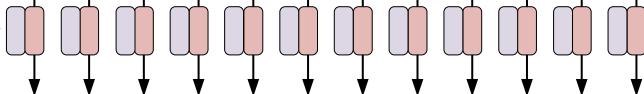


Linear Projection of Flattened Patches

Position Embedding



$L_1 \times$

Transformer Block

Normalization

Encoder Final Hidden States

$L_2 \times$

Transformer Block

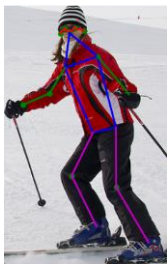
Decoupled Block

$L_3 \times$

Deconvolution Block

Heatmap

Predictor



DVPose Backbone

DVPose Head