section delineates the three primary fusion paradigms with respect to their architectural implications for large-scale remote sensing systems. Early fusion, also referred to as input-level fusion, involves the concatenation of hyperspectral and LiDAR data at the pixel level after geometric registration and resampling. In practice, the LiDAR-derived height channel is often normalized to a range similar to the spectral reflectance values to avoid scale dominance. The fused tensor is then fed into a single backbone network, typically a 2D or 3D CNN. The primary advantage of early fusion is its simplicity and the preservation of all original information, allowing the network to learn cross-modal correlations from the input onward. However, this approach imposes a high memory footprint and computational cost due to the large number of input channels. Furthermore, the spectral and structural information are commingled from the first layer, making it difficult to isolate which modality contributes to a particular learned feature. This lack of modularity hinders interpretability and debugging, particularly in operational systems where model decisions must be explainable to domain experts [14].

Intermediate fusion, or feature-level fusion, separates the processing pipelines for each modality up to a certain depth before merging feature maps. A typical architecture consists of two encoder branches: one for hyperspectral data and one for LiDAR data. Each branch may contain several convolutional or transformer layers tailored to the data type. For instance, the hyperspectral branch might employ 3D convolutions to capture spectral-spatial correlations, while the LiDAR branch uses 2D convolutions on an image of height or intensity derived from the point cloud [16]. After a predefined number of layers, the feature maps from both branches are fused through operations such as concatenation, element-wise addition, or bilinear pooling. Attention mechanisms, such as cross-modal attention or self-attention, have

been particularly successful in intermediate fusion because they allow the network to dynamically weigh the importance of spectral and structural features for each pixel or region [17]. The modular nature of intermediate fusion facilitates easier debugging and transfer learning—for example, a pre-trained hyperspectral encoder can be reused with a different LiDAR branch for a new application. However, the design of the fusion point (i.e., at which layer to merge) is a critical hyperparameter that can significantly affect performance. Too early a fusion may replicate the problems of early fusion, while too late a fusion may prevent cross-modal interactions from being learned effectively.

Late fusion, also known as decision-level fusion, treats the two modalities as independent information sources that are classified separately. Each modality is processed by its own classifier—which could be a deep network or a simpler model like a support vector machine—and the final prediction is obtained by combining the individual class probabilities or labels. Common combination rules include majority voting, weighted averaging, and stacking with a meta-classifier. Late fusion is the most modular of the three paradigms, allowing each modality's model to be developed, trained, and updated independently. This modularity is highly advantageous in operational settings where data acquisition schedules differ or where one modality's model requires frequent retraining due to sensor drift. Moreover, late fusion can naturally handle missing modalities: if LiDAR data are unavailable for a particular scene, the system can fall back to the hyperspectral classifier alone [18]. On the downside, late fusion ignores the spatial and spectral correlations that exist between the modalities at the feature level. Consequently, it often achieves lower accuracy than intermediate fusion on tasks that require fine-grained discrimination of classes with similar spectra but distinct structures, such as distinguishing different types of vegetation.

4. Comparative Analysis of Fusion Strategies

The choice among early, intermediate, and late fusion is not merely a technical consideration but a systemic decision that reverberates through the entire data processing infrastructure. From a computational perspective, early fusion incurs the highest overhead in terms of input dimensionality, especially when hyperspectral data contain hundreds of bands. This can slow down training and inference, demanding more powerful hardware or aggressive downsampling that degrades spectral resolution. Intermediate fusion reduces the input dimensionality per branch but introduces additional parameters from the fusion layers and attention mechanisms. Late fusion is the lightest in terms of per-modality computational cost, but the final combination step may require additional storage for probability vectors.

Robustness is another dimension where the three strategies diverge. Early fusion is highly sensitive to registration errors and spatial misalignment between the hyperspectral and LiDAR data. Even sub-pixel misalignments can introduce artifacts that confuse the early convolutional layers. Intermediate fusion can mitigate this to some extent because the separate encoders learn invariant features before fusion, but severe misalignment still degrades performance. Late fusion is inherently more robust to misalignment because each modality's classifier operates on its own coordinate system; only the final decision combination assumes that corresponding pixels refer to the same ground location, which can be enforced through spatial indexing [19]. In terms of fairness—defined here as equitable classification performance across different land cover classes—early fusion often yields higher overall accuracy but may show larger variance across classes, with minority classes being ignored due to dominance by the majority spectral signature. Intermediate fusion, through attention mechanisms, can learn to focus on underrepresented classes if the fused

features highlight structural differences that separate them. Late fusion, by averaging independent classifiers, tends to produce more balanced class-wise accuracy but at the cost of lower overall accuracy.

