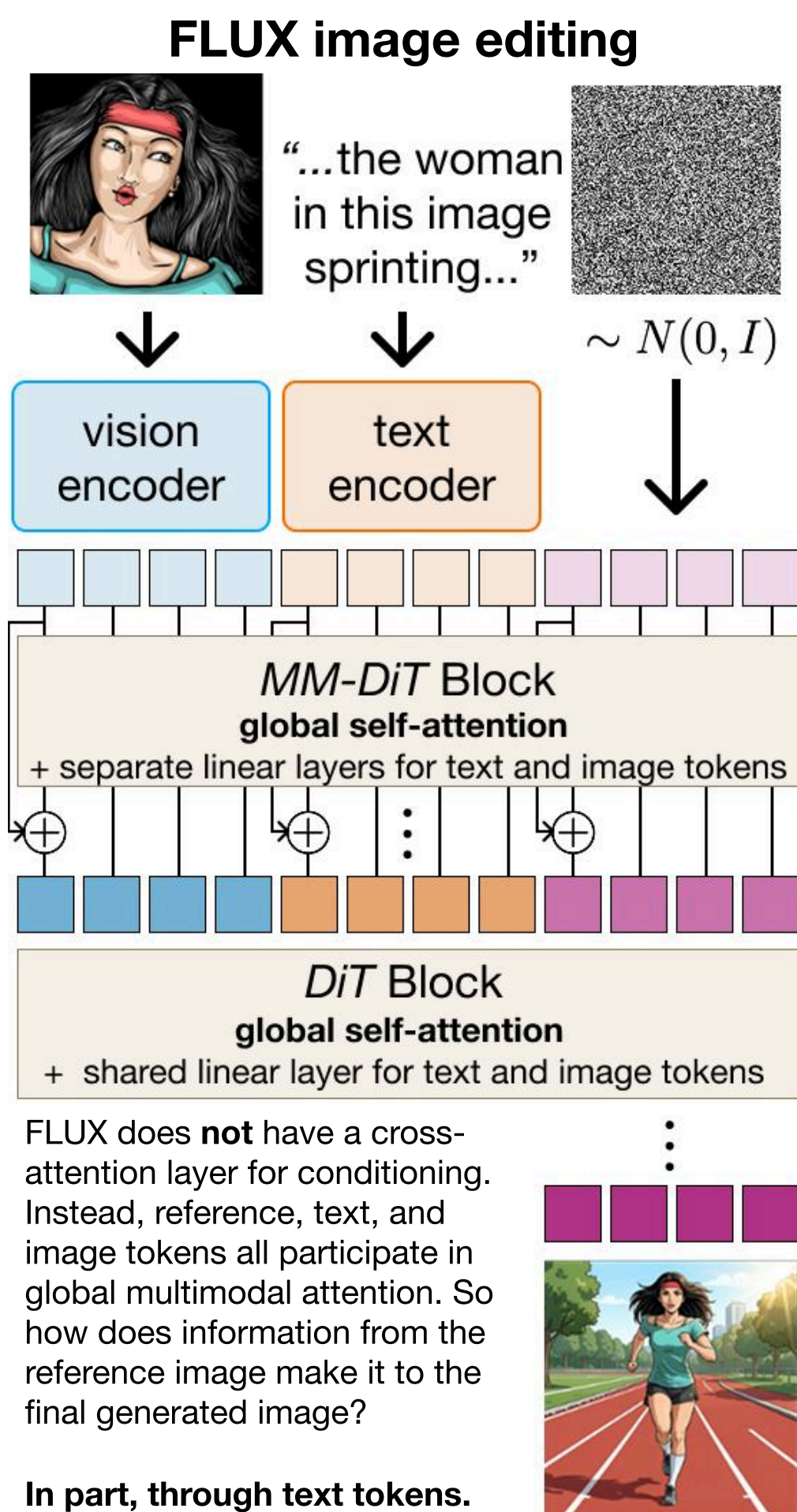


Vision-Language Binding in In-Context Image Generation

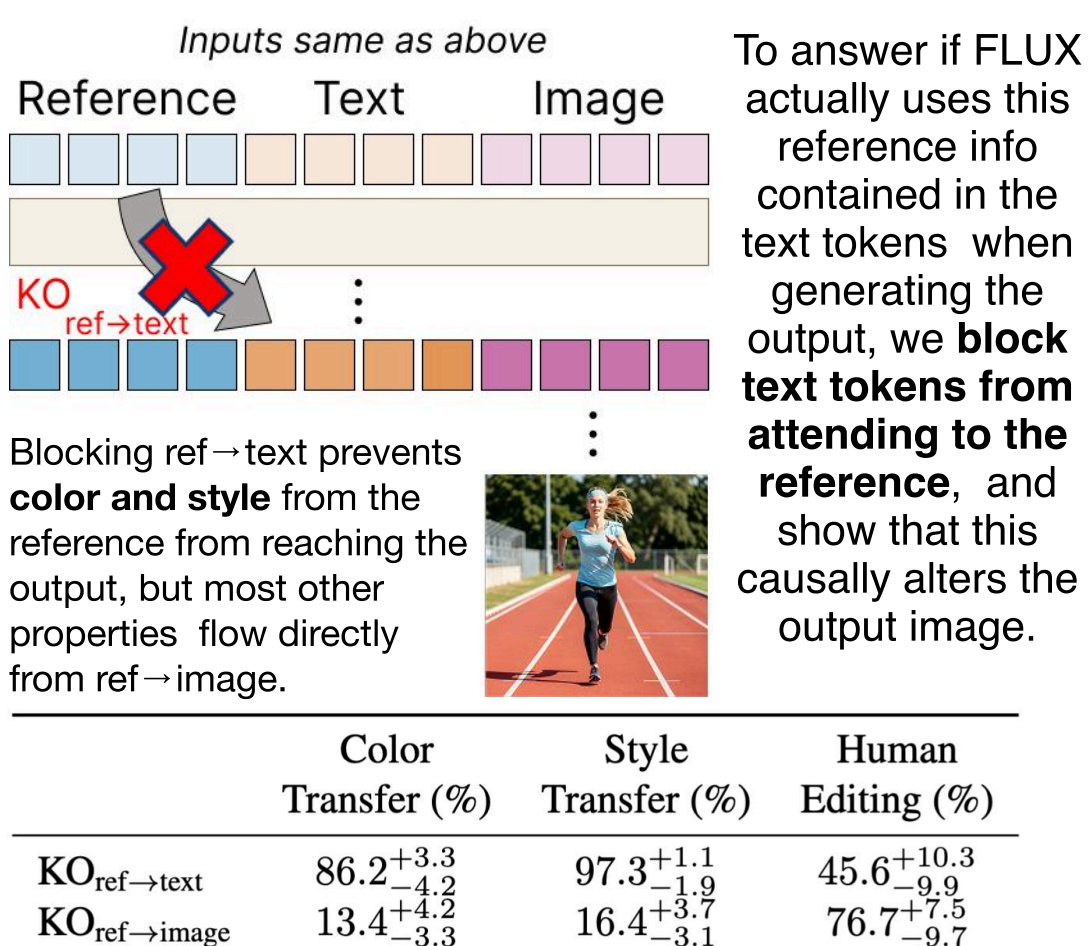
Chris Ge, Rohit Gandikota, Antonio Torralba, Tamar Rott Shaham

When editing a reference image, FLUX.2 writes the reference image content through text tokens rather than writing directly to the output image.

