

CNN

\mathbf{x}_{FC7}

(with *estimated* attribute, i.e., $r(\mathbf{x})=t_i$)

Encoder

Decoder

\mathbf{x}_{FC7}

(with *desired* attribute, i.e., $r(\mathbf{x})=t_j$)

Transformer