(a) Distorted Image Projector

Distorted Image $\mathbf{I}$

Distorted Image Encoder: Conv → BatchNorm → ReLU → Conv → BatchNorm → ReLU

$\mathbf{X}_D$ → Concatenation → $\hat{\mathbf{X}}$

Distorted Image Decoder: DeConv → BatchNorm → ReLU → DeConv → Tanh

Enhanced Image $\hat{\mathbf{I}}$
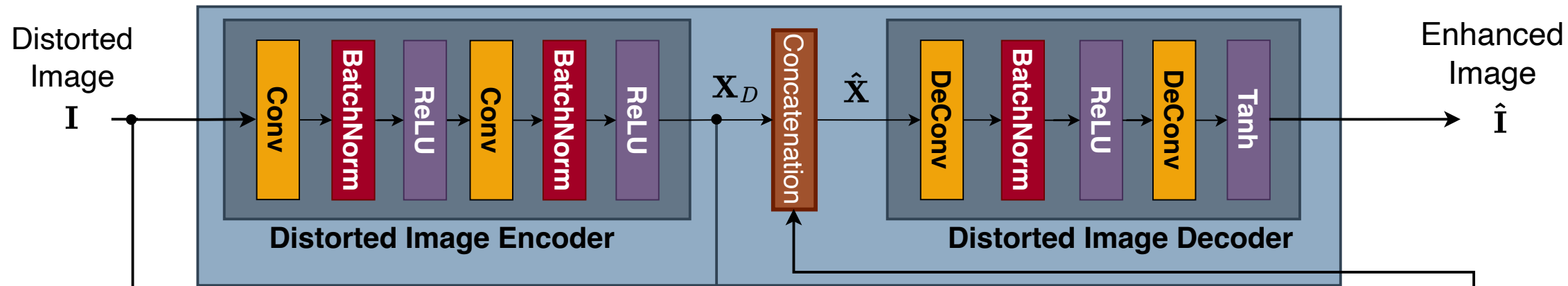
(b) Distortion Level Encoder

Conv → BatchNorm → ReLU → Conv → BatchNorm → ReLU

$\mathbf{X}_{dl}$
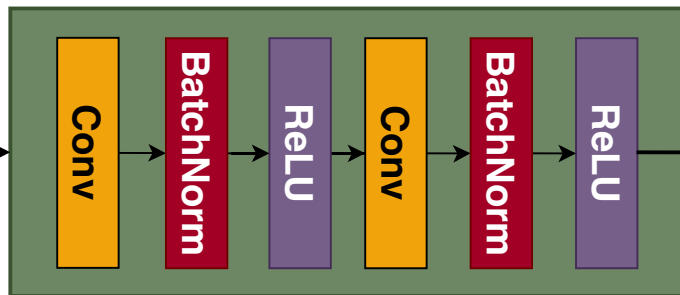
(c) Variational Cross-Attention Module

Patch extraction with positional encoding → $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ → Multi-head Variational Cross-Attention (MVCA) → $\mathbf{Z}$ → $\oplus$ → Layer Norm → $\odot$ → $\mathbf{X}_{af}$ Attention-fused Features