Proceedings of the 14th International Scientific and Practical Conference «Global and Regional Aspects of Sustainable Development»

(September 16-18, 2026). Copenhagen, Denmark

No 309

INFORMATION AND WEB TECHNOLOGIES

Regularized Correlation-Based Integration of Deep Image Descriptors for Remote Sensing Scene Recognition

Khandamov Yigitali[1]

[1]PhD student, Digital Technologies and Artificial Intelligence;
Digital Technologies and Artificial Intelligence Development Research Institute;
Uzbekistan

Abstract. Complex-object recognition in remote sensing imagery is complicated by high intra-class variability, strong inter-class similarity and the heterogeneous statistical properties of deep descriptors extracted from different convolutional models. This paper presents a regularized framework for integrating several image descriptors in a common discriminative space. Label-aware cross-correlation blocks are formed between descriptor matrices, Tikhonov regularization is applied to the block covariance matrix, and projection directions are obtained from a generalized eigenvalue problem. The projected components are concatenated into one integrated descriptor and classified by a support vector machine. Regularization is especially important when the descriptor dimension is comparable to or larger than the training sample size, because covariance blocks may become singular or ill-conditioned. Published benchmark results on NWPU-RESISC45, AID and PatternNet confirm the practical value of correlation-based multi-descriptor fusion for remote sensing scene recognition.
Keywords: *image descriptor integration, remote sensing scene classification, multi-view learning, correlation analysis, regularization, feature fusion, support vector machine.*

1. Introduction. Remote sensing scene classification is a representative problem of complex-object recognition because a single image may contain many objects, textures and spatial structures that must be summarized by one semantic label. Modern benchmarks make this task difficult through variations in scale, viewpoint, illumination and background. NWPU-RESISC45 contains 31,500 images in 45 classes and is characterized by high within-class diversity and between-class similarity [4]. AID and PatternNet provide additional large-scale benchmarks for aerial and remote sensing imagery [5, 6].

Deep convolutional networks provide strong semantic descriptors, but representations obtained from different architectures or feature levels are not identical. They may be complementary, while direct concatenation preserves redundant coordinates and enlarges the final feature space. Multi-view learning therefore seeks a common representation that preserves useful cross-view information while reducing inconsistency among views [1]. Discriminative correlation-based methods extend this idea by using class labels so that the learned common space is both correlated and useful for classification [2, 3]. The purpose of this paper is to present a regularized descriptor-integration procedure and to explain why numerical stabilization is necessary for high-dimensional deep features.

2. Research Approach. Let the training set contain n labeled images. For every image, P descriptors are formed by independent feature extractors and arranged by columns in source-specific descriptor matrices. Each source is centered and, when necessary, standardized. Class labels are represented by a membership matrix from which a label-dependent association matrix H is formed. H emphasizes within-class correspondence and removes the global mean component. For descriptor sources q and s, the supervised cross-correlation block is computed as

$$C_{qs} = \frac{X^{(q)}\, H\, \left(X^{(s)}\right)^T}{n - 1} \qquad (1)$$

The full correlation matrix is assembled from all pairwise blocks. Its diagonal companion D contains the within-source covariance blocks. In high-dimensional problems these blocks may be singular because the effective rank of a descriptor matrix can be smaller than its dimension. This is common when deep feature vectors contain many coordinates but the available training set is limited. Instead of directly inverting such a matrix, a Tikhonov term is added to the block-diagonal covariance matrix.

$$D_\varepsilon = D + \varepsilon_{reg}\, I,\ \ \varepsilon_{reg} > 0 \qquad (2)$$

The regularized matrix is positive definite for a positive regularization value, improves conditioning and reduces unstable projection coefficients. A scale-dependent regularization level can be selected relative to the magnitude of the covariance matrix and then validated on training/validation data. The projection directions are subsequently obtained from the generalized eigenvalue problem

$$A\,w = \rho\, D_\varepsilon\, w \tag{3}$$

where A contains the off-diagonal discriminative cross-correlation structure and the generalized eigenvalue measures the importance of a projection direction. The leading eigenvectors are partitioned into projection matrices, one for each descriptor source. If the same number of latent components is retained from each source, the integrated representation for the i-th image is

$$z_i = \left[\left(W^{(1)}\right)^T x_i^{(1)};\ \dots\ ;\left(W^{(P)}\right)^T x_i^{(P)} \right] \tag{4}$$

3. Regularized Descriptor Integration. The integrated descriptor is then supplied to a support vector machine, which provides a margin-based decision rule in the projected space [8]. All centering parameters, regularization values and projection matrices must be estimated only from the training data. Test descriptors are projected by the already learned matrices and are never used to update the common space. This separation prevents information leakage and preserves the validity of the reported generalization performance.

The main numerical advantage of the procedure is that descriptor fusion and stability control are handled in the same model. Without regularization, very small covariance eigenvalues can amplify noise and make the solution sensitive to minor changes in the training sample. The additive diagonal term shifts the spectrum away from zero and stabilizes the generalized eigenproblem. The method also does not require the original descriptor sources to have equal dimensions: each source is projected separately and the final representation is controlled by the chosen latent dimension. This is useful when heterogeneous CNN descriptors or features from different network levels are combined.

