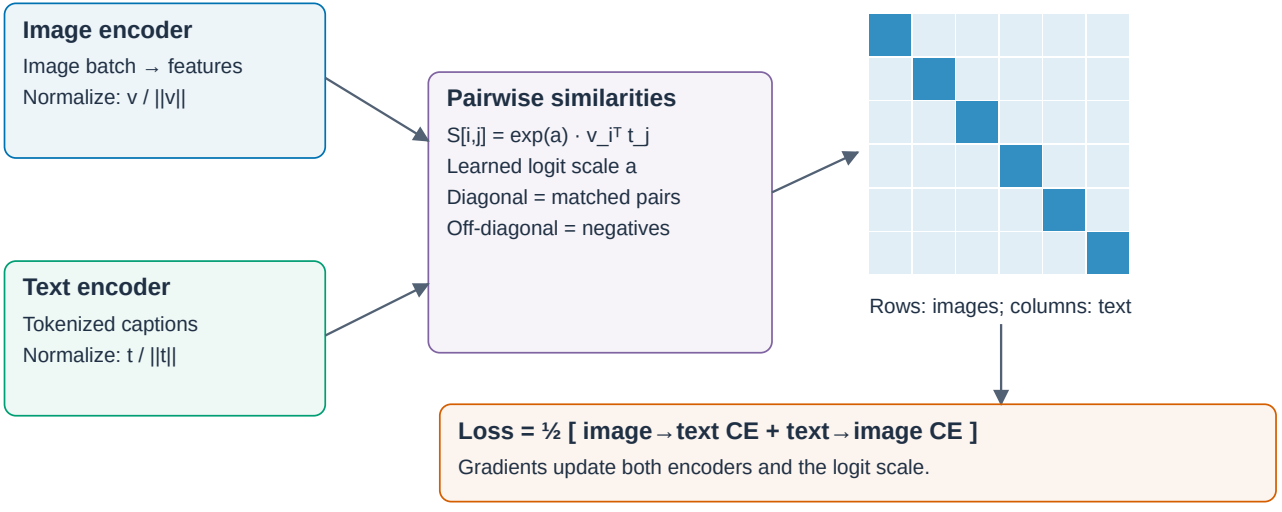
CLIP: align paired images and text in a shared space

Symmetric contrastive training over a minibatch of N image–text pairs (Radford et al., 2021).



Original educational redraw · schematic values only · SVG text and objects remain editable