A. APPENDIX

A.1. Visualization for Qualitative Results of Ablation Cases

Figure 6 presents visualizations of the ablation cases discussed in the main paper. Overall, the results show that the absence of SEM or CEM leads to fuzzier and less clear reconstructions, particularly in high-frequency regions. Notably, in (c) img098 of Urban100 in Figure 6, the CEM and SEM variants display less distinct patterns in the cement tiles compared to the ESCANet-Base. For luminance reconstruction, the proposed ESCANet-Base model outperforms the other ablation variants, especially in its ability to reconstruct the illuminated regions of the image.

A.2. Ablation Studies Extension

Table 3 presents the ablation cases for all five benchmark datasets on $\times 4$ SR. Across all ablation cases, removing the SEM or CEM layers significantly decreases performance. For example, removing the CEM layers results in a drop of 0.89dB and 0.57dB in Set5 and Set14, respectively. Additionally, the results show that both weight sharing and the removal of LayerNorm layers have similarly detrimental effects on performance.

A.3. Additional Visualization with State-of-the-art Methods

Figure 7 shows additional qualitative visualizations from Urban100, comparing our ESCANet with other state-of-the-art methods. We further examined the reconstruction of test images from Urban100. Compared to other methods, our proposed model demonstrates superior reconstruction, particularly in repetitive edge patterns. Specifically, in (a) img005 from Urban100, our model provides better reconstruction of the repeating window patterns at the corners.

A.4. Explanation of Linear Complexity

$X \in \mathbb{R}^{H \times W \times C}$ as an input tensor and Q and V are identical features with X, thus $X = Q = V$. k is the key vector obtained by GAP over H and W, thus involves summing over $H \times W$ for each of the C channels.

$$O_k = O(H \times W \times C) \tag{5}$$

The cross-correlation coefficient C_{Qk} is calculated using equation 2. The numerator term is the element-wise multiplication followed by a sum across channels for each spatial location. The denominator has two terms, each involving a sum of squared difference.

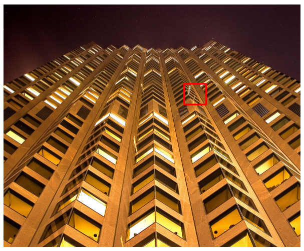
$$\begin{aligned} O_{num} &= O(H \times W \times C) \\ O_{den} &= O(H \times W \times C) + O(C) \\ O_{C_{Fk}} &= 2 \times O((H \times W \times C)) + O(C) \\ &= O(H \times W \times C) \end{aligned} \tag{6}$$

Attention weight matrix $\mathcal{A}$ is calculated as $\mathcal{A} = (1 - \sigma(C_{Qk}))^{\alpha}$ where $\sigma(C_{Qk})$ represents the sigmoid function applied element wise to C_{Qk}, thus having a complexity $O(H \times W)$. The power α is also an element-wise operation and the complexity remains $O(H \times W)$.

Thus, the overall complexity for calculating Efficient Self-Attention Block is:

$$\begin{aligned} O_{\mathcal{A}} &= O(H \times W) \\ O_{overall} &= O(H \times W \times C) + O(H \times W) \\ &= O(H \times W \times C) \end{aligned} \tag{7}$$

This is linear with respect to the spatial dimensions H and W and the number of channels C. Thus, our overall complexity is $O(N)$ or linear complexity.

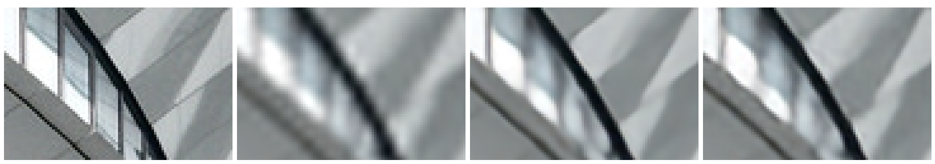


(a) img015 from Urban100

(a) HR patch (b) Bicubic (c) SEM ✗ (d) CEM ✗

(e) ESAB ✗ (f) LayerNorm ✗ [7] (g) ESCANet-Base

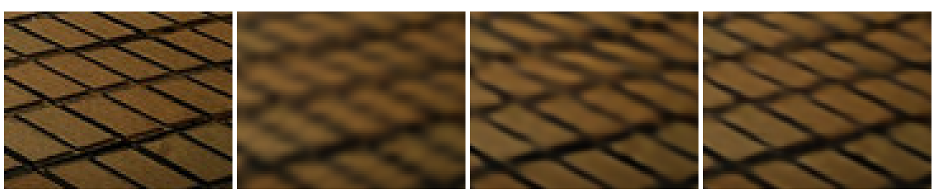
(b) img052 from Urban100



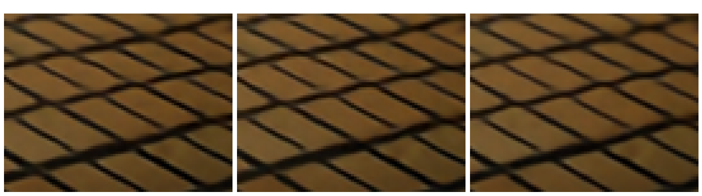
(a) HR patch (b) Bicubic (c) SEM ✗ (d) CEM ✗

(e) ESAB ✗ (f) LayerNorm ✗ (g) ESCANet-Base

(c) img098 from Urban100



(a) HR patch (b) Bicubic (c) SEM ✗ (d) CEM ✗



(e) ESAB ✗ (f) LayerNorm ✗ (g) ESCANet-Base

Fig. 6: Visual comparison of ESABNet and its ablation variants.

