Table 1: Overview of the encoder networks for the different architectural and modeling techniques considered. APE: Absolute Positional Encoding. OPE: Object Positional Encoding. RPE: Relative Positional Encoding. VT: Visual Tokens. Registers: Register tokens. PEMixer: Positional Encoding Mixer, where ”vec weighted sum” signifies a vector-weighted sum.

	Architectural			Modeling	
Encoder	APE	RPE	PEMixer	VT	Register
ResNet	—	—	—	—	—
Vanilla ViT	learned	—	sum	—	—
Grid ViT	2D-sincos w/OPE	RoPE	vec weighted sum	✓	✓
LLaDA	—	RoPE	—	✓	—