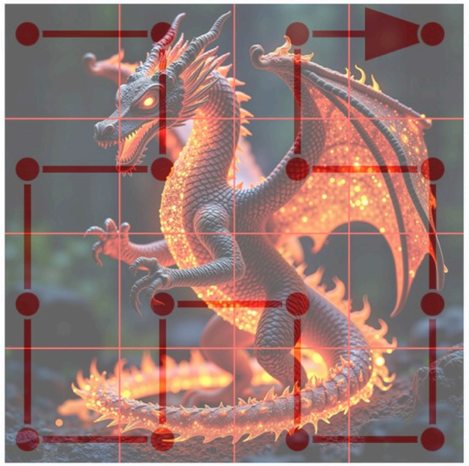

High-Res X
Partitioned to Local Windows

Hilbert-Order Permutation

Local Self-Attention with Contiguous Mask Layout

Guidance Via Attention

Low-Res X_{lr}

Position Scaling