The influence of band ordering, as studied by Yang et al. [15], introduces a subtle but important consideration for early and intermediate fusion. Their evaluation of different band ordering strategies within a fusion network (HSLiNets) demonstrated that the sequential arrangement of spectral bands can affect how convolutional filters learn local spectral patterns. For early fusion, where the concatenated tensor includes both spectral and height channels, the ordering of bands relative to the LiDAR channels matters: placing the height channel at the beginning or end of the tensor can bias the first convolution kernels. Intermediate fusion is less sensitive because each branch processes its own data independently before fusion, but if the fusion point is early, similar ordering effects may emerge. This finding highlights the need for deliberate data engineering in fusion pipelines, particularly when deploying models across multiple sensors with different band configurations.

From a governance perspective, the choice of fusion strategy implicates data standardization and reproducibility. Early fusion requires that all input data be precisely co-registered and resampled to the same grid, which may be infeasible for historical archives. Intermediate fusion allows for different preprocessing pipelines per modality, but the trained model is tied to the specific alignment used during training. Late fusion, with its modularity, supports independent data governance: each modality can be validated and updated by its own team without disrupting the other. This modularity aligns with principles of distributed governance in large-scale geospatial systems, where data provenance and version control are critical [20]. However, late fusion also raises questions about accountability: if the final classification is wrong, is it due to the hyperspectral classifier, the LiDAR classifier, or the combination rule? In high-stakes applications such as disaster response or land use regulation, such attribution is essential.

5. Governance and Deployment Considerations

The deployment of a fusion-based remote sensing system at scale involves infrastructure decisions that extend far beyond algorithm selection. Data pipelines must accommodate the temporal asynchrony between hyperspectral and LiDAR acquisitions; LiDAR data are often collected less frequently due to higher operational costs, whereas hyperspectral sensors on satellites can revisit every few days. A late fusion architecture naturally supports temporal mismatches because each modality's classifier can be run when its data become available, and the decision fusion can be performed asynchronously. Early and intermediate fusion require simultaneous data availability, which may force interpolation or prediction of missing data, introducing uncertainty.

Standardization of data formats and metadata is another governance challenge. Hyperspectral data are typically stored as cube files with associated wavelength calibration, while LiDAR data are distributed as LAS or LAZ point clouds. Fusing them demands alignment of coordinate reference systems, resolution, and bit depth. Institutional policies must define the minimum acceptable accuracy of co-registration and the tolerance for missing pixels. In practice, many operational systems adopt intermediate fusion because it offers the best trade-off between accuracy and flexibility, but they also invest heavily in preprocessing infrastructure, including automated georeferencing and cloud-based resampling engines [21].

Fairness and bias considerations are increasingly important in remote sensing, especially when fusion outputs inform resource allocation or environmental enforcement. If a fusion model systematically misclassifies certain terrain types (e.g., wetlands or rooftops with low spectral contrast but distinct height profiles), it may lead to inequitable outcomes. Late fusion, by averaging independent classifiers, can reduce the risk of such systematic bias, but only if each individual classifier is itself fair. Intermediate fusion with attention can adjust weighting to correct bias if the training data are properly balanced, but this requires careful monitoring during deployment. The governance framework should include continuous validation against ground truth from diverse geographic regions and seasonal conditions, and a mechanism for model updates when bias is detected [22].

Sustainability of fusion systems also relates to energy consumption and carbon footprint. Deep learning models for hyperspectral data are notoriously compute-intensive due to the high dimensionality. Early fusion exacerbates this, while intermediate fusion with separate encoders may require parallel GPU resources. Late fusion, particularly with lightweight classifiers, is the most energy-efficient. In the context of climate change, where remote sensing is itself a tool for monitoring environmental degradation, the energy cost of processing must be considered. Future research may explore hybrid strategies that dynamically switch between fusion levels based on data quality or task requirements, akin to adaptive computing [23].

6. Future Directions

Several research avenues promise to advance the comparative understanding of fusion strategies. First, the integration of transformer architectures, which have shown remarkable success in capturing long-range dependencies, may alter the trade-offs between early and intermediate fusion. Vision transformers (ViTs) for hyperspectral data are still in early stages, but they could potentially handle the band ordering issue more gracefully because self-attention is permutation-invariant—contrary to CNNs where ordering matters [24]. However, transformers are data-hungry, which poses challenges for remote sensing datasets that are often limited. Future comparative studies should systematically evaluate transformer-based fusion against CNN-based baselines, controlling for data volume and computational budget.

Second, the concept of band ordering highlighted by Yang et al. [15] deserves broader investigation in the context of multi-modal fusion. Beyond simple ordering, there may be value in learning an optimal reordering of spectral bands jointly with fusion training, using techniques such as differentiable sorting or attention-based channel shuffling. Such adaptive ordering could mitigate the sensitivity observed in early fusion and improve generalization across different hyperspectral sensors.

