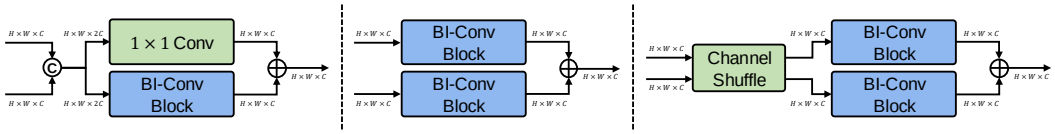

Figure 1: **Illustration of different fusion modules. (a) BFB** (Binary Fusion Block), from BRVE [3], uses one binary convolution (BI-Conv) block for dimension transformation and a full-precision 1×1 convolution (Conv) to maintain full-precision information. BFB fuses features but adds extra computation with 1×1 Conv. **(b) BiFD** (Binarized Fusion Block), by BiSRNet [4], uses two BI-Conv blocks to process and combine two inputs. It maintains the dimensions of the BI-Conv, enabling the use of bypass shortcuts for transferring full-precision information. However, independent Conv blocks struggle with huge feature discrepancies, resulting in insufficient fusion. **(c) CS-Fusion** (Channel-Shuffle Fusion), developed in our method, merges features via channel shuffle before BI-Conv, effectively fusing varied features without extra parameters. [To **Reviewer mEsa**]

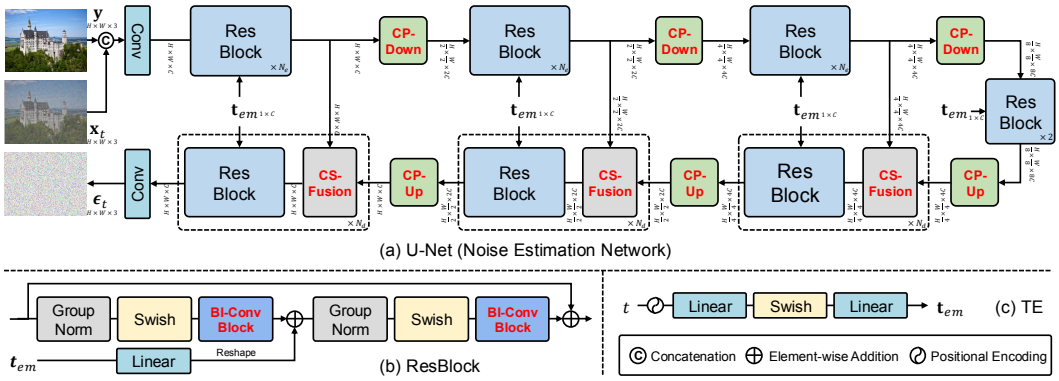

Figure 2: **Overview of the proposed BI-DiffSR.** Newly designed modules are highlighted in **red**, including: CP-Down, CP-Up, CS-Fusion, and BI-Conv Block. **(1) CP-Down/Up** (Consistent-Pixel-Downsample/Upsample), address the issue of dimension mismatch in the UNet architecture. **(2) CS-Fusion** (Channel-Shuffle Fusion), effectively merges features in skip connections. **(3) BI-Conv Block** (Binary Convolution Block), employs timestep-aware redistribution (TaR) and activation function (TaA), enhancing the capabilities of the binarized module. [To **Reviewer KWS7**]

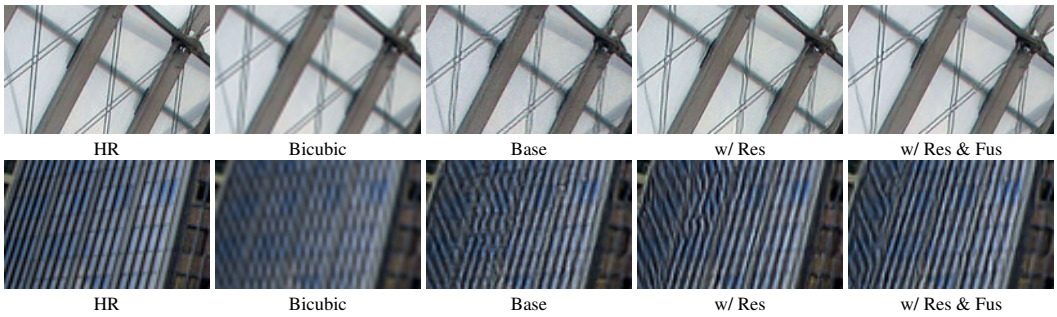

Figure 3: **Visual results ($\times$2).** We compare visual results under different designs to further analyze **challenge I:** dimension mismatch and **challenge II:** fusion difficulty. **Base:** The basic model uses binarized convolution (BI-Conv) without the residual connection and directly adds two features in the skip connection for fusion. **w/ Res:** Uses the residual (Res) connection within BI-Conv. **w/ Res & Fus:** Builds on the w/ Res model by further using CS-Fusion (Fus) to merge different features in the skip connection. **Analysis:** Comparing the results of these three models, w/ Res restores more details than Base, while w/ Res & Fus achieves the best restoration results. These outcomes further support the observations of challenges I and II. [To **Reviewer KWS7**]