Low-Rank Spectral Parallelism for Efficient Hyperspectral Classification

Carlos Cañada-Rostro[1], María B. García-Flores[1], Mercedes E. Paoletti[1], and Juan M. Haut[1]

[1]University of Extremadura, Cáceres, Spain

Hyperspectral image (HSI) classification must exploit hundreds of spectral bands under limited supervision. SPARC introduces spectral parallelism as an alternative scaling axis: multiple learnable residual spectral transformations generate complementary views of the same HSI patch, every view is evaluated by a shared backbone, and the resulting class logits are combined through sample-adaptive aggregation. Unlike a conventional ensemble, SPARC does not replicate the complete classifier, allowing additional spectral-view capacity while sharing the feature extractor.

The framework is evaluated on Pavia University, Houston 2013, WHU-Hi HanChuan, and WHU-Hi HongHu with five SPARC-compatible backbones and 15 paired seeds. Under leakage-controlled evaluation, 16 of 20 paired mean effects are positive. The strongest absolute result among the Holm-significant configurations is obtained on WHU-Hi HanChuan with the MoE backbone, where SPARC reaches $81.92 \pm 2.72\%$ overall accuracy (OA), compared with $77.57 \pm 3.85\%$ for the single-stream counterpart. On the same scene, SPARC improves the 3D-CNN and ViT-Lite backbones by 9.62 and 10.68 percentage points, respectively. Matched-capacity and training controls indicate that these gains are not explained by parameter growth alone, while ablations identify the learned spectral prefix as the most consistent contributor.

A second result is that the learned dense spectral-prefix operators become strongly low-rank after training. A globally frozen rank-16 truncated-SVD representation preserves all 20 final SPARC configurations without retraining. Across 300 checkpoint pairs, the pooled paired OA change is only -0.0033 percentage points and dense-versus-compressed prediction agreement is 99.94%. Table 1 shows that this preservation is consistent across all four datasets while reducing prefix parameters by $3.22\times$–$8.56\times$.

Table 1. Rank-16 post-training accuracy preservation across five backbones and 15 seeds per dataset.

Dataset	Preserved	Mean ΔOA (pp)	Agreement	Prefix comp.
Pavia University	5/5	-0.0004	99.998%	$3.22\times$
Houston 2013	5/5	$+0.0008$	99.994%	$4.50\times$
WHU-Hi HanChuan	5/5	-0.0098	99.840%	$8.56\times$
WHU-Hi HongHu	5/5	-0.0036	99.932%	$8.44\times$

The compact representation also translates into measured deployment gains on an NVIDIA H100 NVL. For the two $P = 8$ configurations in Table 2, rank-16 SPARC reduces batch-1 latency by 21.1–29.3%, raises throughput by 18.9–41.5%, and lowers peak memory, while retaining the selected dense-SPARC prediction function. This yields a practical accuracy–efficiency trade-off: spectral parallelism provides the classification gain, while post-training factorization removes most of the redundant spectral-prefix cost with essentially unchanged predictions.

Table 2. Measured rank-16 deployment gains over dense SPARC on NVIDIA H100 NVL.

Configuration	Batch-1 latency	Throughput	Peak memory
Hybrid, $P = 8$	$0.871 \rightarrow 0.616$ ms	$2431 \rightarrow 2891$ p/s	$432 \rightarrow 350$ MB
ViT-Lite, $P = 8$	$1.163 \rightarrow 0.917$ ms	$4448 \rightarrow 6293$ p/s	$452 \rightarrow 353$ MB