Local Geospatial Attention Extraction

CNN Encoder

Overhead Image

CNN Encoder

Overhead Feature

Grid Level

Orientation

Haversine Distance

Orientation Feature

Distance

Compass Bearing

Distance Feature

Near Panorama

CNN Encoder

F_i

Near Panorama Feature map

Max Pooling

Avg Pooling

$h \times w \times c$

$3 \times 3 Conv$

$5 \times 5 Conv$

Concat

Concatenated Feature Map

$1 \times 1 Conv$

$Softmax$

$A_{i,t}$

Frobenius Inner Product

Reshape

$P_{i,t}$

Upsample

Local Geospatial Attention

Resnet50 Pretrained model

LayerNorm

Relu

CNN Encoder