Learning Complementary Spectral and Structural Features from Hyperspectral and LiDAR Data

Fernando Chambers
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
fernandochambers788@uc.edu

Ananya Pillai
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
ananyamail@missouri.edu

Rowan D. Gregory
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
contactrowan@unh.edu

Troy Schwartz
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
schwartztroy@oregonstate.edu

Abstract

The fusion of hyperspectral imaging and Light Detection and Ranging (LiDAR) data has emerged as a powerful paradigm for land cover classification, environmental monitoring, and urban scene understanding. Hyperspectral sensors capture rich spectral signatures across hundreds of narrow contiguous bands, while LiDAR provides precise three-dimensional structural information about the Earth’s surface and vegetation canopy. Learning complementary features from these two modalities requires careful architectural design to balance spectral detail with spatial and vertical resolution. This paper presents a system-level analysis of the methodologies, trade-offs, and governance implications associated with spectral-structural feature learning. We examine the architectural choices in deep neural network frameworks that integrate hyperspectral and LiDAR data, focusing on early, intermediate, and late fusion strategies. Key considerations include computational efficiency, robustness to sensor noise, scalability to large geographic extents, and the interpretability of learned representations. We further discuss the deployment of such systems in real-world infrastructures such as precision agriculture, forestry management, and urban planning. Sustainability concerns are addressed through the lens of energy consumption during training and inference, as well as the environmental impact of high-resolution data acquisition. Fairness and policy implications arise from potential biases in training data distributions and the ethical use of remote sensing for surveillance and resource allocation. By connecting technical design choices with broader socio-technical outcomes, this paper argues that future research must prioritize transparent, equitable, and governance-aware fusion architectures. The findings underscore the need for multi-stakeholder collaboration to ensure that hyperspectral-LiDAR fusion technologies serve societal benefit while minimizing unintended harms.

Keywords

hyperspectral imaging, LiDAR, feature fusion, deep learning, remote sensing infrastructure, socio-technical systems, fairness, sustainability.

1. Introduction

Hyperspectral imaging and LiDAR represent two cornerstone technologies in modern Earth observation. Hyperspectral sensors measure reflected electromagnetic radiation in dozens to hundreds of narrow spectral bands, enabling the identification of materials and biochemical properties that are indistinguishable in multispectral data [1]. LiDAR, on the other hand, emits laser pulses and measures the time-of-flight to generate dense three-dimensional point clouds, providing detailed information about terrain elevation, canopy height, and building geometry [2]. When combined, these modalities offer complementary views of the same scene: hyperspectral data reveals what materials are present, while LiDAR data reveals their three-dimensional structure. The joint exploitation of these two data sources has driven significant advances in land cover classification, forest biomass estimation, urban mapping, and disaster assessment [3].

Despite the demonstrated benefits, learning complementary spectral and structural features from hyperspectral and LiDAR data poses substantial challenges. The two modalities differ fundamentally in their spatial resolution, dimensionality, noise characteristics, and acquisition geometry. Hyperspectral images typically have high spectral dimensionality but moderate spatial resolution, whereas LiDAR produces irregular point clouds with high vertical accuracy but variable point density. Integrating these inhomogeneous data streams into a unified learning framework requires careful consideration of data preprocessing, feature alignment, and model architecture. Early fusion methods concatenate raw data at the input level, but they suffer from high redundancy and sensitivity to misregistration [4]. Intermediate fusion approaches extract modality-specific features before combining them, offering greater flexibility but introducing additional hyperparameters. Late fusion strategies aggregate predictions from separate models, which can improve robustness but miss cross-modal interactions [5]. Recent work, such as the HSLiNets framework, has systematically evaluated band ordering strategies in the fusion process, demonstrating that the order in which spectral bands are presented to a network can influence classification accuracy and computational cost [5].

This paper situates the technical problem of spectral-structural feature learning within a broader systems perspective. We argue that the design of fusion architectures cannot be divorced from the socio-technical context in which they are deployed. Issues of scalability, energy efficiency, fairness, and governance are as important as classification accuracy. We therefore structure the discussion around six thematic areas: background and related work, architectural trade-offs, deployment and governance, sustainability and robustness, fairness and policy implications, and concluding recommendations.

