

Multi-camera Images



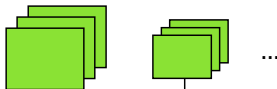
Positional
embedding



query

Encoder

Image features



+

+

Cross
attention

Conv

Cross
attention

Conv

Decoder

