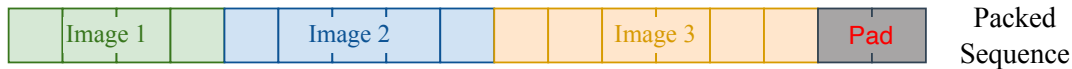
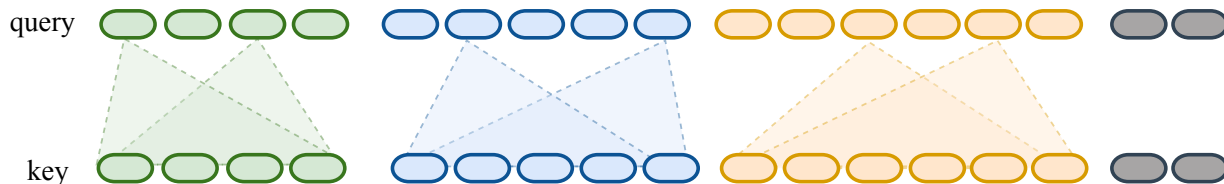


Self-Attention

Each token's receptive field is restricted to the tokens within the same example

Pad tokens are masked out
from attention



Data preprocessing

