(a) NeRV with position embedding (PE)

t: Image index
PE: Position Encoding

(b) CNeRV with content-adaptive embedding (CAE, ours)

Step 1) Dividing image into blocks

$$\mathbb{R}^{C \times H \times W} \rightarrow \mathbb{R}^{C \times M \times \frac{H}{M} \times N \times \frac{W}{N}}$$

Step 2) Compute embedding for each block:

$$\mathbb{R}^{C \times P \times Q} \quad where \Gamma(c, p, q) = \sum_{x,y} \cos(b^p \pi x) \cos(b^q \pi y) Img(c, x, y)$$

Step 3) Concat block embedding:

$$\mathbb{R}^{CPQ \times M \times N}$$

Step 4) Reduce embedding length:

$$\mathbb{R}^{L \times M \times N}$$

(c) CNeRV encoder: Content-adaptive Encoding

Raw Embedding

Conv 1×1

Image Embedding

(d) CNeRV decoder

Image Embedding → Conv 1×1 → Reshape → Conv 3×3 ∘∘∘ → Image Output

$\mathbb{R}^{L \times M \times N}$ → $\mathbb{R}^{dhw \times M \times N}$ → $\mathbb{R}^{d \times Mh \times Nw}$ - - - → $\mathbb{R}^{C \times H \times W}$

Block-wise decoding   Image-wise decoding