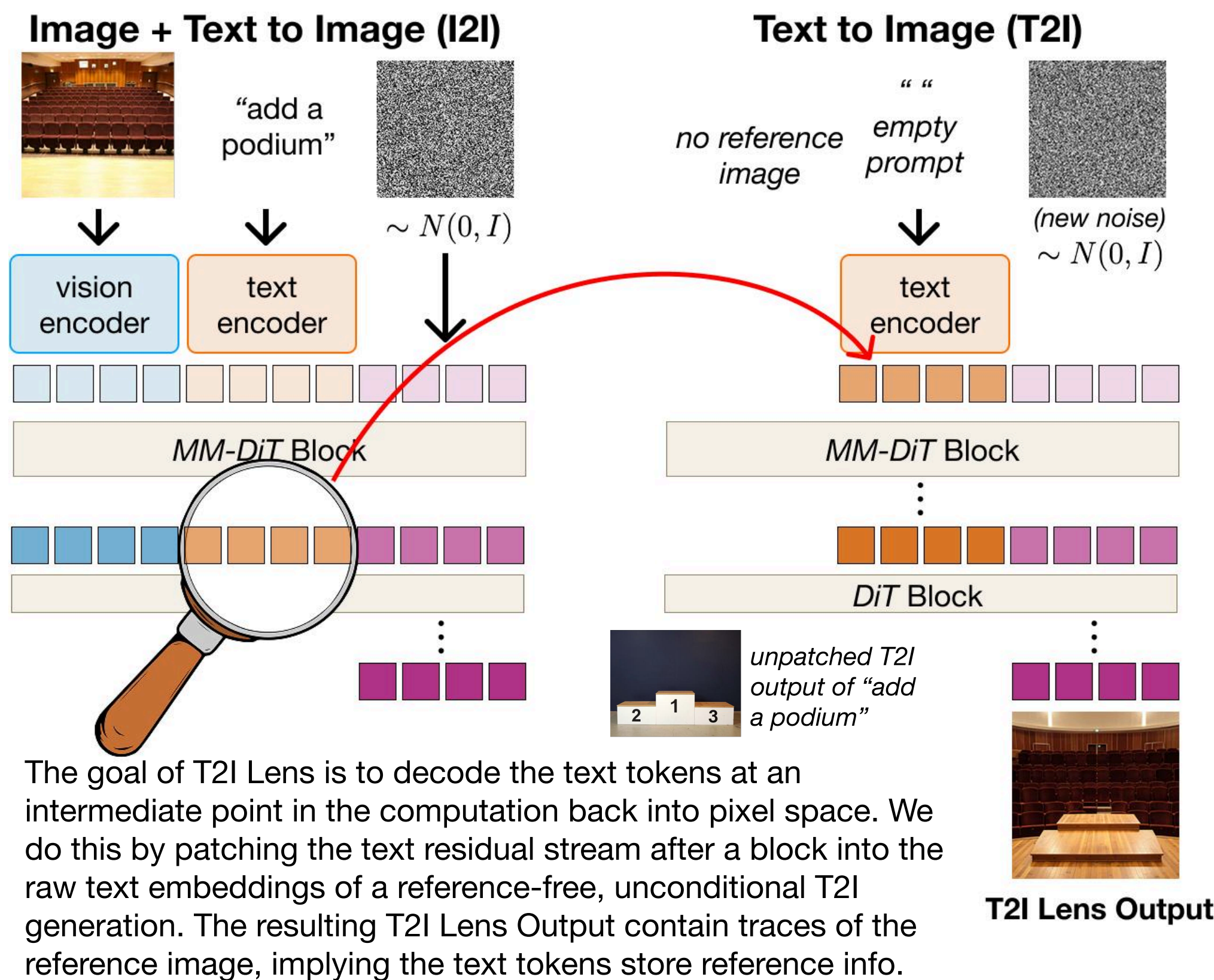
1 Background



3 Causal Evidence 1: Attention Knockout



2 Qualitative Evidence: T2I Lens



4 Causal Evidence 2: I2I-to-I2I Patching