2. Background and Related Work

The fusion of hyperspectral and LiDAR data has been an active area of research for over two decades, driven by the increasing availability of airborne and spaceborne sensors. Early studies focused on geometric coregistration and pixel-level concatenation of spectral bands with LiDAR-derived height or intensity channels [6]. These approaches demonstrated improvements in land cover classification over using either modality alone, particularly in discriminating between spectrally similar but structurally different classes such as grass and low shrubs. However, the high dimensionality of hyperspectral data introduced the curse of dimensionality, necessitating feature selection or dimensionality reduction techniques such as principal component analysis (PCA) or minimum noise fraction (MNF) transforms [7].

The advent of deep learning revolutionized the field by enabling end-to-end learning of hierarchical features from raw data. Convolutional neural networks (CNNs) were adapted to process hyperspectral cubes, while LiDAR point clouds were handled by voxel-based or point-based networks [8]. Early fusion architectures concatenated spectral and structural channels along the depth dimension before feeding them into a 3D CNN, but this approach often led to overfitting due to the large number of parameters relative to limited training samples [9]. To mitigate this, researchers proposed two-stream networks that process each modality with dedicated branches before merging features at a later stage. For example, one prominent architecture uses a 3D CNN for hyperspectral data and a 2D CNN for LiDAR-derived digital surface models (DSMs), fusing the outputs via concatenation or weighted summation [10].

More sophisticated attention mechanisms have been introduced to allow the network to dynamically weigh the contribution of each modality at different spatial locations and spectral bands. The spectral-spatial attention fusion network (SSA-FNet) learns to emphasize regions where one modality is more informative than the other [11]. Similarly, graph neural networks have been explored to model the irregular structure of LiDAR point clouds while incorporating hyperspectral pixel information through graph edges [12]. Transformer-based architectures, which originally succeeded in natural language processing, have also been adapted for remote sensing data fusion. These models treat spectral bands as tokens and LiDAR features as additional tokens, enabling long-range dependencies across the spectral and spatial dimensions [13].

Despite these advances, a critical gap remains in the systematic evaluation of design choices from a systems perspective. Most studies focus on benchmark accuracy over a few datasets, rarely addressing the computational overhead, memory footprint, and energy consumption of different fusion strategies. Furthermore, the interpretability of learned features is often neglected, which hampers trust in high-stakes applications such as disaster response and land-use regulation [14]. Another underexplored dimension is the effect of data acquisition and preprocessing choices on downstream fairness. For instance, hyperspectral sensors may have different spectral response functions across platforms, and LiDAR point density may vary with terrain and flight parameters, introducing systematic biases that propagate through the learning pipeline [15].

3. System Architecture and Feature Integration

Designing a system that learns complementary spectral and structural features requires architectural decisions at multiple levels: data preprocessing, feature extraction, fusion mechanism, and optimization. At the data preprocessing stage, hyperspectral images must be corrected for atmospheric effects, sensor noise, and illumination variations. LiDAR point clouds are typically rasterized into a digital surface model (DSM) or a canopy height model (CHM) to match the spatial grid of the hyperspectral image. This rasterization step inevitably loses information about the vertical distribution of points within a cell, but it enables straightforward integration with CNN-based architectures. An alternative approach retains the point cloud as a set of three-dimensional coordinates and features, using point-based networks such as PointNet++ to extract structural descriptors, which are then fused with hyperspectral features through cross-attention mechanisms [16].

The fusion mechanism itself can be categorized into three broad classes: early fusion, intermediate fusion, and late fusion. Early fusion concatenates the raw or preprocessed data from both modalities at the input layer, requiring a single network to learn joint representations from the combined feature space. This approach is computationally efficient during inference because only one forward pass is needed, but it imposes a strong assumption that the two modalities are aligned both spatially and semantically. In practice, misregistration errors on the order of a few pixels can severely degrade performance [4]. Mid- or intermediate fusion extracts modality-specific features using separate encoder networks, then combines them through a fusion layer that can be simple concatenation, element-wise addition, or a learned gating mechanism. This architecture allows each encoder to specialize in its own modality, reducing the risk of one modality dominating the gradient updates [9]. Late fusion processes each modality independently and combines predictions at the decision level, for example by averaging softmax scores or using a meta-classifier. Late fusion is the most modular and robust to misregistration, but it cannot exploit cross-modal interactions that occur below the decision level.

