A. Implementation details

Our proposed SIA adopts 10 block-level transformation, namely VShift, Hshift, VFlip, HFlip, Rotate, Scale, Add Noise, Resize, DCT, and Dropout. Here we provide the implementation details of these transformations, respectively. In particular, for an given image block $\boldsymbol{x} \in \mathbb{R}^{3\times H\times W}$, we can implement the transformations as follows:

- **Vertical shift (VShift)**: We roll the image block $\boldsymbol{x}$ along the vertical axis by a randomly selected length $h < H$.
- **Horizontal shift (Hshift)**: We roll the image block $\boldsymbol{x}$ along the horizontal axis by a randomly selected length of $w < W$.
- **Vertical flip (VFlip)**: We flip the image block $\boldsymbol{x}$ vertically along the horizontal axis, in which the top of the image block becomes the bottom, and the bottom becomes the top.
- **Horizontal flip (HFlip)**: We flip the image block $\boldsymbol{x}$ horizontally along the vertical axis, in which the left of the image block becomes the right, and the right becomes the left.
- **Rotate**: We turn the image block $\boldsymbol{x}$ clockwise by 180° around its center point, in which the top-left of the image block becomes the bottom-right, and the top-right becomes the bottom-left.
- **Scale**: We multiply a random scale factor $\alpha \in (0, 1)$ with the pixel in the image block to scale $\boldsymbol{x}$ into $\alpha \cdot \boldsymbol{x}$.
- **Add Noise**: We add a uniform noise $r \in [0, 1]^{3\times H\times W}$ to the image block x and clip them into $[0, 1]$ to obtain the transformed image block $\mathrm{Clip}(\mathrm{x} + \mathrm{r}, 0, 1)$.
- **Resize**: We resize $\boldsymbol{x}$ into x' with the size of $3 \times w \times h$ ($w < W$ and $h < H$) and resize x' back into the size of $3 \times W \times H$ using bilinear interpolation.
- **DCT**: We first transform x to the frequency domain using Discrete Cosine Transformation (DCT). Then we mask the top 40% highest frequency with 0 and recover the image in the time domain using Inverse Discrete Cosine Transformation (IDCT).
- **Dropout**: We randomly mask 10% pixels of $\boldsymbol{x}$ with 0 to obtain the new image block.