a) OLMo-7B + CLIP • Layer 0

LN-Lens

1. "...ont of [bulky]moni..." (0.13) — from LLM L8
2. "...th two [clot]hs" (0.13) — from LLM L8
3. "...olored [babu]shka" (0.13) — from LLM L8
...

Embedding Matrix

1. "," (0.09)
2. .object (0.08)
3. _x (0.08)
...

LogitLens

1. iller (4.88)
2. Staff (4.32)
3. hil (4.27)
...
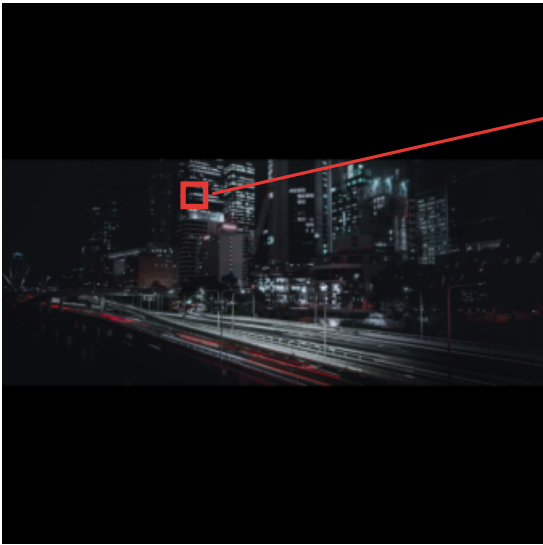
b) LLaMA3-8B + DINOv2 • Layer 4

LN-Lens

1. "...g at a [building]" (0.23) — from LLM L8
2. "...all of [building]" (0.22) — from LLM L8
3. "...e pink [facade]of a..." (0.22) — from LLM L8
...

Embedding Matrix

1. ins (0.08)
2. ayer (0.08)
3. peated (0.07)
...

LogitLens

1. itsu (10.44)
2. -origin... (9.41)
3. Integr (9.39)
...

c) Qwen2-7B + SigLIP • Layer 16

LN-Lens

1. "...id skys[crap]ers,..." (0.33) — from LLM L16
2. "...of skys[crap]ers" (0.33) — from LLM L16
3. "...ny skys[crap]ers ..." (0.31) — from LLM L16
...

Embedding Matrix

1. repair ... (0.09)
2. racuse (0.08)
3. TSRMLS (0.08)
...

LogitLens

1. ([[(15.04)
2. rvine (14.46)
3. ("- (14.09)
...

d) OLMo-7B + SigLIP • Layer 24

LN-Lens

1. "... white [dots.]" (0.19) — from LLM L8
2. "...ular ro[ck] wit..." (0.19) — from LLM L8
3. "... black [spec]kles" (0.19) — from LLM L8
...

Embedding Matrix

1. ueur (0.08)
2. atat (0.08)
3. Ting (0.08)
...

LogitLens

1. holes (10.45)
2. cke (8.17)
3. hole (8.00)
...