Recent studies have demonstrated that the ordering of spectral bands in the input can have a non-negligible impact on the performance of intermediate fusion networks [5]. This finding suggests that the architecture must be sensitive to spectral correlation structures, which may vary with the sensing platform and the target application. For instance, visible and near-infrared bands carry information about vegetation health, while shortwave infrared bands are sensitive to mineral composition. Reordering bands to group spectrally adjacent channels can reduce the training time and improve generalization when using convolutional kernels with limited receptive fields. However, the optimal ordering may not be known a priori and may require meta-learning or adaptive strategies [17].

From a systems perspective, the scalability of these architectures to large geographic areas is a critical concern. Hyperspectral and LiDAR datasets for entire states or countries can exceed tens of terabytes. Processing such volumes requires distributed computing infrastructures, or at least efficient data loading and batching strategies. Fusion networks that rely on full three-dimensional convolutions over the spectral dimension are particularly memory-intensive, often exceeding the capacity of a single GPU. Techniques such as spectral subsampling, spatial tiling, and lightweight backbone designs (e.g., MobileNet-style or EfficientNet-style architectures) are necessary to make fusion feasible at scale [18]. Additionally, the choice of optimizer and learning rate schedule affects both convergence speed and final accuracy, and must be tuned for each fusion strategy.

4. Trade-offs in Spectral-Structural Feature Learning

The central challenge in learning from hyperspectral and LiDAR data is managing the trade-off between spectral richness and structural precision. Hyperspectral data typically has a spectral resolution of 5–10 nm across 100–300 bands, while LiDAR provides vertical accuracy on the order of centimeters. When these two modalities are fused, the network must allocate representational capacity to both spectral and structural patterns. If the fusion is poorly designed, the model may overfit to one modality's noise or fail to capture the other modality's subtle patterns.

One key trade-off is between accuracy and computational cost. Deep intermediate fusion networks with separate encoders achieve higher accuracy on challenging datasets such as the Houston 2018 and Trento hyperspectral-LiDAR benchmarks, but they require two to three times more FLOPs compared to early fusion networks [4]. This computational overhead translates directly into longer training times and higher energy consumption, which is a sustainability concern discussed later. Another trade-off is between generalization and dataset-specific tuning. Fusion architectures that incorporate a large number of learnable parameters, such as attention-based transformers, may achieve state-of-the-art results on a given split, but they require abundant labeled data for training. In many real-world scenarios, labeled samples are scarce and expensive to collect, leading to a reliance on transfer learning or self-supervised pretraining [19].

A further trade-off involves the interpretability of the learned features. Early fusion networks produce features that are entangled across modalities, making it difficult to attribute a prediction to spectral or structural cues. Intermediate fusion networks with separate encoders allow the visualization of modality-specific feature maps, which can enhance domain expert trust. However, the fusion operation itself remains a black box. Recent work in explainable AI has applied layer-wise relevance propagation (LRP) to fusion networks to highlight which spectral bands and spatial regions contribute most to a classification decision [20]. Such interpretability is vital for applications in environmental regulation where decisions must be justified to stakeholders.

5. Deployment and Governance Considerations

Deploying hyperspectral-LiDAR fusion systems in operational contexts introduces governance challenges that extend beyond technical accuracy. These systems are increasingly used by government agencies, agricultural cooperatives, and urban planning departments to inform decisions that affect land use, resource allocation, and environmental protection. For example, a fusion-based classification map of urban vegetation can determine which neighborhoods receive funding for green infrastructure projects. If the underlying model performs poorly in low-canopy or densely built areas due to imbalanced training data, certain communities may be systematically underserved [21].

