

Input Data

CNN Block

3×3 Conv

Batch Norm

ReLU

Max Pool

Residual Block (2 consecutive blocks)

3×3 Conv

Batch Norm

ReLU

3×3 Conv

Batch Norm

+

Self Attention Block

key Conv

Transpose

×

SoftMax

Attention Map

query Conv

value Conv

+

×

Fully Connected Layer

Predictions

