# In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging

## Supplementary Material

This supplementary material provides more details and results that are not included in the main paper due to space limitations. The content of this supplementary material involves:

A Discusstion of Intra-Similarity
B Discusstion of Inter-Similarity
C Computational Complexity Analysis of In2AB

## A. Discusstion of Intra-Similarity

In the main paper, we propose that the intra similarity observed in the panchromatic (PAN) image serves as an effective approximation of the intra similarity present in the hyperspectral image (HSI). The correlation map comparison results shown in Figure 1 demonstrate the visual effects of Table 2 in the main paper. It can be observed that the spatial correlation maps of the HSI and PAN images exhibit high similarity, demonstrating the high consistency in intra similarity patterns between them. This verifies our hypothesis that the intra similarity of the PAN image can effectively approximate the inherent spatial similarity within the HSI itself, thereby providing an enhanced content prior for HSI reconstruction.

## B. Discusstion of Inter-Similarity

In the main paper, we propose that the semantic features of the PAN image align with those of the corresponding HSI region. Specifically, texture and shape attributes should demonstrate a degree of consistency across modalities in corresponding image regions. Leveraging inter-similarity as a robust indicator for accurately reconstructed regions, we establish a correlation: higher feature weight scores correspond to more precisely reconstructed regions.

These high score areas can provide reliable cues to recover remaining uncertain areas by propagating intra-similarity patterns within the HSI itself. This allows reconstructing the highly ill-posed inverse problem in a more confident supervised manner, which provides new contextual information for HSI reconstruction.

## C. Computational Complexity Analysis of In2AB

To analyze the computational complexity of the proposed Intra-Inter Attention Block(In2AB), we omit the negligible position embedding and focus on comparing the attention calculation schemes. The inputs to In2AB are HSI features $\mathbf{X}$ and PAN features $\mathbf{G}$ with dimensions $\mathbb{R}^{H\times W\times C}$ and $\mathbb{R}^{H\times W\times \frac{C}{2}}$ respectively. After tokenization, the dimensions become $\mathbb{R}^{B\times N\times C}$ and $\mathbb{R}^{B\times N\times \frac{C}{2}}$, where $B$ is the number of token groups, $N$ is the number of tokens, and $C$ and $\frac{C}{2}$ represent the token dimensions.According to Eq. (13) and Eq. (14) in the main paper, the computational complexity of the intra-attention is:

$$O(\text{IntraAttention}) = B(N^2\frac{C}{2} + N(\frac{C}{2})^2). \quad (1)$$

And according to Eq. (17) in the main paper, the complexity of the inter-attention is:

$$O(\text{InterAttention}) = \frac{3}{2}BNC. \quad (2)$$

Therefore, the overall computational complexity of In2AB is:

$$O(\text{In2AB}) = BN^2\frac{C}{2} + BN(\frac{C}{2})^2 + \frac{3}{2}BNC, \quad (3)$$

where $BN^2\frac{C}{2}$ is the complexity of the intra-attention on spatial dimensions, $BN(\frac{C}{2})^2$ is for the intra-attention on spectral dimensions, and the last term is for the inter-attention.

The key advantages of these three attention components are their computational simplicity and efficiency:

- The intra-attention mechanism on spatial dimensions works similarly a window-based attention, but is computed within the PAN image and applied in HSI. In contrast to the window-based attention, this results in a reduction in computational complexity from $BN^2C$ to $BN^2\frac{C}{2}$. As demonstrated in Section A, it has been verified that the intra-similarity on the spatial dimensions of PAN effectively approximates the intra-similarity on the spatial dimensions of the HSI. Moreover the reconstructed HSI is not available in measurements.
- The intra-attention mechanism on spectral dimensions works like a global channel attention, capturing global dependencies along the spectral dimensions. Compared to the window-based attention, this better models spectral correlations. Additionally, considering the guidance is a grayscale PAN image, it lacks channel dependencies and color information along spectral dimensions. The intra-attention mechanism on spectral dimensions facilitates the explicit utilization of long-range spectral correlations within the HSI.