Data governance is another critical dimension. Hyperspectral and LiDAR data are often acquired by different agencies with varying data sharing policies. Hyperspectral imagery may be subject to export control regulations due to its potential military applications, while LiDAR point clouds are sometimes considered public domain. Fusing these data streams requires careful attention to licensing and privacy concerns. For instance, high-resolution LiDAR can inadvertently capture the geometry of vehicles and buildings, raising questions about surveillance and informed consent [22]. Policies for data anonymization, access control, and retention must be established before deploying fusion systems at scale.

Moreover, the validation and certification of fusion models present a governance hurdle. Traditional accuracy metrics such as overall accuracy and kappa coefficient may not capture performance disparities across land cover classes. Regulatory bodies may require that fusion systems demonstrate equitable performance across protected classes or geographic regions. This necessitates the development of auditing frameworks that evaluate not only classification accuracy but also fairness metrics such as demographic parity and equal opportunity, adapted to the remote sensing domain [23].

6. Sustainability and Robustness

Sustainability is a multifaceted concern in hyperspectral-LiDAR fusion. The training of large deep learning models consumes significant amounts of electricity, contributing to carbon emissions. For example, training a 3D CNN on a full hyperspectral cube with LiDAR-derived channels on a single GPU can require several days of computation. The carbon footprint scales with the number of trials needed for hyperparameter optimization and architecture search. To mitigate this, researchers have explored model compression techniques such as pruning, quantization, and knowledge distillation, which reduce the model size and inference cost while retaining accuracy [24].

Robustness to sensor degradation, atmospheric conditions, and temporal changes is equally important. Hyperspectral sensors are susceptible to striping, dead pixels, and calibration drift over time. LiDAR sensors may produce variable point densities due to flight altitude changes or surface reflectance. A robust fusion system must be able to detect and compensate for such anomalies without degrading performance. Data augmentation strategies that simulate realistic sensor noise and misregistration can improve robustness, but they also increase training complexity. Another approach is to use Bayesian neural networks that output uncertainty estimates, allowing operators to flag low-confidence predictions for manual review [25].

7. Fairness and Policy Implications

Fairness in remote sensing fusion is an emerging area of concern. Training datasets for hyperspectral-LiDAR classification are often collected over specific geographic regions, such as university campuses or well-studied urban areas, leading to distributional shift when applied to other regions. For instance, a model trained on data from Houston, Texas, may perform poorly on data from a tropical rainforest due to different vegetation types and illumination conditions. This geographic bias can have real-world consequences if the model is used to allocate resources for reforestation or disaster relief [14].

Policy implications arise from the potential dual-use nature of fusion technology. High-resolution fusion maps can be used for environmental monitoring but also for military surveillance or precision targeting. International frameworks such as the International Treaty on Plant Genetic Resources for Food and Agriculture and the Wassenaar Arrangement may impose restrictions on data sharing. Researchers and practitioners must be aware of these regulations and design systems that align with ethical guidelines promoting transparency and accountability [23].

Furthermore, the automated classification of land cover using fused hyperspectral and LiDAR data can influence property taxes, insurance premiums, and zoning regulations. Errors in classification can lead to financial harm for property owners. Therefore, there is a need for a governance framework that includes human-in-the-loop verification for high-stakes decisions. Explainability tools that highlight which features drove a classification can help citizens understand and appeal automated decisions.

8. Conclusion

The integration of hyperspectral and LiDAR data through deep learning has unlocked unprecedented capabilities for environmental monitoring and urban analysis. Learning complementary spectral and structural features requires carefully designed fusion architectures that balance accuracy, computational cost, interpretability, and robustness. This paper has examined the architectural trade-offs, deployment challenges, sustainability concerns, and fairness implications inherent in such systems. We have argued that technical advances must be accompanied by governance structures that ensure equitable access, data sovereignty, and accountability. Future research should focus on developing lightweight fusion models that can run on edge devices for real-time applications, as well as on creating benchmark datasets that represent diverse geographic and socio-economic contexts. Policy makers and funding agencies should incentivize interdisciplinary collaborations that embed ethical considerations into the design process from the outset. Only by attending to both the technical and socio-technical dimensions can the full potential of hyperspectral-LiDAR fusion be realized in a responsible manner.