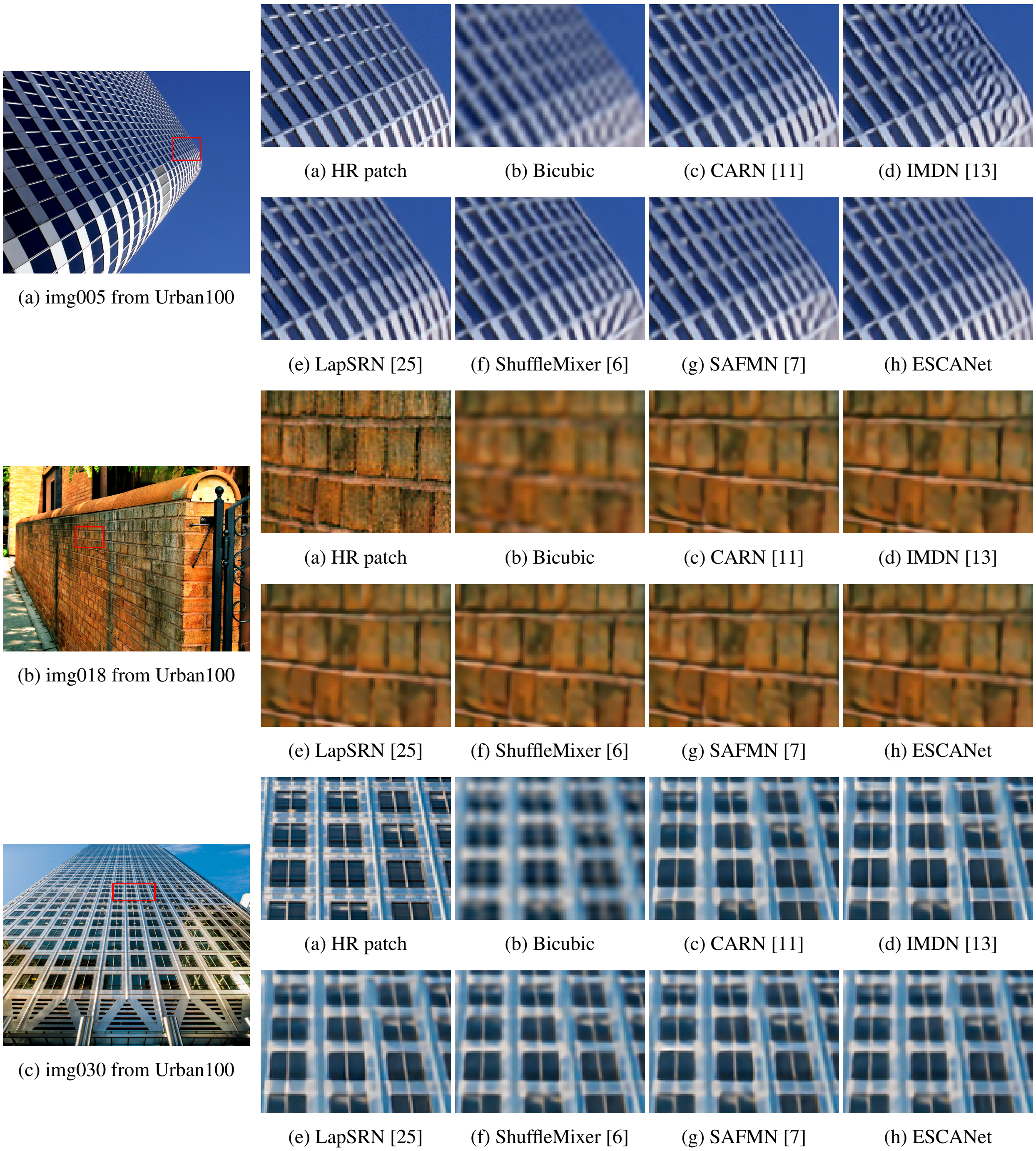


Fig. 7: Additional visual comparisons for ×4 SR on the Urban100 dataset.s Our proposed method reconstructs more details of patterns and edges.

Table 3: Ablation study on variations of blocks for ESCANet on $\times 4$ SR Set5, Set14, B100, Urban100, and Manga109 datasets.

Ablation Case	**Variations**	**#Params [K]**	**#FLOPs [G]**	**Set5**	**Set14**	**B100**	**Urban100**	**Manga109**
Base Model	SEM ✓ CEM ✓	354	19	32.35/0.8965	28.72/0.7839	27.65/0.7380	26.20/0.7871	30.84/0.9109
Core Blocks	SEM ✗ CEM ✓	233	13	31.83/0.8900	28.39/0.7770	27.43/0.7307	25.55/0.7658	29.90/0.8900
	SEM ✓ CEM ✗	147	7	31.46/0.8823	28.15/0.7686	27.29/0.7255	27.28/0.7253	29.17/0.8877
Self-Attention	ESAB ✗	354	19	32.31/0.8959	28.70/0.7832	27.63/0.7379	26.16/0.7862	30.75/0.9100
Layer Normalization	LN ✗	354	19	32.26/0.8956	28.67/0.7827	27.61/0.7368	26.11/0.7840	30.65/0.9088
CEM Projection Weight Sharing	Weight Sharing ✓	304	19	32.26/0.8956	28.64/0.7823	27.59/0.7367	26.13/0.7844	30.69/0.9090