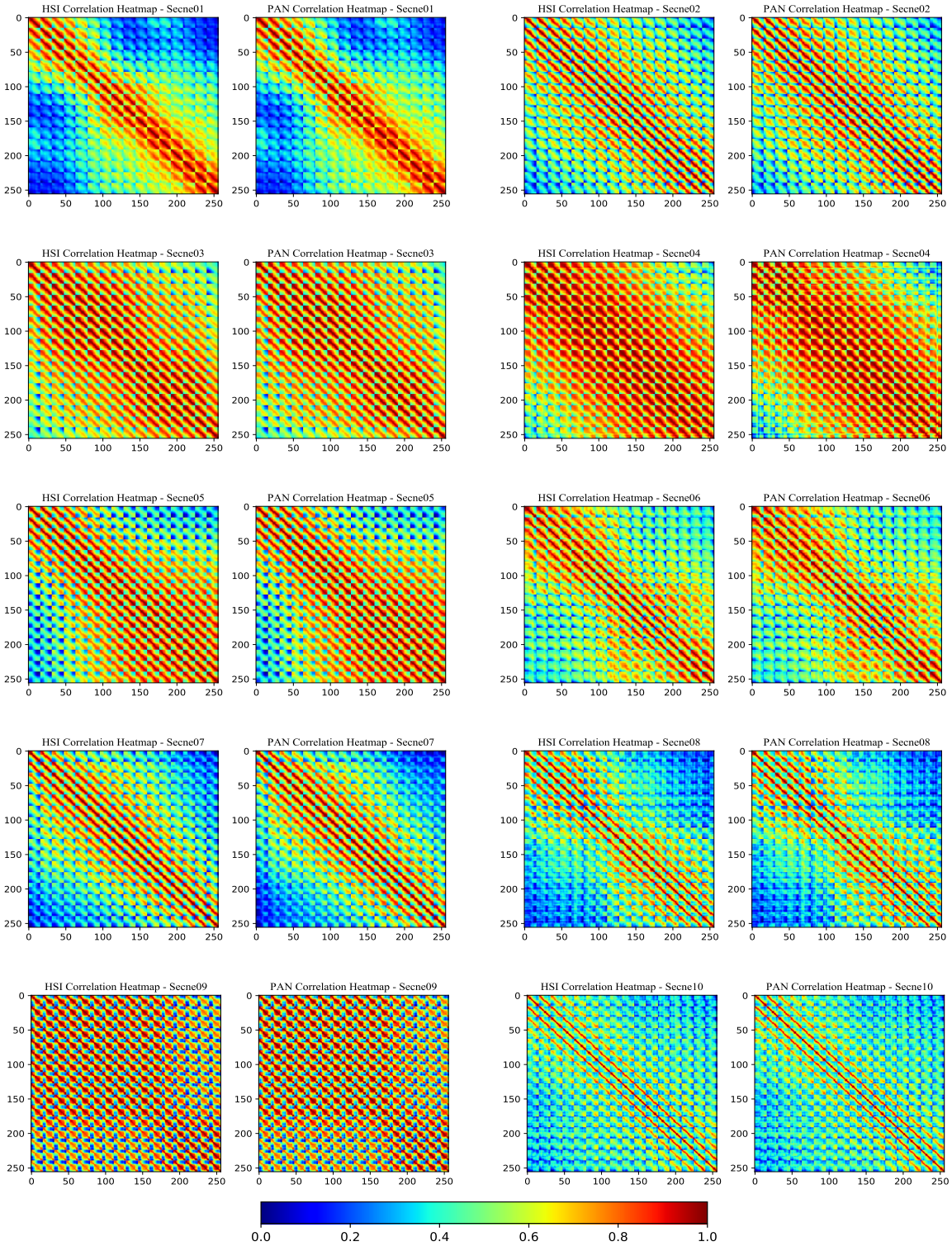


Figure 1. Visualization of correlation maps for HSIs and PAN images, demonstrated across 10 scenes of KAIST dataset.