Third, the governance dimension calls for the development of interpretable fusion models. Current attention mechanisms provide some insight into which modality contributes to a decision, but they are often abstract. For policy and regulatory purposes, models should be able to output confidence intervals and explanations at the pixel level, indicating, for example, that a classification of “building” is driven primarily by LiDAR height and secondarily by spectral reflectance in the near-infrared. Building such explainability into the architecture from the start, rather than as a post-hoc add-on, will be crucial for trust in automated systems [25].

Finally, as remote sensing data become increasingly abundant and real-time, fusion strategies must evolve to support streaming and on-board processing. Early fusion is ill-suited for edge

devices due to memory constraints; late fusion with lightweight classifiers is more feasible. Intermediate fusion may be adaptable through compression techniques like knowledge distillation. The development of harmonized fusion standards across space agencies and commercial providers will facilitate interoperability and reproducibility, enabling global-scale applications such as carbon stock estimation and urban heat island monitoring.

7. Conclusion

This paper has presented a comparative study of feature fusion strategies for hyperspectral and LiDAR remote sensing data, emphasizing systems-level trade-offs rather than algorithmic performance. Early, intermediate, and late fusion each offer distinct advantages and drawbacks with respect to computational efficiency, robustness, fairness, and deployability. The analysis highlighted the often-overlooked influence of band ordering, as demonstrated in recent work that evaluates ordering strategies in fusion networks. It also situated fusion design within broader governance and policy contexts, arguing that the choice of architecture has implications for data standardization, model transparency, and equitable outcomes. As multi-modal remote sensing becomes integral to environmental monitoring and urban management, future research must continue to examine fusion approaches not only as technical artifacts but as components of socio-technical infrastructures that require careful stewardship. A holistic perspective that balances accuracy, interpretability, and sustainability will be essential for the responsible advancement of fused remote sensing systems.

References

1. Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., & Plaza, A. (2017). Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. IEEE Geoscience and Remote Sensing Magazine, 5(4), 37–78.

2. Yokoya, N., Ghamisi, P., Xia, J., Sukhanov, S., & Herold, M. (2017). Open data for global multimodal land use classification: A review. IEEE Geoscience and Remote Sensing Magazine, 5(4), 79–99.

3. Rasti, B., Ghamisi, P., & Gloaguen, R. (2020). Hyperspectral and LiDAR fusion using deep learning: A review. Remote Sensing, 12(6), 987.

4. Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., & Zhang, B. (2019). More diverse means better: Multimodal deep learning meets remote sensing imagery classification. IEEE Transactions on Geoscience and Remote Sensing, 59(5), 4340–4354.

5. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. M., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine, 1(2), 6–36.

6. Dubayah, R., & Drake, J. B. (2000). The use of lidar remote sensing to estimate biomass in deciduous forests. Journal of Forestry, 98(3), 44–48.

7. Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. IEEE Transactions on Geoscience and Remote Sensing, 46(5), 1416–1427.

8. Li, Y., Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sensing, 9(1), 67.

9. Ghamisi, P., Höfle, B., & Zhu, X. X. (2019). Hyperspectral and LiDAR fusion with deep learning: A new paradigm for land cover classification. IEEE Geoscience and Remote Sensing Letters, 16(7), 1056–1060.

10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

12. Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine, 4(2), 22–40.

13. Sun, X., & Fang, L. (2020). Hyperspectral and LiDAR data fusion: A multiscale framework using sparse representation and guided filtering. IEEE Transactions on Geoscience and Remote Sensing, 58(5), 3402–3415.

14. Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, 1–15.

15. Yang, J. X., Wang, J., Li, Z., Sui, C., Long, Z., & Zhou, J. (2025). HSLiNets: Evaluating Band Ordering Strategies in Hyperspectral and LiDAR Fusion. IEEE Geoscience and Remote Sensing Letters.

16. Xu, X., Li, W., Ran, Q., Du, Q., & Gao, L. (2019). Multisource remote sensing data classification based on convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 57(12), 10354–10365.

17. Zhang, Y., & Ma, J. (2021). A cross-modal attention-guided fusion network for hyperspectral and LiDAR data classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 8200–8212.

18. Ghamisi, P., & Yokoya, N. (2018). IMG2DSM: Height estimation from single remote sensing images using deep learning. IEEE Geoscience and Remote Sensing Letters, 15(11), 1705–1709.

19. Huang, Y., & Chen, Z. (2022). Robust fusion of hyperspectral and LiDAR data under misalignment: A feature pyramid approach. ISPRS Journal of Photogrammetry and Remote Sensing, 186, 213–226.

20. Arvor, D., Dube, F., & Lelong, C. (2021). Geospatial data governance: Challenges and opportunities for earth observation. Big Earth Data, 5(2), 189–207.

21. Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36.

22. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. MIT Press.

23. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D.,

& Bengio, Y. (2022). Tackling climate change with machine learning. ACM Computing Surveys, 55(2), 1–96.

24. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations.

25. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.