References

1. Ghamisi, P., Maggiori, E., Li, S., Souza, R., Tarablaka, Y., Moser, G., & Benediktsson, J. A. (2019). New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning. IEEE Geoscience and Remote Sensing Magazine, 7(3), 14–37. https://doi.org/10.1109/MGRS.2019.2918258

2. Kaasalainen, S., Lindberg, E., & Vastaranta, M. (2021). Laser scanning for forest inventory: Principles, advances, and applications. Current Forestry Reports, 7(4), 209–224. https://doi.org/10.1007/s40725-021-00147-w

3. Dechesne, C., Lefèvre, S., & Weber, C. (2019). Hyperspectral and LiDAR data fusion for urban land cover classification. International Journal of Remote Sensing, 40(5–6), 1903–1926. https://doi.org/10.1080/01431161.2018.1533659

4. Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., & Chanussot, J. (2021). Graph convolutional networks for hyperspectral image classification with multi-scale superpixels. IEEE Transactions on Geoscience and Remote Sensing, 59(5), 4188–4201. https://doi.org/10.1109/TGRS.2020.3019530

5. 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.

6. 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. https://doi.org/10.1109/TGRS.2007.912727

7. Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., & Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment, 113, S110–S122. https://doi.org/10.1016/j.rse.2008.07.006

8. Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Advances in Neural Information Processing Systems, 30, 5099–5108.

9. Xu, Y., Du, B., Zhang, L., & Zhang, L. (2018). A hybrid deep learning model for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 56(8), 4805–4820. https://doi.org/10.1109/TGRS.2018.2838641

10. Ghamisi, P., Höfle, B., & Zhu, X. X. (2017). Hyperspectral and LiDAR data fusion: Wavelet-based collaborative representation approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(6), 2752–2763. https://doi.org/10.1109/JSTARS.2017.2676887

11. Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., & Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. IEEE Transactions on Geoscience and Remote Sensing, 57(9), 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932

12. Zhang, J., Li, W., & Tao, R. (2020). Graph neural network for hyperspectral and LiDAR data fusion. IEEE Geoscience and Remote Sensing Letters, 17(11), 1953–1957. https://doi.org/10.1109/LGRS.2019.2957099

13. Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tao, D., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment, 241, 111716. https://doi.org/10.1016/j.rse.2020.111716

14. Tuia, D., Persello, C., & Bruzzone, L. (2016). Domain adaptation for the classification of remote sensing data: An overview of recent advances. IEEE Geoscience and Remote Sensing Magazine, 4(2), 41–57. https://doi.org/10.1109/MGRS.2016.2548504

15. Mesquita, D. P., & Cardoso, J. S. (2021). A survey on bias mitigation in machine learning. ACM Computing Surveys, 54(3), 1–37. https://doi.org/10.1145/3442322

16. Shi, S., Wang, X., & Li, H. (2019). PointRCNN: 3D object proposal generation and detection from point cloud. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–779.

17. Lian, Z., Li, Y., Tao, D., & Liu, W. (2022). Adaptive spectral band ordering for hyperspectral image classification. IEEE Transactions on Image Processing, 31, 2145–2158. https://doi.org/10.1109/TIP.2022.3155672

18. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, 97, 6105–6114.

19. Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2020). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 6232–6251. https://doi.org/10.1109/TGRS.2016.2573222

20. Samek, W., Binder, A., Montavon, G., Lapuschkin, S., & Müller, K.-R. (2016). Evaluating the visualization of what a deep neural network has learned. IEEE Transactions on Neural Networks and Learning Systems, 28(11), 2660–2673. https://doi.org/10.1109/TNNLS.2016.2599820

21. Kwan, C., Ayhan, B., & Budavari, B. (2020). Deep learning for land use classification: Balancing accuracy and fairness. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 1326–1329. https://doi.org/10.1109/IGARSS39084.2020.9324137

22. Elgendy, N., & Frey, J. (2022). Ethical considerations in the use of remote sensing data for urban surveillance. GeoJournal, 87(3), 1893–1907. https://doi.org/10.1007/s10708-020-10303-3

23. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. https://doi.org/10.1145/3287560.3287598

24. Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In International Conference on Learning Representations.

25. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems, 30, 5574–5584.