4. Numerical Stability and Computational Considerations. The singularity problem can be described directly from the rank of the within-source covariance blocks. Let d_q denote the dimensionality of descriptor source q. After centering n training samples, the rank of the corresponding covariance matrix cannot exceed either the descriptor dimension or $n-1$. Therefore, the following upper bound holds:

$$rank(C_{qq}) \leq min(d_q, n - 1) \quad (5)$$

Consequently, if $d_q > n - 1$, the covariance block is necessarily singular. Even when d_q is slightly smaller than n, near-linear dependencies between descriptor coordinates can make the matrix ill-conditioned. In this situation, direct inversion or an unregularized generalized eigenvalue decomposition may become sensitive to small perturbations in the training data. Tikhonov regularization shifts all eigenvalues of the covariance structure in the positive direction and therefore provides a controlled lower bound for the spectrum. The regularization parameter should be scale-aware rather than chosen as an arbitrary absolute constant. A convenient parameterization is

$$\varepsilon_{reg} = \alpha \frac{tr(D)}{d}, \ \alpha > 0 \quad (6)$$

where d is the total dimension of the block covariance matrix and α is a dimensionless control coefficient. This form scales the diagonal correction to the average variance represented by D. Small values of α preserve the original correlation geometry more closely, whereas larger values improve conditioning at the cost of stronger shrinkage. In practice, α and the latent dimension r should be selected using only the training and validation subsets. The test set must not participate in either parameter selection or estimation of the projection matrices.

The computational cost is governed mainly by the construction of cross-correlation blocks and the solution of the generalized eigenvalue problem. If the q-th descriptor has dimension d_q, computing all pairwise cross-covariance blocks requires operations proportional to n times the products d_q d_S over descriptor pairs. Let $d = \sum_q d_q$ be the total dimension before projection. A dense eigensolution may require cubic time in d and quadratic memory, which can become expensive for very high-dimensional CNN descriptors. This observation motivates three practical measures: selecting only complementary descriptor sources, keeping the latent dimension r substantially smaller than the original dimensions, and using regularization to avoid repeated factorization failures caused by singular matrices. Once the projection matrices have been learned, however, the cost of processing a new image is much lower because each descriptor is only multiplied by its fixed projection matrix and the projected components are concatenated.

5. Experimental Evidence. Published experiments with a related multi-descriptor correlation-based fusion and SVM framework reported strong performance on three remote sensing benchmarks [7]. The overall accuracies are summarized in Table 1. These values provide experimental evidence for the practical value of integrating complementary deep descriptors rather than relying on a single representation.

Table 1

Reported classification accuracy of the integrated descriptor framework

Dataset	Classes	Images	Accuracy, %
NWPU-RESISC45	45	31,500	92.75
AID	30	10,000	93.92
PatternNet	38	30,400	99.35

Source: benchmark characteristics [4-6] and integrated-model accuracy reported in [7].

6. Discussion. The results indicate that correlation-based integration is useful when different deep features capture complementary scene information. PatternNet yields the highest reported accuracy, whereas NWPU-RESISC45 remains more difficult because of its pronounced within-class diversity and between-class similarity [4]. The result does not imply that every descriptor combination is beneficial. Highly redundant sources can increase computation without adding discriminative information. Descriptor selection and integration should therefore be treated jointly: complementary sources are preferred, the latent dimension should be validated, and the regularization level should stabilize the solution without suppressing the useful covariance structure.

Conclusion. The presented procedure provides a compact framework for integrating heterogeneous deep image descriptors in a common discriminative space. Its key elements are label-aware cross-correlation analysis, source-specific projection, Tikhonov regularization of the covariance structure and final classification using the integrated descriptor. Regularization directly addresses rank deficiency that arises when high-dimensional descriptors are estimated from a limited number of training samples. The benchmark evidence reported for NWPU-RESISC45, AID and PatternNet confirms that multi-descriptor fusion can support highly accurate scene recognition. Future work can extend the same principle to nonlinear kernels, adaptive descriptor weighting and descriptor selection before integration.

References:

[1] Yan X., Hu S., Mao Y., Ye Y., Yu H. Deep multi-view learning methods: A review. Neurocomputing. 2021; 448: 106-129. DOI: 10.1016/j.neucom.2021.03.090.

[2] Kumar D., Maji P. Discriminative Deep Canonical Correlation Analysis for Multi-View Data. IEEE Trans. Neural Netw. Learn. Syst. 2023; 35: 14288-14300. DOI: 10.1109/TNNLS.2023.3277633.

[3] Yuan B., Sehra S.S., Chiu B. Multi-Scale and Multi-Network Deep Feature Fusion for Discriminative Scene Classification of High-Resolution Remote Sensing Images. Remote Sens. 2024; 16: 3961. DOI: 10.3390/rs16213961.

[4] Cheng G., Han J., Lu X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proc. IEEE. 2017; 105: 1865-1883. DOI: 10.1109/JPROC.2017.2675998.

[5] Xia G.-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L., Lu X. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. IEEE Trans. Geosci. Remote Sens. 2017; 55: 3965-3981. DOI: 10.1109/TGRS.2017.2685945.

[6] Zhou W., Newsam S., Li C., Shao Z. PatternNet: A Benchmark Dataset for Performance Evaluation of Remote Sensing Image Retrieval. ISPRS J. Photogramm. Remote Sens. 2018; 145: 197-209. DOI: 10.1016/j.isprsjprs.2018.01.004.

[7] Fazilov S., Yusupov O., Khandamov Y., Eshonqulov E., Khamidov J., Abdieva K. Remote Sensing Scene Classification via Multi-Feature Fusion Based on Discriminative Multiple Canonical Correlation Analysis. AI. 2026; 7: 5. DOI: 10.3390/ai7010005.

[8] Cortes C., Vapnik V. Support-vector networks. Machine Learning. 1995; 20: 273-297. DOI: 10.1007/BF00994018.