Vision Token Interpretability: Static NN vs LN-Lens

Static NN

(Input Embeddings)

LN-Lens (Ours)

(Contextual Embeddings)

1. "Michele"

2. "blinds"

3. "rien"

4. "White"

5. "handwriting"

1. "ouver" → "steeple white louvered windows."

2. "ouver" → "siding of building with louvered..."

3. "ouver" → "a dark green louver"

4. "ouver" → "kitchen window with black curtai..."

5. "-window" → "gabled roof on multi-windowed br..."