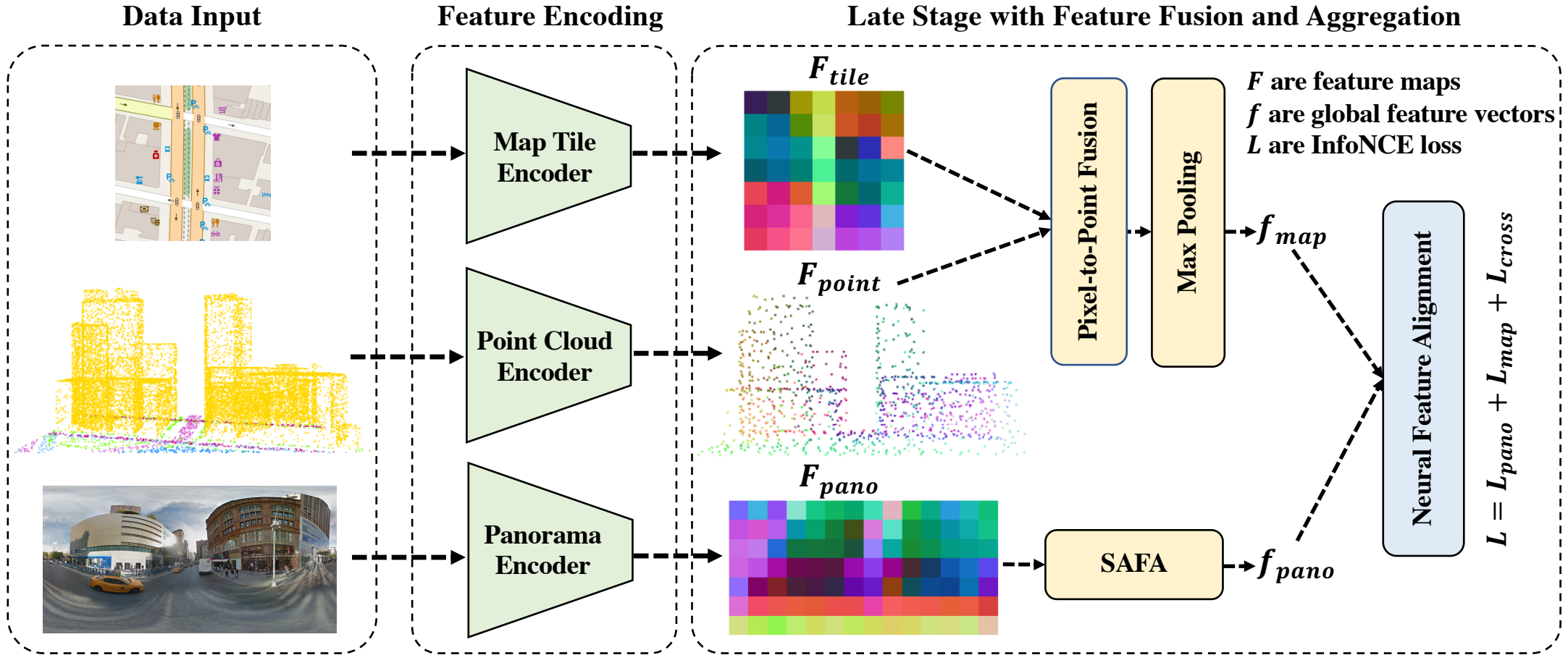
Data Input
Feature Encoding
Late Stage with Feature Fusion and Aggregation
Map Tile Encoder
Point Cloud Encoder
Panorama Encoder
F_{tile}
F_{point}
F_{pano}
Pixel-to-Point Fusion
Max Pooling
SAFA
f_{map}
f_{pano}
Neural Feature Alignment
F are feature maps
f are global feature vectors
L are InfoNCE loss
$L = L_{pano} + L_{map} + L_{cross}$